Published in final edited form as:

Ear Hear. 2013 ; 34(3): 300–312. doi:10.1097/AUD.0b013e31826fe77a.

Using a vocoder-based frequency-lowering method and spectral enhancement to improve place-of-articulation perception for hearing-impaired listeners

Ying-Yee Kong^{a),b),*} and Ala Mullangi^{b)}

^{a)}Department of Speech Language Pathology & Audiology, Northeastern University, Boston, MA 02115, USA

^{b)}Bioengineering Program, College of Engineering, Northeastern University, Boston, MA 02115, USA

INTRODUCTION

High-frequency sensorineural hearing loss is the most common type of hearing loss. Recognition of speech sounds that are dominated by high-frequency information, such as fricatives and affricates, is challenging for listeners with this hearing loss configuration. Furthermore, perception of place of articulation is difficult because listeners rely on high-frequency spectral cues for the place distinction, especially for fricative consonants (e.g., labiodental fricative/f/ vs. alveolar fricative/s/ vs. palatal fricative/ʃ/) (Harris 1958). Individuals with a steeply sloping severe-to-profound (> 70 dB HL) high-frequency hearing loss may receive limited benefit for speech perception from conventional amplification at high frequencies (e.g., Hogan & Turner 1998; Vickers et al. 2001; Baer et al. 2002). This could be attributed to the loss of inner hair cells and loss of neural function in the region that has severe or profound hearing loss.

One approach to this problem is *frequency lowering*, which involves conveying high-frequency acoustic information to a lower-frequency region where the hearing loss is less severe. A number of signal processing methods have been proposed for frequency lowering, including channel vocoding (e.g., Lippmann 1980), slow playback (e.g., Beasley et al. 1976), frequency transposition (e.g., Velmans 1974), and frequency compression with proportional frequency shift (e.g., Turner & Hurtig 1999) or nonproportional frequency shift (e.g., Reed et al. 1983). Comprehensive reviews of these processing methods are provided by Braida et al. (1979) and Simpson (2009). Currently, the two commonly used methods to achieve frequency lowering are frequency transposition and frequency compression. A brief description of each method and the corresponding speech recognition results is provided below.

Frequency Transposition

In the frequency transposition approach, high-frequency acoustic signals are shifted to a lower-frequency region and then the transposed signal is added to the unprocessed low-frequency signal. This approach is currently used by the Widex hearing aid company. Widex's system ("Audibility Extender") first identifies the most prominent peak in the signal spectrum in the source octave above a pre-determined frequency (start frequency) and then linearly lowers its frequency by one octave (e.g., Kuk et al. 2009). Sounds below the

*Corresponding author: Address: 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.

start frequency are not modified. To reduce masking from the transposed signal, the transposed signal is bandpass filtered around the transposed peak frequency with a one octave bandwidth. The transposed signal is then amplified and mixed with the original signal. Despite the bandpass filtering around the transposed peak in Widex's system, the transposed signal can still mask useful low-frequency speech cues.

Auriemma et al (2009) showed that consonant recognition was better with frequency transposition than with conventional amplification for hearing-impaired children with severe-to-profound high-frequency hearing loss when speech stimuli were presented at a soft level of 30 dB HL. However, no benefit was evident when speech stimuli were presented at a conversational level of 50 dB HL. Kuk et al. (2009) reported significant frequency-lowering benefit in quiet and in noise for a group of adult listeners with a severe-to-profound sensorineural hearing loss above 2000 Hz. However, the participants in Kuk et al.'s study received two months of training with the frequency transposition algorithm and no training was provided for the control condition (i.e., conventional amplification). It is possible that the transposed benefit would have been reduced if extensive training had also been provided for the control condition (Fullgrabe et al. 2010). Thus, it is unclear if the observed benefit was really due to the use of frequency transposition or was simply the result of a learning effect.

Apart from the algorithm used by Widex, the frequency transposition method has also been explored by other researchers recently. Robinson et al. (2007) estimated the edge frequency (f_e) of dead regions in hearing-impaired listeners, a frequency corresponding to the characteristic frequency of the functioning inner hair cells and/or neurons immediately adjacent to the dead region (Moore 2004). High-frequency components ($2f_e$ to $2.7f_e$) were linearly transposed to a destination band (f_e to $1.7f_e$). The transposed signal was then superimposed onto the original speech. Frequencies below the edge frequency were amplified without modification. Robinson et al. used a conditional frequency transposition method in which only speech signals that were dominated by high frequencies, determined by the ratio of power in the high-frequency region to the power in the low-frequency region, were subjected to frequency transposition. Seven subjects with high-frequency sensorineural hearing loss and dead regions above 800 to 1500 Hz were tested on a consonant identification task in quiet. As a group, significant transposition benefit was observed for affricate consonant identification compared to the control condition (i.e., lowpass [LP] filtering at $1.7f_e$ without frequency transposition). Information analysis revealed no significant difference for the perception of voicing, manner, and place-of-articulation features between the transposed and control conditions.

Frequency Compression

In the frequency compression approach, the bandwidth of the original speech signal is reduced by downwardly shifting the high-frequency components. One form of frequency compression involves nonproportional shifting with increasing amounts of frequency lowering for relatively high input frequencies. This nonproportional shifting approach, known as "nonlinear frequency compression" (Simpson et al. 2005) is currently used by the Phonak hearing aid company. Phonak's system ("Sound Recover") compresses a wide range of frequencies above a pre-determined cutoff frequency into a narrower frequency range of output signals. Frequencies below the cutoff frequency are amplified without modification. A loudness balance procedure is performed between a sound at the reference frequency below the cutoff frequency and a sound at a test frequency in the compressed signal. Frequencies within the compressed signal are amplified to produce equal loudness. Unlike frequency transposition, there is no spectral overlap between the shifted and unshifted signals for frequency compression, reducing the risk of masking. The disadvantage of nonlinear frequency compression is that the frequency ratios of components in the high-

frequency signal are not preserved, reducing the spectral contrast of speech sounds. This could potentially have a negative effect on speech recognition.

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 about half of their subjects showed better phoneme recognition with the nonlinear frequency compression scheme than with conventional amplification, but the other half did not. Simpson et al. (2006) tested seven listeners with steeply sloping severe-to-profound hearing loss at frequencies above 1000 Hz, but normal or near-normal hearing at low frequencies, on a consonant identification task. They reported no significant difference in performance between nonlinear frequency compression and conventional amplification. A more recent study by Glista et al. (2009) reported significant improvement with nonlinear frequency compression compared to conventional amplification for consonant identification and plural recognition in some of their adult and child participants with high-frequency hearing loss.

In summary, frequency transposition and frequency compression provided benefit for speech recognition compared to conventional amplification for some individuals with sloping high-frequency sensorineural hearing loss. The improvement was mainly found for fricative and affricate perception. The perception of other consonant classes (e.g., stops, nasals, and semivowels), consonant features (e.g., place of articulation) and vowels was largely unaffected.

The frequency-lowering system presented in this paper differs from the frequency transposition and compression systems described above in several ways: First, we used a vocoder-based method similar to that described by Posen et al. (1993) instead of the fast Fourier Transform (FFT) method used in other systems. Second, the system involves a combination of frequency transposition and frequency compression. Third, unlike systems used in the Widex and Phonak devices, in which frequency lowering is activated for all frequency components above a start frequency, frequency lowering is conditional in the system used here. Fourth, we incorporated a novel spectral enhancement method in the system. This was designed to enhance the spectral differences between fricative consonants that differ in place of articulation.

The new vocoder-based system was first described by Kong and Mullangi (2012). They provided perceptual data from normal-hearing listeners with simulated severe-to-profound high-frequency hearing loss to demonstrate the potential benefit of the frequency-lowering system for speech perception. The system used in the present study is similar to that described by Kong and Mullangi (2012), but with a significant difference in the method used for enhancement of spectral cues. In the present paper, we first describe the modified vocoder-based frequency-lowering system. We then present perceptual data from a group of hearing-impaired listeners. The perceptual study also evaluated the effect of degree of high-frequency hearing loss on the benefit from the frequency-lowering system. The results showed significant frequency-lowering benefit for the perception of frication, affrication, and place of articulation.

A VOCODER-BASED FREQUENCY-LOWERING SYSTEM

System of Posen et al (1993)

A vocoder-based frequency-lowering system was described by Posen et al. (1993). Their system divided the high-frequency (1000 to 5000 Hz) speech signals into four analysis bands. The system measured the output level from each analysis band to determine the level of four bands of low-frequency noise. The four high-frequency analysis bands and the four low-frequency synthesis filters (center frequencies from 397 to 794 Hz) were monotonically

related in that the lowest analysis band controlled the level of the lowest synthesis band, the second-lowest analysis band controlled the level of 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. The four low-frequency narrow-band noise signals that carried the high-frequency speech information in the original signal were summed and added to the original speech signal. In this system, frequency lowering was conditional; it was only activated for speech signals that were dominated by high frequencies.

Modified System

Overview—We made two modifications to the system of Posen et al.: We used a different conditional frequency-lowering rule and employed a method that enhances spectral differences among fricatives differing in place of articulation. Figure 1 shows a block diagram of the analysis and processing stages in this system. In the system, there are two analysis stages. First, the system separates speech sounds into two classes, sonorants and nonsonorants, based on periodicity detection over the frequency region above 400 Hz. Only the nonsonorant sounds which have aperiodic high-frequency energy are subjected to further analyses and frequency-lowering processing (conditional lowering). The second analysis stage involves separation of high-frequency frication sounds into three groups based on the spectral information related to place of articulation (classification of nonsonorant sounds). The spectrum of the low-frequency transposed signal used in the system is then modified based on the classification results (spectral enhancement). Finally, four or six bands of low-frequency noise signals are added to the original speech (frequency lowering). Signal processing was performed offline. Apart from the periodicity detection analysis, which used a 40-ms analysis window, the remaining analysis (nonsonorant sound classification) and processing (spectral enhancement and frequency lowering) steps were performed every 20 ms, similar to the system of Posen et al.

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 made by Shannon et al. (1999) and the two child speakers were recorded in our laboratory. The original speech was digitally recorded at a sampling rate of 44,100 Hz. We divided the stimuli into two sets: design set and test set. The design set contained stimuli from three adult males and three adult females. The test set contained the stimuli from the remaining speakers. Rationale and criteria for this division can be found in Kong and Mullangi (2012). Classification of speech sounds, as described below, was performed for the design set to determine 1) the frequency region for identifying the speech signals that would undergo frequency-lowering processing, and 2) the number of acoustic features and the classification technique for dividing the high-frequency frication sounds into three groups. We then processed the stimuli in the test set using these parameters.

Conditional Lowering—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, the system used here only applied frequency lowering to consonants that were classified as nonsonorants (i.e., stop, fricative, and affricate consonants). The algorithm, using a conditional frequency-lowering rule, identified the consonants in the high-frequency region (>400 Hz) that contained aperiodic signals as determined by the

autocorrelation-based pitch-extraction algorithm in PRAAT (Boersma and Weenink, 2009). The pitch-extraction algorithm used a 40-ms window size and a time step of 10 ms. The pitch analysis results from two consecutive 10-ms segments of the speech signal were combined to form a 20-ms frame for the frequency-lowering process in the later stage. If the two 10-ms segments contained different periodicity detection results, the system classified the 20-ms signal as aperiodic. Classification accuracy for individual sounds was defined as greater than 50% of the speech segment that indicated the presence of periodicity for sonorant sounds, or absence of periodicity for nonsonorant sounds at high frequencies. This method detected voiced and voiceless nonsonorant consonants having aperiodic high-frequency energy with 95% accuracy for both the design and test sets. Sonorant sounds, including vowels, nasals, and semivowels, were identified as having periodic high-frequency energy with 98% accuracy for both the design and test sets. The overall accuracy was 97%.

Classification of Nonsonorant Sounds—Sounds classified as nonsonorant consonants were subjected to further analysis and the frequency-lowering processing. The purpose of this analysis stage was to separate high-frequency frication sounds into three groups. Spectral information related to fricative consonants was used for this classification. Fricatives were classified into three groups that differ in place of articulation (labial and interdental/f, θ, v, ð/vs. alveolar/s, z/vs. palatal/ʃ, ʒ/). We included the labial and interdental fricatives in the same group because previous reports did not find any combinations of static properties that could separate these two groups of fricatives with a high degree of accuracy (e.g., Onaka & Watson 2000; Ali et al. 2001; Fox & Nissen 2005). Unlike the earlier version of the system, which used spectral slope and spectral peak location for the classification of fricatives (Kong & Mullangi 2012), the present system used a classification technique known as Linear Discriminant Analysis (LDA) (Fisher 1936). Fricative consonants were bandpass filtered by 20 one-third-octave filters with center frequencies from 125 to 10,079 Hz. After examining the spectral characteristics of all fricative consonants in our sample (see Kong & Mullangi 2012, Fig. 2, for the spectral shape of the eight fricatives in our sample), as well as spectral characteristics of fricatives described in previous studies (Jongman et al. 2000; Onaka & Watson 2000; Ali et al. 2001), we decided to use eight features, which were the normalized outputs of eight filters with center frequencies from 2000 to 10,079 Hz. The stimuli consisted of all fricative consonants in the design set, plus fricative consonants recorded from one male adult and one female adult in our laboratory. Additionally, we added speech-shaped noise to the fricative consonants at +10 dB SNR and included these in the LDA training. Overall classification accuracy for the test set was 88% correct in quiet and 85% correct in noise.

Spectral Enhancement—We modified the system of Posen et al. by enhancing the difference of the spectral shapes of the transposed signals for the nonsonorant sounds based on the classification results for fricatives described above. Three spectral patterns of low-frequency transposed noise were used to signal the different groups. Nonsonorant sounds were first bandpass filtered into 20 one-third-octave analysis bands, and the output levels from each band were measured. We created two sets of low-frequency synthesis filters for the transposed signal. In one set, we used four synthesis filters (4-band condition) with center frequencies from 394 to 794 Hz, similar to those used in the system of Posen et al. The second set consisted of six synthesis filters (6-band condition) with center frequencies from 397 to 1260 Hz. The use of four or six synthesis filters in the system depended on the severity of hearing loss above 1000 Hz for each individual. For individuals who had severe-to-profound hearing loss above 1000 Hz (i.e., residual hearing cutoff frequency = 1000 Hz), four synthesis filters were used in the system. For those who had severe-to-profound hearing loss above 2000 Hz (i.e., residual hearing cutoff = 2000 Hz), six synthesis filters were used, because individuals with milder hearing loss at 1500 and 2000 Hz would be able to perceive

the transposed signal in the two highest frequency bands (1000 and 1260 Hz) and could receive a greater frequency-lowering benefit than with four synthesis bands.

1. For nonsonorant sounds classified as Group 1 (/f, θ, v, ð/), the outputs of eight one-third-octave filters from the original speech with center frequencies from 2000 to 10,079 Hz were first combined to form four analysis bands, similar to the method used in the system of Posen et al. The levels of these four analysis bands were then used to determine the levels of the low-frequency noise bands. For the 4-band condition, the high-frequency analysis filters and the low-frequency synthesis filters (397 – 794 Hz) were monotonically related in that the lowest analysis band determined the level of the lowest synthesis band, the second-lowest analysis band determined the level of the second-lowest synthesis band, and so on. For the 6-band condition, the four high-frequency analysis filters and the four middle synthesis filters (500 – 1000 Hz) were monotonically related. For both 4- and 6-band conditions, the output levels of the low-frequency synthesis bands were set 10 dB lower than those at the output of the corresponding high-frequency analysis bands. Also, for the 6-band condition, the noise levels in the lowest (397 Hz) and highest (1260 Hz) synthesis bands were set 40 dB lower than the output levels of the highest and lowest analysis bands, respectively.
2. For nonsonorant sounds classified as Group 2 (/ʃ, ʒ/), the noise level in the lowest synthesis band (397 Hz for both 4- and 6-band conditions) corresponded to the highest output level of the one-third-octave analysis filters from 1000 to 10,079 Hz in the original speech. The noise level in the second-lowest synthesis band (500 Hz for both 4- and 6-band conditions) was set 10 dB lower than that in the lowest synthesis band. The noise levels in the remaining bands were set 40 dB lower than that in the lowest synthesis band.
3. For nonsonorant sounds classified as Group 3 (/s, z/), the noise level in the highest synthesis band (794 Hz for the 4-band condition or 1260 Hz for the 6-band condition) corresponded to the highest output level of the one-third-octave analysis filters from 1000 to 10,079 Hz in the original speech. The noise level in the second-highest synthesis band (630 Hz for the 4-band condition or 1000 Hz for the 6-band condition) was set 10 dB lower than that in the highest synthesis band. The noise levels in the remaining bands were set 40 dB lower than that in the highest synthesis band.

We analyzed the noise level in each of the low-frequency synthesis bands for each group of fricatives for the 4-band (left panel) and 6-band (right panel) conditions (Fig. 2). As shown in Figure 2, the spectral contrast between the three groups of transposed noise is further enhanced in the system with six synthesis bands compared to the system with four bands. Preliminary data from some of our hearing-impaired subjects, who had residual hearing cutoff frequency at 2000 Hz, showed better fricative identification in the 6-band condition than in the 4-band condition.

Vocoder-Based Frequency Lowering—Similar to the system of Posen et al. (1993), for each group of nonsonorant sounds, the low-frequency narrow-band noise signals that carried the high-frequency speech information in the original signal were summed and then added to the original speech.

SPEECH RECOGNITION BY HEARING-IMPAIRED LISTENERS

The primary purpose of this study was to evaluate the effect of the frequency-lowering system on speech recognition by hearing-impaired listeners. Experiment 1 examined the benefit of the system over conventional amplification on place-of-articulation perception for

fricative consonants. Experiment 2 investigated the benefit of the system on consonant identification. We were interested in knowing the effect of severity of hearing loss above 1000 Hz on frequency-lowering benefit. Given that nonsonorant consonants have greater energy above 2000 Hz, we hypothesized that the system would provide benefit in situations when listeners receive little or no information above 2000 Hz.

Experiment 1: Identification of Fricative Consonants

This experiment was designed to evaluate the benefit of the processing for fricative place-of-articulation perception.

Methods

Subjects: Six hearing-impaired subjects (S1 – S6; two male), aged 14 to 58 years (mean = 26 years) with steeply sloping sensorineural hearing loss participated. Relevant information for each subject is shown in Table I. In this experiment, each subject was tested using one ear only. The test ear was chosen as the ear that had greater (> 10 dB) hearing loss at high frequencies. If the two ears had similar degrees of hearing loss, we tested the right ear, except for S3. Subject S3 was tested using the better ear due to the fact that she experienced a sudden drop of hearing in the worse ear at 1000 Hz a few months prior to the study. We were concerned that she might experience further deterioration of hearing in that ear during the course of the study.

Figure 3 (left) shows the unaided thresholds for the test ear for each subject. The dotted horizontal line at 70 dB HL indicates the cutoff level above which the loss is classified as severe. Thresholds above 70 dB HL indicate severe to profound hearing loss. Subjects S1 – S4 had a cutoff at 1000 Hz, whereas subjects S5 and S6 had a cutoff at 2000 Hz.

Stimuli: Eight fricative consonants /f, θ, s, ʃ, v, ð, z, ʒ/ in VCV context from the test set were used. All three repetitions spoken by each of the six speakers were used, resulting in a total of 432 tokens. All stimuli were scaled to have equal root-mean-squared (RMS) amplitude. Each stimulus was then processed by the frequency-lowering system with four or six synthesis bands. Four synthesis bands were used for subjects with residual hearing cutoff at 1000 Hz (S1 – S4). Six synthesis bands were used on remaining two subjects with residual hearing cutoff at 2000 Hz (S5 and S6). Stimuli were presented at an RMS level of 65 dB A. To compensate for the loss of audibility due to hearing loss, amplification was provided up to 8000 Hz for each subject. The NAL-RP prescription method (Byrne & Dillon 1986; Byrne et al. 1990) was used for calculating the gain. Since the NAL method does not prescribe gain for 8000 Hz, we used the formula for 6000 Hz for the gain calculation at 8000 Hz. The same gain settings were used for stimuli with and without frequency lowering. Due to the maximum output limit (about 110 dB SPL from 500 to 4000 Hz; and about 90 dB SPL above 4000 Hz) of the transducer (E-A-R TONE 3A insert earphones), the maximum gain applied to the stimuli was limited to 30 dB at each frequency. With this gain setting, all subjects received the prescribed NAL gain for frequencies at which the hearing loss was 75 dB or lower, except for subject S2 who received 30 dB of gain at 1000 Hz instead of the 34-dB gain prescribed by the NAL method. Also, the 30-dB gain provided for hearing loss between 80 and 95 dB HL was only slightly (1–5 dB) lower than the prescribed gain. With a 30-dB gain, some of the speech tokens had maximum amplitude that approached the output limit of our transducer.

Additionally, the unprocessed and frequency-lowered speech underwent LP filtering with an attenuation rate of 90 dB/octave. The LP cutoff frequency used for each subject was determined by their residual hearing cutoff frequency (i.e., 1000 Hz for S1 – S4; 2000 Hz for S5 and S6). Amplification was provided up to the cutoff frequency. We included the LP

condition in this experiment to determine if providing amplification at frequency regions with severe to profound hearing loss would affect speech recognition performance. Based on results from previous studies (Hogan & Turner 1998; Vickers et al. 2001; Baer et al. 2002), we hypothesized that LP filtering would either have no effect on speech intelligibility or would enhance speech perception compared to no filtering for individuals with severe-to-profound high-frequency hearing loss, and that the amount of frequency-lowering benefit would be similar between the wideband and LP filtered speech conditions.

Procedures: Subjects were tested using conventional amplification and frequency lowering with wideband and LP filtered speech. Conventional amplification was considered as the control condition. The wideband speech condition was tested before the LP filtered speech condition for half of the subjects, and the order was reversed for the remaining subjects. The order of testing of the two processing conditions within each speech condition was counterbalanced across subjects.

The experiment consisted of a familiarization phase and a test phase. One repetition spoken by four speakers (one adult male, one adult female, one boy, one girl) was used in the familiarization phase and the remaining two repetitions by each of the six speakers (two adult males, two adult females, one boy, one girl) were used in the test phase. For each processing and speech condition, each subject first practiced with a minimum of three blocks of 96 trials (eight fricatives $\times$ four speakers $\times$ three vowels) until performance appeared to level off, defined as performance within 3 percentage points in two consecutive practice blocks, up to a maximum of 5 practice blocks. After practice, each subject was tested with six blocks of 144 test trials (eight fricatives $\times$ six speakers $\times$ three vowels). Each one of the two repetitions by each speaker was tested for three blocks. 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 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.

At the beginning of the familiarization phase, subjects were instructed to report to the experimenter if the stimuli were perceived as uncomfortably loud. For subjects S1, S2, and S4, the level of the stimuli was reduced by 3–5 dB after they indicated that the amplified stimuli slightly exceeded their comfortable listening level.

Wideband white noise was added to the nontest ear for subject S1. In the nontest ear, subject S1 had normal hearing at 2000 Hz and below and moderate to moderately severe hearing loss above 2000 Hz. The spectrum level of the noise presented to mask “cross-hearing” was 35 dB. This level was determined by consideration of the worst (lowest) interaural attenuation level for insert earphones (i.e., 70 dB) and the maximum level of the amplified speech at each frequency band.

Results—Overall percent correct fricative identification scores and percent information transmission (Miller & Nicely 1955) for voicing and place features, across three vowel contexts, were calculated for the individual and group data for each processing and speech condition (see Fig. 4). Individual data were computed from confusion matrices combined across different runs for each subject, and the group data were computed by averaging the scores across subjects. Left panels represent performance with LP filtered speech and right panels with wideband speech.

For the individual data, standard errors and significant differences between two speech recognition scores were approximated on the basis of Bernoulli fluctuations. For example, at a performance level of 70 percent correct and 864 trials, the critical difference for $p < 0.05$ is about 4 percentage points. We found that all subjects showed a significant difference between the control and the frequency-lowering condition for both wideband and LP filtered speech.

For the group data, paired-t tests were performed on the overall percent correct and percent information transmission scores to compare performance between the control and frequency-lowering conditions. Fricative identification scores were higher for the frequency-lowering condition than for the control condition for both LP filtered and wideband speech [LP: $t(5) = 4.7$, $p < 0.01$; wideband: $t(5) = 6.5$, $p < 0.005$]. The improvement was 14 and 12 percentage points for the LP filtered and wideband speech conditions, respectively. Paired-t tests confirmed that the frequency-lowering benefit was due to enhanced place-of-articulation perception, while voicing perception remained essentially the same ($p > 0.05$) for both frequency-lowering and control conditions and for both LP filtered and wideband speech. The improvement for place perception was 25, and 22 percentage points for the LP filtered [$t(5) = 5.1$, $p < 0.005$] and wideband [$t(5) = 6.2$, $p < 0.005$] speech conditions, respectively.

Fricative identification performance was similar between LP filtered and wideband speech for five of the six subjects for the control condition, but not for S1, who showed 37 percentage points better performance with wideband speech than with the 1000 Hz LP filtered speech. This could be due to S1 having less severe hearing loss (75 dB HL loss) at high frequencies (4000 to 8000 Hz) than the rest of the subjects, who had profound hearing loss (> 90 dB HL) above 2000 Hz. Consistent with findings in previous studies (Hogan & Turner 1998; Vickers et al. 2001), this pattern of results further suggests that amplification at frequencies with profound hearing loss has no effect on speech intelligibility. More importantly, all subjects showed improved fricative consonant perception with frequency-lowered speech compared to conventional amplification, and, on average, the amount of frequency-lowering benefit was similar between the LP filtered and wideband speech conditions.

Experiment 2: Consonant Identification

The aim of this experiment was to investigate the effect of the frequency-lowering system for the perception of a larger group of consonants. While we have demonstrated the benefit of the system for fricative perception, it is important to evaluate the impact of frequency lowering on the perception of other classes of consonant sounds, including stops, nasal, and semivowels. We were also interested in determining candidacy for this system. The amount of frequency-lowering benefit for this system could be affected by (1) the degree of hearing loss below the residual hearing cutoff and (2) the residual hearing cutoff frequency above which the hearing loss is too severe to receive benefit from conventional amplification. To evaluate the effect of hearing loss on frequency-lowering benefit, the present study included additional subjects who had mild to moderate hearing loss in the frequency region of the transposed signals. We expected an effect of degree of low- and mid-frequency hearing loss on the amount of frequency-lowering benefit, due to the possible reduced ability to perceive spectral differences in the low-frequency transposed signals. We hypothesized that as the residual hearing cutoff frequency increases, the amount of low-frequency benefit decreases, because the listeners could receive more high-frequency speech information via conventional amplification. As demonstrated in experiment 1, LP filtering at the residual hearing cutoff produced similar fricative identification scores compared to wideband speech in the control condition for individuals with severe-to-profound high-frequency hearing loss (the target patient population for the system). The amount of frequency-lowering benefit was

also similar between LP filtered and wideband speech. In practice, high amounts of gain provided by a hearing aid in the severe to profound hearing loss region would cause distortion and acoustic feedback. Thus, the LP condition could be representative of what could be achieved with a conventional hearing aid (Robinson et al. 2007), especially for individuals with unaidable hearing loss at high frequencies, such as subjects S2—S6 in the present study.

Methods

Subjects: The same six subjects (S1 – S6) as in experiment 1 and three additional subjects (S7 – S9; two female) participated (see Table I). Subjects S1 – S6 were tested using the same ear as in experiment 1. Subject S7 was tested using the worse ear and subject S8, who had symmetrical hearing loss was tested using his right ear. Subject S9 had moderately-severe to severe hearing loss from 500 to 3000 Hz in the worse ear. However, the dynamic range for that ear was only about 20 dB. Given that we did not incorporate compression in the system, we decided to test her better ear, which had a greater dynamic range. Unlike the sloping hearing loss configuration for subjects S1 – S6, subject S7 – S9 had flat or mid-frequency (“cookie-bite”) hearing loss. These subjects had hearing loss in the frequency region of the transposed signals (400 – 1260 Hz). In the steeply sloping hearing loss group, some subjects had normal or near-normal thresholds at 500 and 750 Hz. The inclusion of the flat/mid-frequency hearing loss subjects in this experiment allowed us to examine the effect of degree of low- and mid-frequency hearing loss on frequency-lowering benefit. Speech stimuli presented to these three subjects were subjected to the same LP filtering procedure as for the steeply sloping hearing loss subjects.

Stimuli: Stimuli consisted of twenty-two consonants in/VCV/utterances with three vowels (/a, i, u/) in the test set. All three repetitions spoken by each of the six speakers were used, resulting in a total of 1,188 tokens. All stimuli were scaled to have equal RMS amplitude. Each stimulus was then processed by the frequency-lowering system. The unprocessed and frequency-lowered speech also underwent LP filtering at cutoff frequencies of 1000, 1500 and 2000 Hz (90 dB/octave rolloff). As for experiment 1, amplification based on the NAL-RP method was provided to each subject up to the cutoff frequency.

Procedures: There were two processing conditions (control and frequency lowering) and three filtering conditions (1000LP, 1500LP, and 2000LP filtered speech). Subjects S1 – S6, who had steeply sloping high-frequency hearing loss, were tested on the LP filtering condition with the LP cutoff frequency set to their residual hearing cutoff frequency (i.e., S1 – S4 at 1000 Hz; S5 – S6 at 2000 Hz). It is noted that hearing loss was at a profound level above the residual hearing cutoff frequency for these subjects, except for S1, who had a severe hearing loss above the residual hearing cutoff. Additionally, subject S5 was tested with 1000LP and 1500LP conditions. Subjects who had a flat or a cookie-bite configuration of hearing loss (S7 – S9) were tested using all three LP conditions. The 2000LP condition was tested first for subjects S5 and S7 – S9, followed by 1500LP, and then 1000LP conditions. For each filtering condition, subjects completed conditions with and without frequency lowering before they continued to the next filtering condition. For the 2000LP condition, frequency lowering was performed with six low-frequency synthesis filters. For the 1000LP and 1500LP conditions, frequency lowering was performed with four low-frequency synthesis filters. The order of testing of the two processing conditions was counterbalanced across subjects. As for experiment 1, masking noise was presented to the non-test ear of S1 and S7 to prevent cross-hearing.

There was a familiarization phase and a test phase. One repetition spoken by each of the four speakers was used in the familiarization phase and the remaining two repetitions by each of

the six speakers were used in the test phase. Different vowel contexts were tested in different sessions. This resulted in a total of 88 trials for the practice blocks (22 consonants $\times$ four speakers $\times$ one repetition $\times$ one vowel context) and a total of 264 trials for the test block (22 consonants $\times$ six speakers $\times$ two repetitions $\times$ one vowel context). This blocking was chosen to avoid very long practice and test blocks, to allow practice and test runs to be performed in one 3-hour session. For each processing and speech condition, each subject first practiced with a minimum of three blocks of practice trials for each vowel context until performance appeared to level off, defined as performance within 3 percentage points in two consecutive practice blocks, up to a maximum of 5 practice blocks. After practice, each subject was tested with five blocks of test trials for each vowel context. 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 (as before). Subjects were given a break every 50 trials. No feedback was provided during the test sessions.

Results—When stimuli were LP filtered at the residual hearing cutoff frequency for sloping profound high-frequency loss subjects (S2 – S6), frequency-lowering benefit was evident for overall consonant identification [$t(4) = 7.5$, $p < 0.005$], as well as for the perception of frication [$t(4) = 4.8$, $p < 0.01$] and place [$t(4) = 7.5$, $p < 0.005$].

Overall percent correct consonant identification (Fig. 5) and percent information transmission for six consonant features (Figs. 6–8), across vowel contexts, were calculated for the individual and group data for each processing and filtering condition. Individual data were computed from confusion matrices combined across different runs, and the group data were computed by averaging the scores across subjects. The method for standard error and significant difference approximations was the same as for experiment 1. Figure 5 shows the overall percent correct scores (top panels) and scores for various groups of consonants, semivowels-nasals (middle panels) and stops-fricatives-affricates (bottom panels), for each LP condition (from left to right). All subjects showed significant frequency-lowering benefit for overall consonant identification and for identification of stops-fricatives-affricates for all LP conditions.

For the group data, pairwise comparisons showed significantly better overall consonant identification with frequency lowering than for the control condition by about 9, 8, and 5 percentage points for the 1000LP [$t(7) = 9.98$, $p < 0.05$], 1500LP [$t(3) = 3.35$, $p < 0.05$], and 2000LP [$t(4) = 6.62$, $p < 0.05$] conditions, respectively. The improvement was mainly due to enhanced identification of stops-fricatives-affricates ($p < 0.05$) by about 6–11 percentage points, while the identification of semivowels-nasals was not significantly affected ($p > 0.05$) by frequency lowering for any LP condition. Information transmission analysis was performed for six features: voicing, stop, nasality, frication, affrication, and place. Results for the 1000LP, 1500LP, and 2000LP conditions are shown in Figs 6, 7, and 8, respectively. For each LP condition, the perception of voicing and nasality with frequency lowering was as good as that without frequency lowering ($p > 0.05$). The frequency-lowering system led to consistent significant improvements in the amount of information transmitted for the perception of frication and place for all LP conditions. For the perception of frication, the improvement was about 11, 5, and 4 percentage points for the 1000LP [$t(7) = 7.76$, $p < 0.0005$], 1500LP [$t(3) = 4.25$, $p < 0.05$], and 2000LP [$t(4) = 12.76$, $p < 0.0005$] conditions, respectively. For the perception of place, the improvement was 10, 11, and 9 percentage points for the 1000LP [$t(7) = 8.30$, $p < 0.0001$], 1500LP [$t(3) = 3.88$, $p < 0.05$], and 2000LP [$t(4) = 9.59$, $p < 0.001$] conditions, respectively. A frequency-lowering benefit was found for the perception of stop [$t(7) = 3.13$, $p < 0.05$] and affrication [$t(7) = 2.43$, $p < 0.05$] for the 1000LP condition, but not for the other LP conditions.

The effect of low- and mid-frequency hearing loss on frequency-lowering benefit was examined using correlation of the pure-tone average (PTA) and the amount of frequency-lowering benefit for place perception. For the 4-band condition, PTA was calculated by averaging the thresholds at 250, 500, and 750 Hz. For the 6-band condition, thresholds from 250 to 1500 Hz were included in the PTA. The PTA ranged from 13 dB HL to 45 dB HL and the amount of frequency-lowering benefit ranged from 6 points to 16 points. There was no significant correlation between PTA and the amount of benefit for place perception ($r = -0.045$, $p > 0.05$).

GENERAL DISCUSSION

Effect of the Frequency-Lowering System on Consonant Perception

Our study demonstrated the benefit of the frequency-lowering system for speech perception in quiet for hearing-impaired listeners with severe-to-profound high-frequency hearing loss. Improvement of the perception of frication and affrication has been reported previously for some hearing-impaired listeners with other frequency-lowering systems (e.g., Kuk et al. 2009; Robinson et al. 2007; Simpson et al. 2005). The present study showed a significant frequency-lowering benefit for place-of-articulation perception for listeners with steeply sloping high-frequency hearing loss, in addition to the benefit for frication and affrication perception. Some studies only reported overall percent correct consonant identification scores, but did not analyze information transmission. For example, Turner and Hurtig (1999) reported that their proportional frequency compression system improved consonant identification for some of their subjects and the amount of benefit was comparable to that found in the present study. However, it is not known if there was any improvement in place-of-articulation perception.

The system used here took advantage of two features of the system of Posen et al.: (1) conditional frequency lowering and (2) reduction of masking from the transposed signal. The system used here employed a conditional lowering rule that separates sonorant and nonsonorant sounds based on detection of periodicity at frequencies above 400 Hz. The perception results showed that the frequency-lowering system did not worsen the identification of semivowels and nasals, presumably because of the conditional frequency lowering. Posen et al. transposed the high-frequency speech energy (1000–5000 Hz) to the low-frequency region (400–800 Hz), where there is little energy in the original signal for fricative and affricate consonants (see Fig. 2 in Kong & Mullangi 2012). This minimized the masking effect due to superimposition. The other advantage is that hearing loss is less severe or absent at the low and mid frequencies for individuals with sloping high-frequency hearing loss. Transposition of the high-frequency speech sounds to the region with relatively good hearing could potentially lead to greater speech recognition benefit than the nonlinear frequency compression method, for which compression takes place mainly at high frequencies (Simpson et al. 2005). This is because spectral resolution in the normal or near-normal hearing region would be better than that in the region with substantial hearing loss (e.g., Glasberg & Moore 1986).

One disadvantage of Posen et al.'s frequency-lowering approach is that it involves frequency compression. This could have a negative impact on speech recognition, especially when listeners use primarily spectral cues to perceive the place-of-articulation distinction for fricative consonants (Harris 1958). The frequency-lowering system used here alleviates the detrimental effect of frequency compression by enhancing spectral cues for fricatives differing in place of articulation. We attribute the improved place perception to the enhanced spectral cues for fricative consonants in the system.

Effect of Hearing Loss on Frequency-Lowering Benefit

In experiment 2, we showed that the amount of frequency-lowering benefit increased as the amount of high-frequency speech information received by the listeners decreased. There was a greater benefit for the 1000LP condition than for the 1500LP and 2000LP conditions. For the 1000LP condition, benefit was observed for the perception of stop, frication, affrication, and place, but benefit was found only for frication and place for the higher cutoff LP conditions. The amount of benefit for the perception of frication was also greater for the 1000LP condition than for the 1500LP and 2000LP conditions. However, the presence of mild to moderate hearing loss at the low- and mid-frequency regions did not seem to affect the frequency-lowering benefit.

In the present study, the majority of the participants had severe-to-profound high-frequency hearing loss. A recent study by Wolfe et al. (2010) showed a significant frequency-lowering benefit for children who had only moderate to moderately-severe hearing loss at and above 2000 Hz when tested with Phonak's nonlinear frequency compression system. For phoneme identification, frequency compression improved/s/and/d/identification, but there was no significant effect for other consonant sounds (/f, k, ʃ, t/). It is not clear if the system used here can enhance speech recognition for this patient population or for individuals who have residual hearing cutoff frequencies above 2000 Hz.

Future Directions

Future work should consider (1) exploring the effect of varying other parameters of the system, including the region and the bandwidth of the low-frequency synthesis bands for individuals with different degrees and slopes of hearing loss (e.g., Robinson et al. 2007; Fullgrabe et al. 2010), and (2) improving classification accuracy for speech sounds. The success of the system relies on the accuracy of classification of consonant sounds. We are currently investigating methods that employ a combination of acoustic features for sonorant vs. nonsonorant sound classification in quiet and in noise using a shorter time window. We will continue to investigate methods for fricative classification using speech samples from more speakers, tests in more natural conditions (e.g., continuous speech, multi-talker babble noise), and at more challenging SNRs. Finally, we will investigate methods that could reliably separate the nonsonorant sounds into three groups: stops vs. fricatives vs. affricates, and then enhance the acoustic cues in the frequency-lowered signal for each group.

Acknowledgments

We are grateful to all the subjects for their participation in this study. We would like to thank Professor Brian Moore, Dr. Monita Chatterjee, and one anonymous reviewer for their helpful comments. We also thank Melissa Early and Jaclyn Dollard for their assistance with data collection. This work was supported by NIH/NIDCD (R03 DC009684-03 and ARRA supplement) to the first author.

References

- Ali AM, Van der Spiegel J, Mueller P. Acoustic-phonetic features for the automatic classification of fricatives. *J Acoust Soc Am*. 2001; 109:2217–2235. [PubMed: 11386573]
- Auriemma J, Kuk F, Lau C, et al. Effect of linear frequency transposition on speech recognition and production of school-age children. *J Am Acad Audiol*. 2009; 20:289–305. [PubMed: 19585959]
- Baer T, Moore BCJ, 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 Am*. 2002; 112:1133–1144. [PubMed: 12243160]
- 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

- Braida LD, Durlach NI, Lippmann RP, et al. Hearing aids: a review of past research on linear amplification, amplitude compression, and frequency lowering. *ASHA Monograph*. 1979; (19)
- Byrne D, Dillon H. The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear Hear*. 1986; 7:257–265. [PubMed: 3743918]
- Byrne D, Parkinson A, Newall P. Hearing aid gain and frequency response requirements for the severely/profoundly hearing impaired. *Ear Hear*. 1990; 11:40–49. [PubMed: 2307302]
- Fisher RA. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*. 1936; 7:179–188.
- Fox RA, Nissen SL. Sex-related acoustic changes in voiceless English fricatives. *J Speech Lang Hear Res*. 2005; 48:753–765. [PubMed: 16378471]
- Fullgrabe 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]
- Glasberg BR, Moore BCJ. Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *J Acoust Soc Am*. 1986; 79:1020–1033. [PubMed: 3700857]
- Glista D, Scollie S, Bagatto M, et al. 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.
- Hogan CA, Turner CW. High-frequency audibility: Benefits for hearing-impaired listeners. *J Acoust Soc Am*. 1998; 104:432–441. [PubMed: 9670535]
- Jongman A, Wayland R, Wong S. Acoustic characteristics of English fricatives. *J Acoust Soc Am*. 2000; 108:1252–1263. [PubMed: 11008825]
- Kong YY, Mullangi A. On the development of a frequency-lowering system that enhances place-of-articulation perception. *Speech Comm*. 2012; 54:147–160.
- Kuk F, Keenan D, Korhonen P, et al. Efficacy of linear frequency transposition on consonant identification in quiet and in noise. *J Am Acad Audiol*. 2009; 20:465–479. [PubMed: 19764167]
- Lippmann RP. Perception of frequency lowered speech. *J Acoust Soc Am*. 1980; 67:S78.
- Moore BCJ. Dead regions in the cochlea: conceptual foundations, diagnosis and clinical applications. *Ear Hear*. 2004; 25:98–116. [PubMed: 15064655]
- Miller GA, Nicely PE. An analysis of perceptual confusion among some English consonants. *J Acoust Soc Am*. 1955; 27:338–352.
- Onaka A, Watson CI. Acoustic comparison of child and adult fricatives. 8th Aust Int Conf Speech Sci & Tech. 2000:134–139.
- Posen MP, Reed CM, Braida LD. Intelligibility of frequency-lowered speech produced by a channel vocoder. *J Rehabil Res Dev*. 1993; 30:26–38. [PubMed: 8263827]
- Reed CM, Hicks BL, Braida LD, et al. Discrimination of speech processed by low-pass filtering and pitch-invariant frequency lowering. *J Acoust Soc Am*. 1983; 74:409–419. [PubMed: 6619418]
- 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, et al. Consonant recordings for speech testing. *J Acoust Soc Am*. 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]
- Turner CW, Hurtig RR. Proportional frequency compression of speech for listeners with sensorineural hearing loss. *J Acoust Soc Am*. 1999; 106:877–886. [PubMed: 10462793]
- Velmans M. The design of speech recoding devices for the deaf. *Br J Audiol*. 1974; 8:1–5.

- Vickers DA, Moore BC, Baer T. Effects of lowpass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies. *J Acoust Soc Am*. 2001; 110:1164–1175. [PubMed: 11519583]
- Wolfe J, John A, Schafer E, et al. 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]

Figure 1. Block diagram of the frequency-lowering system and the consonant classification algorithms. LDA stands for linear discriminant analysis.

Figure 2. The relative noise level of each of the low-frequency synthesis bands for three groups of fricative consonants for the four low-frequency synthesis bands (left) and six low-frequency synthesis bands (right) conditions.

Figure 3.

Unaided thresholds for the test ears of the six subjects in the steeply sloping hearing loss group (left) and the three subjects in the flat/mid-frequency hearing loss group (right).

Figure 4.

Percent correct fricative identification scores and percentage of information transmission of the voicing and place features. Error bars represent ± 1 standard error. Significant differences between processing conditions are indicated by *. Left panels represent the LP filtered speech condition, whereas the right panels represent the wideband speech condition.

Figure 5.

Percent correct scores for the overall set of 22 consonants and for two subgroups (semivow-nasal and stop-fric-affric) of consonants. Error bars represent ± 1 standard error. Significant differences between processing conditions are indicated by *. Each column represents a different speech condition.

Figure 6. Percentage of information transmission for individual and group data for each of the six features in the 1000LP condition. Error bars represent ± 1 standard error. Significant differences between processing conditions are indicated by *.

Figure 7.
As Fig. 6 but for the 1500LP condition.

Figure 8.
As Fig. 6 but for the 2000LP condition.

Table 1

Subject demographic information

Subject	Age	Gender	Ear tested	Probable Etiology	Age onset [*]	Hearing aid? ^{**}	Sloping HL?
S1	20	F	L	Unknown	9	No	Yes
S2	23	F	R	Unknown	birth	Yes	Yes
S3	18	F	L	Genetic	6	Yes	Yes
S4	14	F	R	Genetic	5	Yes	Yes
S5	58	M	L	Noise-induced	20s	Yes	Yes
S6	22	M	L	Genetic	birth	No	Yes
S7	22	F	L	Pendred Syndrome	birth	Yes	No
S8	57	M	R	Genetic	20s	No	No
S9	21	F	R	Genetic	10	Yes	No

^{*} Age hearing loss was diagnosed^{**} Hearing aid use in the test ear