Time efficient contrast-enhancing frequency shaping and multiband compression in hearing aids

Shahab U. Ansari (sansari@grads.ece.mcmaster.ca)**, Harjeet S. Bajaj, Kamran Mustafa, and Ian C. Bruce**
Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada

1. Introduction

Contrast-enhancing frequency shaping (CEFS) proposed by Miller et. al. in 1999 resolves the second formant (F2) at the auditory nerve fiber (AN) of an impaired ear without amplifying the harmonics between formants F1 and F2 (Fig 1). Bruce et. al. in 2004 has shown that CEFS can be employed with multiband compression scheme when used in series without counteracting one another. In this poster we present a combination of CEFS amplification and multiband compression (M-CEFS) in a single frequency-domain filterbank implementation, thus reducing the computational complexity and the signal delay. Also, the M-CEFS scheme improves neural representation of F2 and F3. The new scheme is tested on the models of normal and impaired ears (Bruce et. al., 2003) and compared with linear amplification and CEFS without compression.

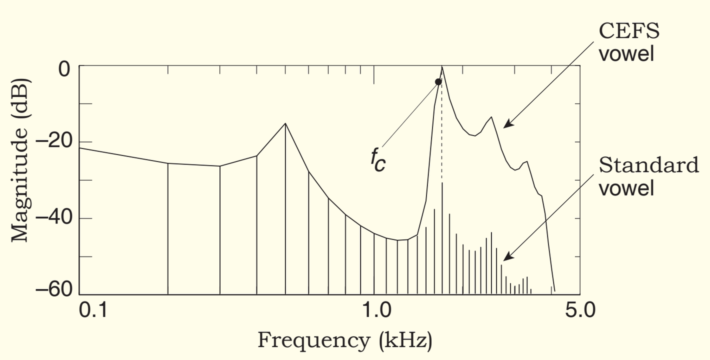


Figure 1. Power spectra of the standard and CEFS versions of the vowel /ε/. The line spectrum shows the unprocessed vowel's spectral shape and the solid line shows the CEFS-modified spectral envelope. The CEFS vowel was obtained by high-pass filtering the standard vowel with a cutofff frequency f_c, which is 50 Hz below the second formant frequency (indicated by the vertical dashed line). Reprinted from Bruce *et al* (2003) with permission from the Acoustical Society of America © (2003).

2. Method

The multiband compression algorithm was implemented in the frequency domain using the FFT overlap-add method. The input signal, sampled at 16 kHz, was divided into small frames using Hanning window of length 128 and zero-padded to avoid time aliasing. The energy in each frame was then calculated in different frequency band by using a filterbank of 15 bandpass filters with center frequencies starting at 250 and 1/3-octave apart. The gain-frequency response is adjusted as a function of the energy in each band to give a compression ratio of 2:1 above 40 dB (kneepoint) at the center frequency of each filter. The amplification gain-frequency response was then realized by interpolation and extrapolation across the frequency bands. The gain in each frame was further emphasized at F2 and F3 by using a time varying highpass filter (M-CEFS) (Fig. 3), whose cutoff frequencies were determined by formant tracker (Mustafa and Bruce, 2004) (see Fig. 2).

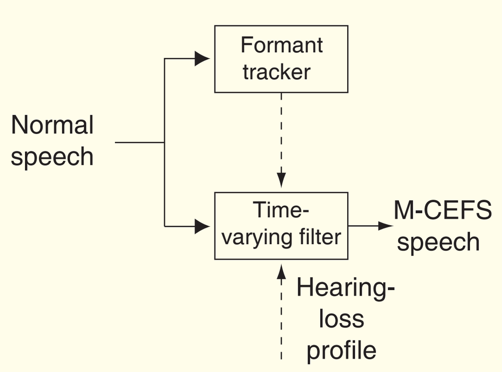


Figure 2. Schema of M-CEFS scheme

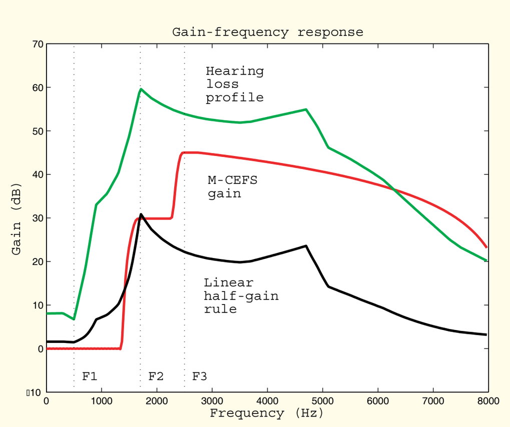
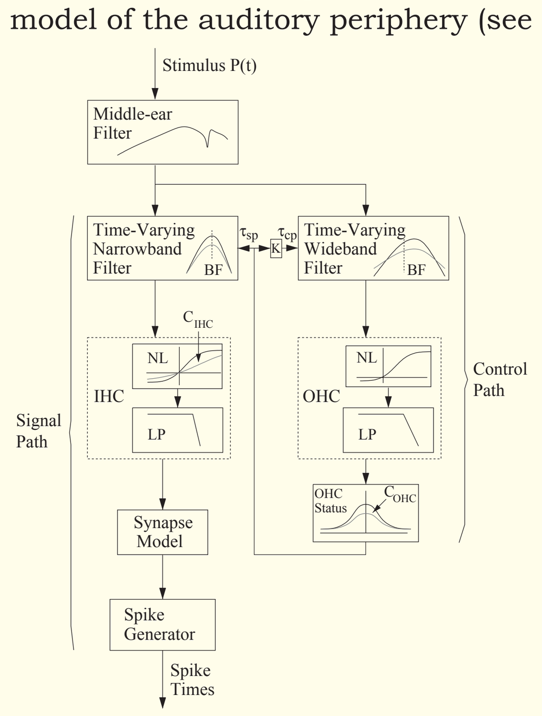


Figure 3. Gain-frequency response

The M-CEFS algorithm was tested for two types of speech: a synthesized vowel /ε/ and a synthesized sentence "*Five women played basket ball*". The model responses to the speech are evaluated for four conditions: (1) unmodified speech presented to a normal ear; (2) linear-amplified ("half-gain rule") speech presented to an impaired ear; (3) CEFS-modified speech presented to an impaired ear; and (4) M-CEFS speech presented to an impaired ear. The neural representation of the speech in normal and the impaired ear was achieved by the Bruce et. al. (2003) model of the auditory periphery (see Fig, 4).



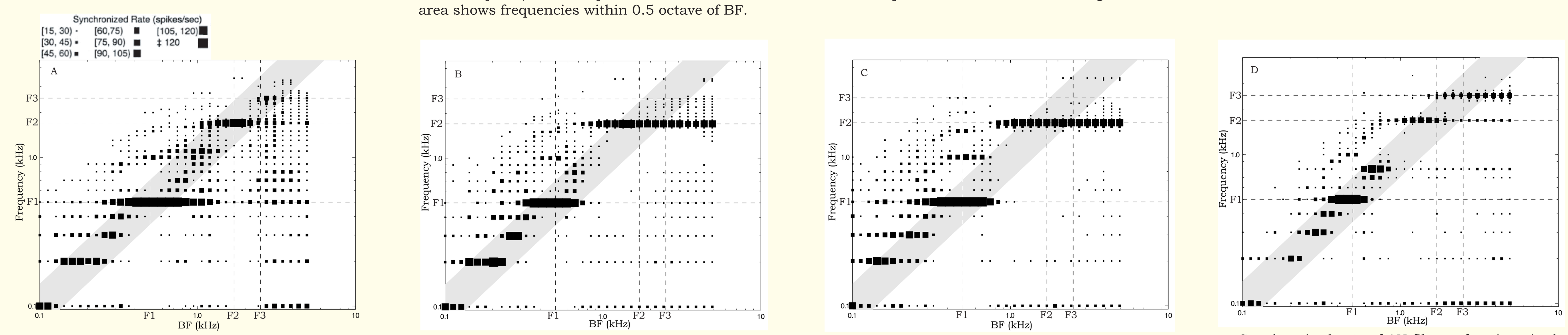
In case of the vowel /ε/, the discharge rate (spikes/s) was plotted for each fiber's best frequency against the signal frequency (Fig. 6). For the test sentence a formant power ratios plot was used to show which AN fibers synchronize to F1, F2 and F3 (Fig. 6). In case of speech, a neurogram plot was also used to show the discharge rate at each AN fiber in response to the speech.

Figure 4. AN model from Bruce *et al* (2003). Abbreviations: outer hair cell (OHC); low-pass (LP) filter; static nonlinearity (NL); inner hair cell (IHC); best frequency (BF); C_{IHC} and C_{OHC} are scaling constants that control IHC and OHC status, respectively. Reprinted with permission from the Acoustical Society of America © (2003).

3. Results

The results for the synthesized vowel are compared using box plots in Fig. 5. For the synthesized test sentence, the spectrogram, the formant power ratios, and the neurogram plotted in Fig. 6 are used for comparison.

Figure 5. Population distribution of synchronized rates in response to a vowel /ε/. The horizontal axis is fiber BF. Each column of squares shows the synchronized rate for a model fiber with a given BF. The size of each square represents the synchronized rate, as defined in the legend above panel A. Response components fewer than 15 spikes/s are not plotted. F1, F2 and F3 are the formant frequencies of the vowel. The diagonal shaded area shows frequencies within 0.5 octave of BF.



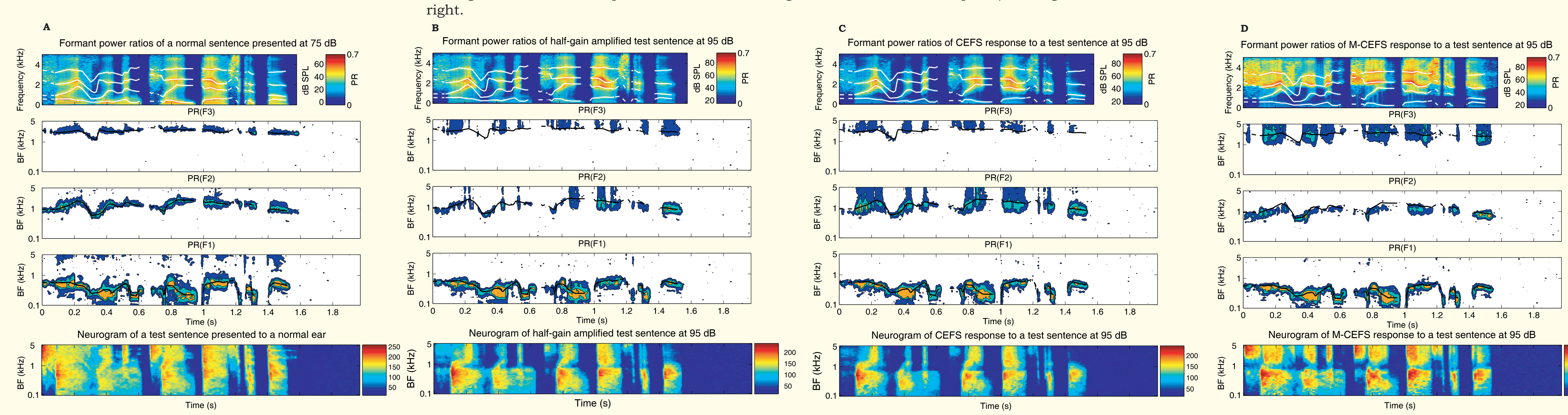
Synchronized rate of AN fibers of a normal ear responding to a vowel /ε/ presented at 65 dB SPL. The synchrony to the formant frequencies occurs at the appropriate (tonotopic) BF regions.

Synchronized rate of AN fibers of an impaired ear responding to a linearly amplified (half-gain rule) vowel /ε/ presented at 95 dB SPL. Synchrony to F1 is localized to the appropriate BF region, but there is a upwards spread of synchrony to F2 and a loss of synchrony to F3.

Synchronized rate of AN fibers of an impaired ear responding to a CEFS-modified vowel /ε/ presented at 95 dB SPL. Synchrony to F1 is localized, but there is still an upwards spread to F2. The contrast between the formant frequencies is improved.

Synchronized rate of AN fibers of an impaired ear responding to a vowel /ε/ modified by M-CEFS and presented at 95 dB SPL. The F1 and F2 are localized to their respective BF regions. Synchrony to F3 is restored at the appropriate BF region, but there is some upwards spread to F3. The contrast between the formant frequencies has been improved significantly.

Figure 6. Spectrogram, formant power ratios and the neurogram in response to a sentence "*Five women played basketball*". The top panel shows the spectrogram of the test sentence. The white lines show the trajectories of the four formant frequencies. Solid lines indicate voiced speech, dashed lines indicate unvoiced speech, and silence is indicated by no lines. The color scale to the right of the spectrogram indicates the sound level in the test sentence in dB SPL. The next three white panels show the formant power ratio (PR) plots for F1, F2, and F3 plotted with BFs along the y-axis against time along the x-axis. The black line in each panel shows the respective formant frequency. The PR plots indicate the strength of synchrony to a particular format. The neurogram in the bottom panel shows the discharge rate of each fiber in spikes/s using the color scale to the right.



Response of the normal ear to the test sentence presented at 75 dB SPL. The spectrogram shows strong formant trajectories during voiced speech segments. The formant power ratios show that the fibers with BFs near the formants are responding heavily. From the neurogram it can be seen that the discharge rate of the fibers is also very prominent at each formant frequencies.

Response of the impaired ear to the test sentence processed by a linear amplification scheme (half-gain rule) presented at 95 dB SPL. The spectrogram shows strong F2 and F3 trajectories during voiced speech segments. The power ratios plot show that the fibers with BFs near the F1 are responding heavily. Synchrony capture at F2 and F3 is deteriorated and there is upwards spread of synchrony to F2 and F3. The neurogram show that the discharge rate of the fibers at higher BFs is decreased noticeably.

Response of the CEFS-modified test sentence presented at 95 dB SPL. The power ratio plots show the synchrony capture at F2 is emphasized, but there is still upwards spread of synchrony to F2.

M-CEFS response of the impaired ear to the test sentence presented at 95 dB SPL. The spectrogram shows enhanced F2 and F3 trajectories during voiced speech segments. The formant power ratio plots show that fibers with BFs near the formants are responding heavily. There is an improvement in synchrony capture of BFs at F3, but there is upwards spread of synchrony to F3. It can be seen from neurogram plot that the discharge rate of the fibers at higher BFs is restored to near-normal.

Comparison

Panel B of Fig. 5 show that linear amplification (half-gain rule) of the vowel before presenting it to an impaired ear helps localize synchrony to F1 and improves the discharge rate of AN fibers with BFs near F2. However, the upwards spread of synchrony to F2 hinders the AN response to F3. The CEFS-modified vowel, shown in Panel C of Fig. 5, has increased contrast between the formants, but does not help in restoring the AN response to F3. In panel D of Fig. 5, the M-CEFS algorithm has successfully restored the AN response to F3 in addition to localizing F1 and enhancing contrast between the formants. Some upwards spread of synchrony to F3 remains.

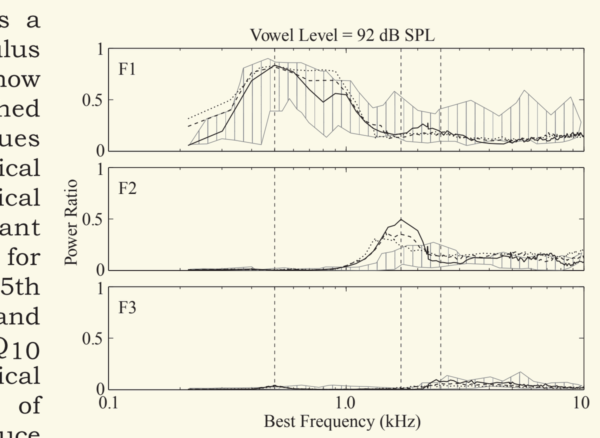
The linear amplification scheme (half-gain rule) applies the most gain to higher formants as can be seen in the spectrogram of panel B of Fig. 6 in response to the test sentence. This helps restrict the upwards spread of synchrony to F1, and improves the synchronization of AN fibers with BFs close to F2 and F3 as shown in the PR plots. However, linear amplification could not do much in curtailing the upwards spread of synchrony to F2 and F3. Panels C of Fig. 6 shows that the CEFS response, in addition to localizing F1, exhibits an increase in synchronization at F2, but there is a upwards spread of synchrony to F2. The M-CEFS response has obviously localized F2 in addition to localizing F1 and has improved synchronization at F3. In M-CEFS, the multiband compression has not altered the contrast between formants. Some upwards spread to F3 is still observed.

4. Discussion

In this poster we have presented M-CEFS, a hearing-aid amplification scheme to compensate for sensorineural hearing loss. The reduced dynamic range of the impaired ear is corrected by multiband compression in M-CEFS. Contrast enhancement of the formants in M-CEFS compensates for the elevated and broadened tuning curves of AN fibers and is implemented by using a time-varying filter. The cutoff frequencies of the time-varying filter are determined by the first three formant frequencies of the speech signal. A formant tracker is used in parallel with the M-CEFS filter to track the formants of the speech in real time. The implementation of M-CEFS has assured the reduction in group delay and the computational complexity by incorporating CEFS into the same FFT-based filterbank used for the compression algorithm. The group delay of M-CEFS algorithm is about 16 ms on average (a 10 ms improvement as compared to series implementation of CEFS and multiband compression), which is still larger than desired in practical hearing aids (typical 10 ms). So, we will investigate further reduction of the time delay of M-CEFS without affecting the current performance. Another improvement, which is achieved with M-CEFS is the response of AN fibers to F3.

In the presentation of the results, the model underestimates the loss of F2 synchrony, which hinders the comparison of amplification schemes (see Fig. 7). Finally, the algorithm still requires human testing to determine its actual performance.

Figure 7. Model predictions of impaired power ratios for F1, F2, and F3 as a function of impaired BF for a stimulus intensity of 92 dB SPL. Thick lines show model predictions and grey hatched areas indicate the range of values observed in the impaired physiological data of Miller et al (1997). Vertical dashed lines show the formant frequencies. Predictions are shown for model Q_{10} values that are at the 75th (solid lines), 50th (dashed lines), and 25th (dotted lines) percentiles of Q_{10} values for the impaired physiological data, i.e., for the three functions of C_{OHC} and with IHC impairment. Bruce et al, (2003). J. Acoust. Soc. Am.



References

Bruce, I. C. (2004). Physiological assessment of contrast-enhancing frequency shaping and multiband compression in hearing aids. Physiol. Meas. 25, 945-956

Bruce, I. C., Sachs, M. B., and Young, E. D. (2003). An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses. J. Acoust. Soc. Am. 113, 369-388

Dillon, H. (2001). Hearing Aids. New York: Thieme Medical Publishers.

Miller, R. L., Calhoun, B. M., Young, E. D. (1999). Contrast enhancement improves the representation of /eh/-like vowels in the hearing-impaired auditory nerve. J. Acoust. Soc. Am. 106, 2693-2708.

Mustafa, K., and Bruce, I. C. (2004). Robust formant tracking for continuous speech with speaker variability. Submitted to IEEE Trans. Speech Audio Processing.

Sachs, M. B., Bruce, I. C., Miller, R. L., and Young, E. D. (2002). Biological basis of hearing-aid design. Ann. Biomed. Eng. 30, 157-168.

Schilling, J. R., Miller, R. L., Sachs, M. B., and Young, E. D. (1998). Frequency-shaped amplification changes the neural representation of speech with noise-induced hearing loss. Hearing Research 117, 57-70

Acknowledgement

The authors would like to thank Brent Edwards for supplying the code for multiband-compression. This work was supported by NSERC Discovery Grant 261736 and the Barber-Gennum Chair Endowment.