

Physiological assessment of contrast-enhancing frequency shaping and multiband compression in hearing aids

Ian C Bruce

Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario L8S 4K1, Canada

E-mail: ibruce@ieee.org

Received 13 February 2004, accepted for publication 8 April 2004

Published 22 July 2004

Online at stacks.iop.org/PM/25/945

doi:10.1088/0967-3334/25/4/013

Abstract

Spectral enhancement is now being used in many hearing aids in an attempt to compensate for broadened cochlear filtering. However, spectral enhancement may be counteracted by multiband-compression algorithms designed to compensate for the reduced dynamic range of the impaired cochlea. An alternative scheme for spectral enhancement, contrast-enhancing frequency shaping (CEFS), has been proposed, which results in an improved neural representation of the first and second formants of voiced speech segments in the impaired ear. In this paper, models of the normal and impaired ear are used to assess the compatibility of CEFS with multiband compression. Model auditory nerve responses were assessed under four conditions: (1) unmodified speech presented to a normal ear; (2) amplified, unshaped speech presented to an impaired ear; (3) CEFS speech presented to an impaired ear; and (4) CEFS+multiband-compression speech presented to an impaired ear. The results show that multiband compression does not reduce the benefits of CEFS, and in some cases multiband compression assists in preventing distortion of the neural representation of formants. These results indicate that the combination of contrast-enhancing frequency shaping and multiband compression should lead to improved perception of voiced speech segments in hearing aid users.

Keywords: spectral enhancement, multiband compression, hearing aids, auditory nerve model, hearing loss, formant representation

1. Introduction

Auditory nerve (AN) fibers are known to exhibit elevated and broadened threshold tuning curves in an ear suffering from cochlear damage (e.g., Kiang *et al* 1976, Robertson 1982, Liberman and Dodds 1984), and a damaged cochlea has a reduced compressive nonlinearity leading to a smaller dynamic range (e.g., Robles and Ruggero 2001). Corresponding elevations in perceptual thresholds, broadening of perceptual tuning curves and reduced dynamic ranges for loudness have been observed in individuals suffering from cochlear hearing loss (Moore 1995). In order to compensate for elevated thresholds, conventional hearing aid amplification schemes utilize gain-frequency responses based on the profile of perceptual threshold elevation versus frequency, referred to as the audiogram, and multiband compression is used to compensate for reduced cochlear compression and perceptual dynamic range as a function of frequency (Dillon 2001).

Recently, strategies to compensate for impaired frequency resolution, i.e., broadened physiological and perceptual tuning curves, have been investigated (Summerfield *et al* 1985, Simpson *et al* 1990, Stone and Moore 1992, Baer *et al* 1993, Lyzenga *et al* 2002). These algorithms apply some form of spectral expansion, i.e., increasing the height of spectral peaks (formants) and/or deepening spectral valleys, to produce greater spectral contrast, in an attempt to offset broadened frequency resolution in the cochlea. However, Franck *et al* (1999) have shown that multiband compression tends to flatten the speech spectrum and consequently counteracts any benefits of spectral expansion schemes.

Miller *et al* (1999) have proposed an alternative spectral enhancement scheme, contrast-enhancing frequency shaping (CEFS). Based on physiological data from cat AN, CEFS attempts to adjust the relative amplitudes of spectral peaks without modifying or distorting spectral valleys, in order to produce a better neural representation of the formants. Figure 1 illustrates how CEFS is applied to the vowel / ϵ / in order to restore a more normal neural representation of the first ($F1$) and second ($F2$) formants in an ear that has cochlear damage in the region of the $F2$ frequency.

In this paper, the effects of combining CEFS amplification with multiband compression are investigated using a computational model of the auditory periphery (Bruce *et al* 2003). First, a prototype scheme for adding multiband compression to CEFS is described, and the CEFS algorithm of Miller *et al* (1999) is generalized to work for running speech, in which the formants vary over time. Second, simulated auditory nerve spike trains are generated in response to a test sentence, and the neural representation of formant trajectories is quantified by the formant power ratios (PRs) for the synchronized rates of a population of auditory nerve fibers (Miller *et al* 1997). It is shown that this neural representation is not degraded by the addition of multiband compression to CEFS, and the implications of these results for hearing aid amplification schemes are discussed.

2. Method

2.1. Test sentence

The test speech signal used in this paper was the synthesized sentence 'Five women played basketball' (courtesy of R McGowan of Sensimetrics Corp, Somerville, MA). This test sentence is phonetically rich, giving a variety of formant trajectories. A synthesized sentence was used so that the exact formant trajectories were known, allowing quantitative analysis of the neural representation of the formants. In the results section, presentation levels for the test

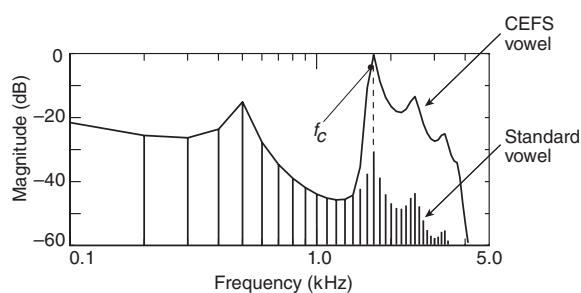


Figure 1. Power spectra of the standard and CEFS versions of the vowel / ϵ /. The line spectrum shows the unprocessed vowel's spectral shape and the solid line shows the CEFS-modified spectral envelope. The CEFS vowel was obtained by high-pass filtering the standard vowel with a cutoff frequency f_c , 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).

sentence are given in dB relative to 20 μ Pa, commonly referred to as dB sound pressure level (SPL), for the highest-amplitude phoneme of the sentence, the 'a' in 'basket'.

2.2. Amplification schemes

In the prototype amplification scheme evaluated in this paper, the test speech signal was first passed through a multiband-compression stage, utilizing an FFT-based filterbank and compressor, and then through a CEFS amplification stage, utilizing a time-varying FIR filter.

2.2.1. Multiband compression algorithm. The multiband-compression algorithm used in this paper utilizes a filterbank consisting of 15 filters spaced at $\frac{1}{3}$ -octave, starting at 250 Hz. Filter bandwidths were approximately two equivalent rectangular bandwidths (ERBs). A sampling frequency of 16 kHz was used. The filterbank was implemented in the frequency domain using the overlap-and-add method, utilizing an 8 ms (128 point) sliding Hann window with 50% overlap. Each windowed speech segment was zero padded to give a 32 ms (512 point) sequence, and the 512 point fast Fourier transform (FFT) of each sequence was taken. The FFT spectra of the windowed sequences were multiplied by the specified filterbank frequency responses, and the inverse FFT of each resulting spectrum was taken. The resulting time-domain sequences were overlapped and added according to which original windowed segment the sequence came from.

Compression was applied independently in each frequency band of the filterbank by measuring the signal power in each band for each windowed segment. For input powers below 40 dB SPL in a frequency band, the gain of that filter remained unchanged, giving a linear filter in that band. For input powers above 40 dB SPL, the gain in that frequency band was reduced to give a steady-state compression ratio of 2:1 in that band. The 40 dB SPL threshold is referred to as the compression 'knee-point.' The gain in each frequency band was adjusted dynamically to give a near-instantaneous attack and a release time of approximately 60 ms.

2.2.2. CEFS amplification. The CEFS amplification scheme of Miller *et al* (1999) was generalized to running speech as illustrated in figure 2. CEFS amplification for running speech was achieved by tracking $F2$ over time and applying a time-varying high-pass filter as a function of (i) the $F2$ frequency estimate at a particular time and (ii) the degree of hearing loss at that $F2$ frequency.

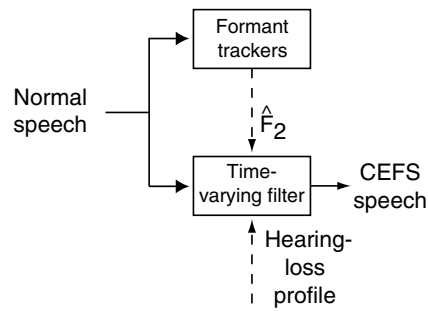


Figure 2. Generalization of the CEFS amplification scheme of Miller *et al* (1999) to running speech.

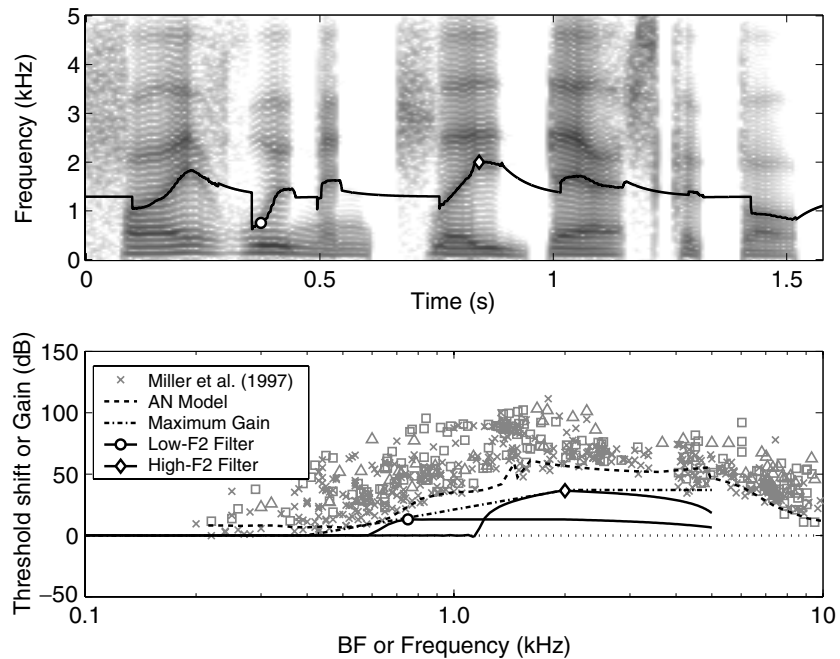


Figure 3. Example of CEFS filtering. Top panel: spectrogram of the test sentence ‘five women played basketball’ with the estimated $F2$ trajectory given by the solid line. The circle indicates an instance when $F2$ had a relatively low frequency and the diamond indicates an instance when $F2$ had a relatively high frequency. Bottom panel: grey symbols are individual AN fiber thresholds in an impaired ear relative to the best threshold curve from normal cats (Miller *et al* 1997). The dashed line shows the threshold shift as a function of BF for the computational model. The dot-dashed line shows the maximum CEFS filter gain as a function of frequency. CEFS filter gain-frequency responses for the $F2$ values at the two different times indicated in the top panel are given by the two solid lines, with their $F2$ frequencies indicated by the circle and the diamond, respectively.

The formant tracker used in this paper was that of Bruce *et al* (2002), which provides reliable and smooth estimates of $F2$ and is robust in the presence of background noise. The average signal delays of the time-varying filter and the formant tracker were approximately matched at ~ 10 ms so that formant estimates were passed to the time-varying filter at the appropriate times. The estimated $F2$ trajectory for the test sentence is shown in the top panel of figure 3, plotted over the spectrogram of the sentence.

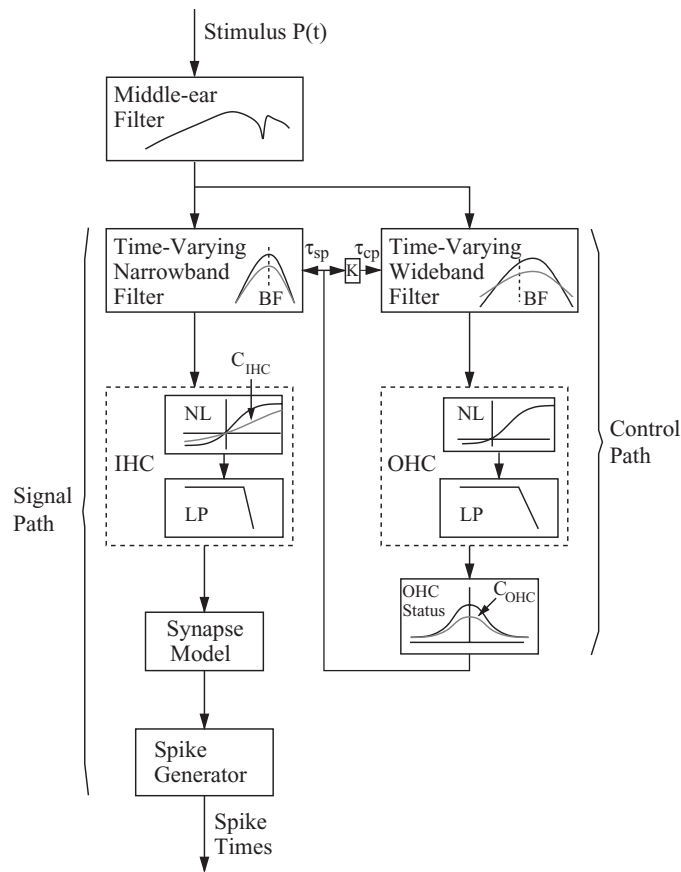


Figure 4. AN model from Bruce *et al* (2003). Abbreviations: outer hair cell (OHC); low-pass (LP) filter; static nonlinearity (NL); inner hair cell (IHC); best frequency (BF). C_{IHC} and C_{OHC} are scaling constants that control IHC and OHC status, respectively. Reprinted with permission from the Acoustical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

The time-varying high-pass filter was a 201 point FIR filter, designed using the least-squares method and implemented at a sampling frequency of 10 kHz. The desired frequency response, updated each sample based on the current $F2$ estimate, had unity gain in the low-frequency ‘stopband’ and a high-frequency ‘passband’ gain based on the $F2$ frequency and the hearing loss profile, as indicated by the dot-dashed line in the bottom panel of figure 3, to give an approximate ‘half-gain rule’ (Dillon 2001). The passband cutoff frequency was 50 Hz below the $F2$ frequency, and the transition region between the stopband and the passband had a slope of approximately 160 dB/decade. Two examples of actual filter frequency responses for different $F2$ values are shown by the solid lines in the bottom panel of figure 3.

2.3. Auditory periphery model

Figure 4 shows the auditory periphery model of Bruce *et al* (2003) used in this paper to obtain a physiological representation of the test sentence following application of the amplification schemes described above. This model has been extensively tested with both simple acoustic stimuli (Zhang *et al* 2001) and speech-like stimuli (Bruce *et al* 2003), and it has been shown

to be predictive of human speech intelligibility (Bondy *et al* 2004). The model takes any arbitrary acoustic waveform as input and produces simulated spike times for an auditory nerve fiber with a specified best frequency (BF) as output.

Anatomical investigations of acoustically traumatized ears (Liberman and Dodds 1984) have shown how damage to the inner (IHC) and outer (OHC) hair cells of the cochlea each contribute to the changes in hearing sensitivity (i.e., elevation in thresholds) and frequency selectivity (i.e., broadening of tuning curves). Two parameters in the computational model, C_{IHC} and C_{OHC} , can be adjusted independently to describe the effects of IHC and OHC impairment (Bruce *et al* 2003). These parameters were varied as a function of the model fiber's BF to produce a moderate–severe high-frequency hearing loss as shown by the dashed line in the lower panel of figure 3.

2.4. Analysis of model spike trains

Forty model auditory nerve fibers with BFs logarithmically spaced between 0.1 and 5 kHz were simulated. For each model fiber, simulated spike trains were collected for 100 presentations of the test sentence with a particular amplification scheme applied. An average poststimulus time histogram (PSTH) was created by binning spike occurrences in 0.1 ms bins, and the PSTH was normalized to have units of spikes/s. The frequency response of the PSTH $p[n]$ over time was evaluated by applying a 25.6 ms moving Hamming window $w[m]$ and calculating the normalized magnitude of the discrete short-time Fourier transform (STFT) $R[n, k]$, referred to as the synchronized rate (Miller *et al* 1997)

$$R[n, k] = \frac{\left| \sum_{m=0}^{255} p[n+m]w[m] \exp\left(-j\frac{2\pi}{256}km\right) \right|}{\sqrt{256 \sum_{m=0}^{255} w[m]^2}}, \quad 0 \leq k \leq 128, \quad (1)$$

where n is the sample (i.e., bin) number of the PSTH, m is the sample number relative to the start of the moving Hamming window and k is the frequency component index. For a 25.6 ms window length, the frequency resolution of the discrete STFT is $39\frac{1}{16}$ Hz, so the equivalent frequency for each frequency component index in Hz is $f = k \times 39\frac{1}{16}$ Hz. In this analysis, the Hamming window was moved in steps of 64 samples, such that $n = r \times 64$, $r = 0, 1, 2, \dots$. The term in the denominator of (1) is to correct for the attenuation of the PSTH energy by the window function $w[m]$.

Following the synchronized rate analysis, the degree to which a model fiber synchronized to a particular formant F_x in the speech spectrum for a window starting at sample n was quantified by the power ratio $\text{PR}[n, F_x]$ (Miller *et al* 1997). The power ratio is defined as the sum of the power in the components of the synchronized rate that can be attributed to a particular formant, divided by the total power in the synchronized rate across all frequency components. A PR of 0 indicates no synchrony to the formant F_x , and a PR of 1 indicates complete synchronization to formant F_x alone.

Spectral smearing caused by the windowing function was accounted for in the following manner. If the frequency of F_x at sample index n exactly matched a discrete-STFT frequency component, i.e., was an integer multiple of $39\frac{1}{16}$ Hz, the synchronized rates from the frequency components to each side of the exact formant frequency were included in the power ratio calculation, along with the exactly matching frequency component, giving a total of three frequency components with indices given by the vector $\mathbf{k}_x[n]$ considered to be resulting from that formant. If the frequency of F_x at sample index n did *not* exactly match a discrete-STFT frequency component, i.e., was *not* an integer multiple of $39\frac{1}{16}$ Hz, the synchronized rates from the two frequency components to each side of the exact formant frequency were included

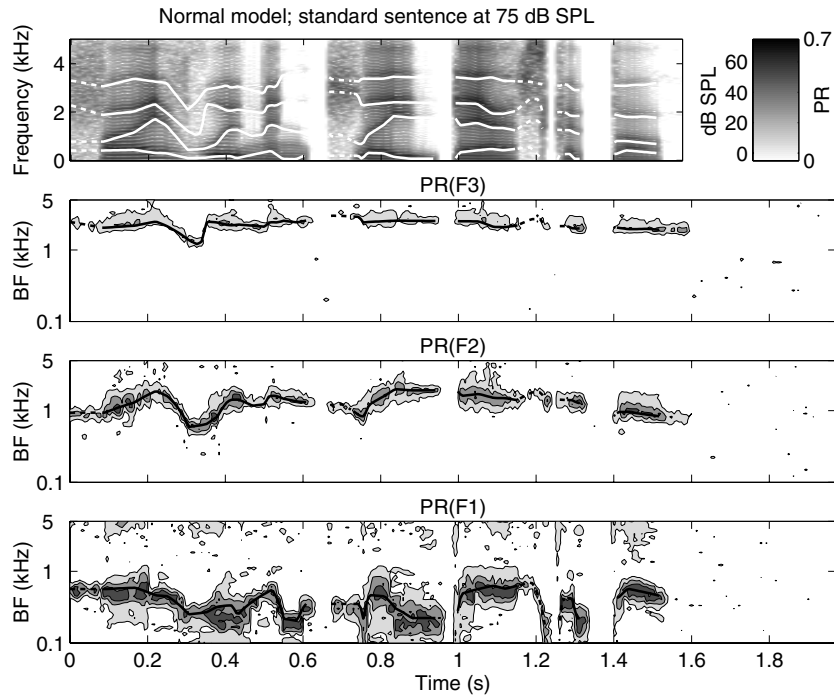


Figure 5. Formant power ratios (PRs) for the normal AN model and the standard sentence at 75 dB SPL.

in the power ratio calculation, giving a total of four frequency components with indices given by the vector $\mathbf{k}_x[n]$ considered to be resulting from that formant.

Thus, the power ratio at sample index n for formant F_x is given by

$$\text{PR}[n, F_x] = \frac{\sum_{k_x} R^2[n, \mathbf{k}_x]}{\sum_{k=2}^{128} R^2[n, k]}, \quad (2)$$

where the frequency used for F_x at sample n takes into account that (i) the Hamming window is centered at sample $n + 128$ and (ii) the average latency of model auditory nerve discharges is approximately 5 ms. Note that the total power given by the denominator of (2) does not include the synchronized rate components for $k = 0$ or 1, which correspond to the DC component of the synchronized rate, i.e., simply the average discharge rate of the model fiber.

3. Results

Figure 5 shows the PR analysis results for the test sentence (without any processing) presented to the normal model of the ear at 75 dB SPL. The top panel shows the spectrogram of the test sentence, with a color scale to the right indicating the scale of the spectrogram in dB SPL. The four white lines overlying the spectrogram indicate the formant trajectories for F_1 , F_2 , F_3 and F_4 , from low to high frequency, obtained from the speech synthesizer parameters for the test sentence. The white lines are solid during voiced speech segments where the synthesizer formant trajectories correspond to actual formant transitions (i.e., spectral peaks). The white lines are dashed during unvoiced speech segments where the synthesizer formant trajectories are maintained for the sake of continuity but do not necessarily correspond to any spectral

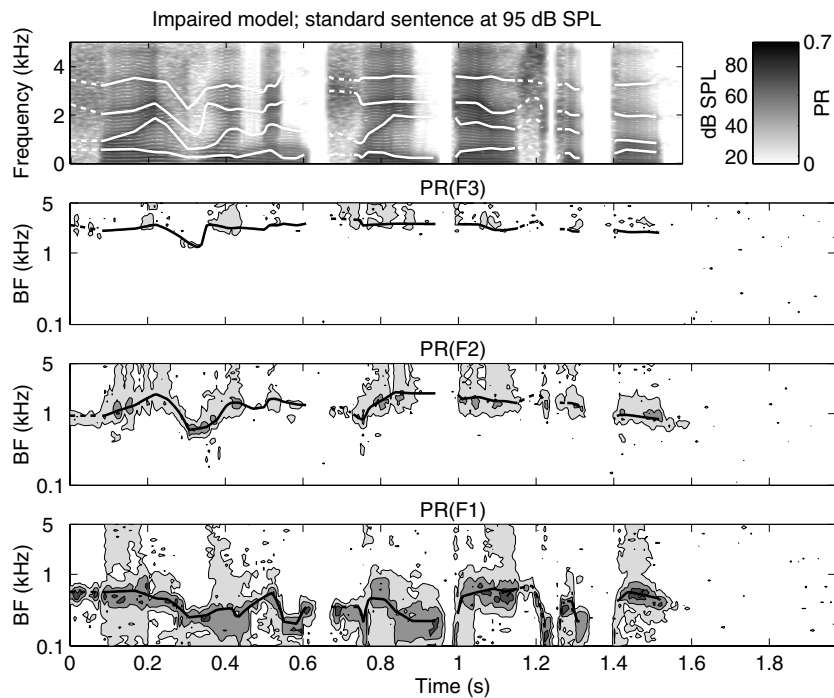


Figure 6. Formant power ratios (PRs) for the impaired AN model and the standard sentence at 95 dB SPL.

peaks. The white lines are missing during periods of silence, but the formant trajectories are assumed to move in a linear fashion to connect the trajectories over these junctions. The PRs of model spike trains were computed over the duration of the entire sentence, so the PRs during unvoiced speech or silent segments give a measure of the ‘noise floor’ of this analysis method.

Shown in the next three panels of figure 5 are the PRs for $F3$, $F2$ and $F1$, respectively, plotted as a function of AN fiber BF on the ordinate and time on the abscissa. The color scale next to the spectrogram also indicates the scale of the PR, and the formant trajectories from the spectrogram are replicated as black lines overlying the PR plots. The PR plots show that normal AN fibers provide a strong representation of formant trajectories during voiced speech segments, by fibers with BFs near the formant frequencies vigorously responding (i.e., synchronizing) to these frequencies. $F1$ exhibits the strongest neural representation, with the magnitudes of the typical PRs somewhat lower for the higher formants.

Figure 6 shows the PR analysis results for the test sentence presented to the impaired model of the ear at 95 dB SPL, using the same plotting conventions as figure 5. The 20 dB increase in presentation level is to compensate for the threshold shifts (i.e., hearing loss) in the high-frequency region and corresponds to flat amplification. The spectrogram appears the same as that of figure 5, but the 20 dB increase in presentation is indicated by the scale bar to the right of the spectrogram. In the lower three panels, we observe weakened representations of the formants, especially $F2$ and $F3$, as measured by the PR magnitude and an upwards spread of synchrony to $F1$ and $F2$ into BF regions well away from the formant frequencies.

Figure 7 shows the PR analysis results for the test sentence with CEFS amplification presented to the impaired model of the ear at 93 dB SPL. The presentation level was reduced

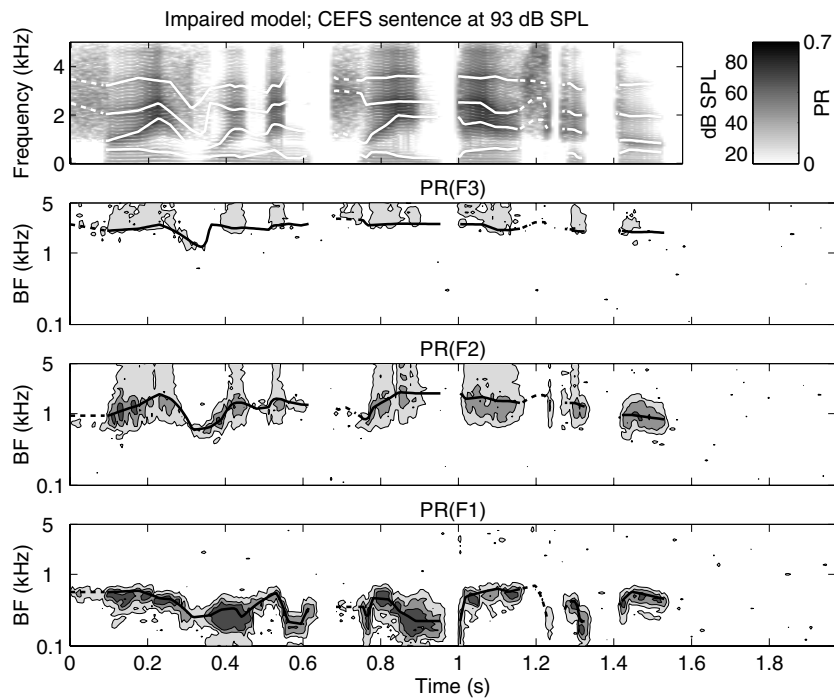


Figure 7. Formant power ratios (PRs) for the impaired AN model and the CEFS sentence at 93 dB SPL.

by 2 dB from that of the flat amplification in figure 6, so that the $F2$ amplitudes during the ‘a’ in ‘basket’ are approximately matched following the two different amplification schemes. The effects of CEFS amplification on the spectrogram in figure 7 is seen by the lower energy for frequency components below the $F2$ trajectory compared to flat amplification (figure 6). In the lower three panels, we observe that CEFS amplification restores a stronger representation of $F1$ and $F2$ than is obtained with flat amplification, and the upwards spread of synchrony to $F1$ is removed. Some upwards spread of synchrony to $F2$ does remain, and synchrony to $F3$ is not restored with CEFS amplification.

Figure 8 shows the PR analysis results for the test sentence with multiband compression followed by CEFS amplification presented to the impaired model of the ear at 93 dB SPL. The effects of applying multiband compression before CEFS amplification on the spectrogram in figure 8 is seen primarily by the reduced variation in the energy from phoneme to phoneme over the duration of the sentence compared to CEFS amplification alone (figure 7). In the lower three panels, it is shown that multiband compression does not adversely affect the benefits of CEFS amplification in restoring a more normal representation of $F1$ and $F2$ and halting the upwards spread of synchrony to $F1$. Furthermore, this implementation of multiband compression acts to equalize the sound intensity across phonemes and to equalize formant amplitudes within phonemes, and consequently a stronger representation is obtained for $F2$ in lower-energy voiced phonemes (e.g., the ‘n’ in ‘women’ at approximately 0.6 s and the ‘a’ in ‘ball’ at approximately 1.5 s) than is produced with CEFS amplification without multiband compression.

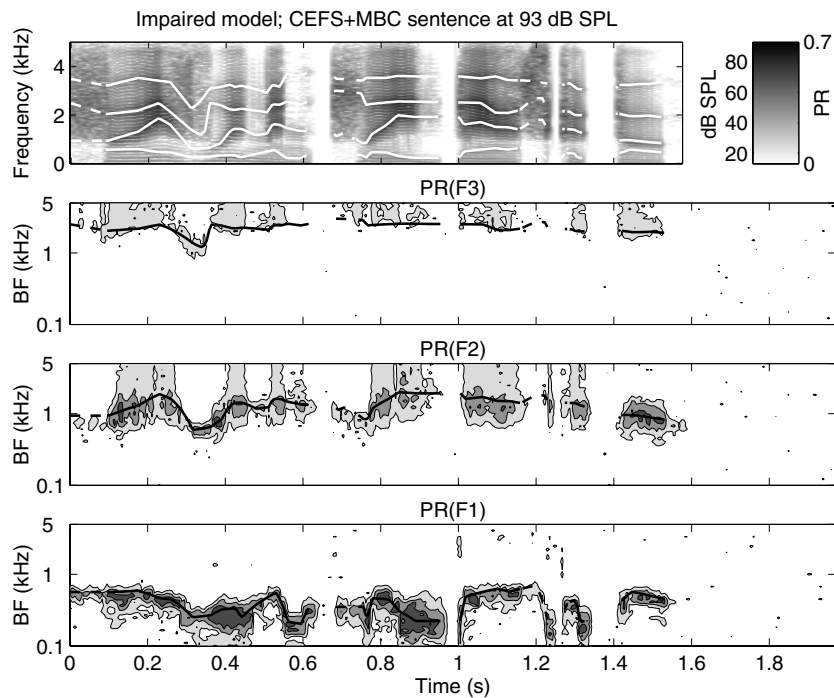


Figure 8. Formant power ratios (PRs) for the impaired AN model and the CEFS+multiband-compression sentence at 93 dB SPL.

4. Discussion

In this paper, a prototype scheme for combining CEFS amplification with multiband compression has been evaluated with a physiological model. The CEFS amplification was applied with a separate time-varying FIR filter from the FFT-based filterbank of the multiband compression algorithm, thereby increasing the overall signal delay of the processing scheme. Long signal delays are known to be disturbing for hearing aid users and can affect both speech production and intelligibility (Stone and Moore 1999, 2002, 2003). We are currently developing an amplification scheme in which CEFS-like filtering can be applied within the FFT-based filterbank of the compression algorithm, thereby reducing the required signal delay.

In addition, the CEFS scheme was designed to restore the normal neural representation of $F1$ and $F2$ only, because these formants are thought to be the primary cues to identifying vowels (Peterson and Barney 1952, Pols *et al* 1969). However, the current CEFS scheme does not prevent the abnormal upward spread of responses to $F2$ and does not restore the neural representation of $F3$, which may be of importance in identifying phonemes characterized by formant transitions. Consequently, we are investigating if the CEFS filter gain-frequency responses can be modified to correct these deficiencies.

Furthermore, only the representation of formants in voiced segments of speech has been considered in this paper. The power-ratio analysis used in this paper does not assess the representation of speech transients such as unvoiced fricatives and plosives, and CEFS has not been designed to improve the representation of such speech transients. Alternative methods for analyzing AN responses to speech transients will be required to develop suitable

transient-enhancement amplification schemes (e.g., Bandyopadhyay and Young 2004, Bondy *et al* 2003, 2004).

The computational AN model will continue to be used as tool for predicting the effects of changes to the CEFS amplification scheme, but evaluation in hearing aid users will be required to determine the actual benefits of CEFS amplification on speech quality and intelligibility for hearing impaired individuals.

Acknowledgments

The author would like to thank Brent Edwards for supplying the code for the multiband-compression algorithm and Eric D Young for discussions on how to analyze the running speech data. This research was supported by NSERC Discovery Grant 261736 and the Barber-Gennum Endowed Chair in Information Technology.

References

- Bandyopadhyay S and Young E D 2004 Discrimination of voiced stop consonants based on auditory nerve discharges *J. Neurosci.* **24** 531–41
- Baer T, Moore B C and Gatehouse S 1993 Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times *J. Rehabil. Res. Dev.* **30** 49–72
- Bondy J, Bruce I C, Becker S and Haykin S 2004 Predicting speech intelligibility from a population of neurons *NIPS 2003 Conference Proceedings: Advances in Neural Information Processing Systems* vol 16, ed S Thrun, L Saul and B Schölkopf (Cambridge, MA: MIT Press) pp 1409–16
- Bondy J, Bruce I C, Dong R, Becker S and Haykin S 2003 Modeling intelligibility of hearing-aid compression circuits *Conference Records of the 37th Asilomar Conference on Signals, Systems and Computers* vol 1, pp 720–4
- Bruce I C, Karkhanis N V, Young E D and Sachs M B 2002 Robust formant tracking in noise *Proc. 2002 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing: ICASSP (Orlando, FL)* vol 1 pp 281–4
- 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** 369–88
- Dillon H 2001 *Hearing Aids* (New York: Thieme Medical Publishers)
- Franck B A, van Kreveld-Bos C S, Dreschler W A and Verschuure H 1999 Evaluation of spectral enhancement in hearing aids, combined with phonemic compression *J. Acoust. Soc. Am.* **106** 1452–64
- Kiang N Y, Liberman M C and Levine R A 1976 Auditory-nerve activity in cats exposed to ototoxic drugs and high-intensity sounds *Ann. Otol. Rhinol. Laryngol.* **85** 752–68
- Liberman M C and Dodds L W 1984 Single-neuron labeling and chronic cochlear pathology: III. Stereocilia damage and alterations of threshold tuning curves *Hear. Res.* **16** 55–74
- Lyzenga J, Festen J M and Houtgast T 2002 A speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity *J. Acoust. Soc. Am.* **112** 1145–57
- Miller R L, Calhoun B M and Young E D 1999 Contrast enhancement improves the representation of /ε/-like vowels in the hearing-impaired auditory nerve *J. Acoust. Soc. Am.* **106** 2693–708
- Miller R L, Schilling J R, Franck K R and Young E D 1997 Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers *J. Acoust. Soc. Am.* **101** 3602–16
- Moore B C 1995 *Perceptual Consequences of Cochlear Damage* (New York: Oxford University Press)
- Peterson G E and Barney H L 1952 Control methods used in a study of the vowels *J. Acoust. Soc. Am.* **24** 175–84
- Pols L C, van der Kamp L J and Plomp R 1969 Perceptual and physical space of vowel sounds *J. Acoust. Soc. Am.* **46** 458–67
- Robertson D 1982 Effects of acoustic trauma on stereocilia structure and spiral ganglion cell tuning properties in the guinea pig cochlea *Hear. Res.* **7** 55–74
- Robles L and Ruggero M A 2001 Mechanics of the mammalian cochlea *Physiol. Rev.* **81** 1305–52
- Simpson A M, Moore B C and Glasberg B R 1990 Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners *Acta Otolaryngol. Suppl.* **469** 101–7
- Stone M A and Moore B C 1992 Spectral feature enhancement for people with sensorineural hearing impairment: effects on speech intelligibility and quality *J. Rehabil. Res. Dev.* **29** 39–56

-
- Stone M A and Moore B C 1999 Tolerable hearing aid delays: I. Estimation of limits imposed by the auditory path alone using simulated hearing losses *Ear Hear.* **20** 182–92
- Stone M A and Moore B C 2002 Tolerable hearing aid delays: II. Estimation of limits imposed during speech production *Ear Hear.* **23** 325–38
- Stone M A and Moore B C 2003 Tolerable hearing aid delays: III. Effects on speech production and perception of across-frequency variation in delay *Ear Hear.* **24** 175–83
- Summerfield A Q, Foster J, Tyler R S and Bailey P J 1985 Influences of formant narrowing and auditory frequency selectivity on identification of place of articulation in stop consonants *Speech Commun.* **4** 213–29
- Zhang X, Heinz M G, Bruce I C and Carney L H 2001 A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression *J. Acoust. Soc. Am.* **109** 648–70