Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Signal Processing 84 (2004) 1239–1253

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

A novel signal-processing strategy for hearing-aid design: neurocompensation

Jeff Bondy[a,*], Sue Becker[b], Ian Bruce[a], Laurel Trainor[b], Simon Haykin[a]

[a] *Department of Electrical and Computer Engineering, McMaster University, 1280 Main Street West, Hamilton, Canada ONL8S 4K1*
[b] *Department of Psychology, McMaster University, 1280 Main Street West, Hamilton, Canada ONL8S 4K1*

Received 10 December 2003; received in revised form 3 April 2004

Abstract

A novel approach to hearing-aid signal processing is described, which attempts to re-establish a normal neural representation in the sensorineural impaired auditory system. Most hearing-aid fitting procedures are based on heuristics or some initial qualitative theory. These theories, such as loudness normalization, loudness equalization or maximal intelligibility can give vastly different results for a given individual, and each may provide variable results for different hearing impaired individuals with the same audiogram. Recent research in characterizing sensorineural hearing loss has delineated the importance of hair cell damage in understanding the bulk of sensorineural hearing impairments. A novel methodology based on restoring normal neural representation after the sensorineural impairment is presented here. This approach can be used for designing hearing-aid signal processing algorithms, as well as providing a general, automated means of predicting the relative intelligibility of a given speech sample in normal hearing and hearing impaired subjects.

Keywords: Speech perception; Hearing-aids; Sensorineural impairment

1. Introduction

There have been several advances in our understanding of the neurophysiological basis of hearing impairment. The insight that hair cell damage alters the auditory system has profound effects on the design of hearing-aid systems to combat sensorineural loss. While conductive loss, which can arise after ossicle damage or an ear drum puncture, can largely be overcome with frequency-shaped linear amplification, the types of impairment associated with inner hair cell (IHC) and outer hair cell (OHC) damage require a new suite of algorithms. Up until the mid-1980s the mechanisms underlying the more prevalent type of impairment, hair cell loss, were not well understood. This led to a group of ad hoc algorithms, largely based on the discerned symptoms (spectrally shaped sensitivity loss, identification in noise problems) as opposed to the mechanisms underlying the symptoms.

The approach to evaluating hearing aids in the audiological field has been solely empirical. By using neuro-physiologically based auditory models, we show predictive measures for offline evaluation. These measures are consistent with experimental data on human performance. Furthermore, using auditory models, one can form a general hearing-aid algorithm design methodology, whose performance in

* Corresponding author.
E-mail address: jeff@soma.ece.mcmaster.ca (J. Bondy).

0165-1684/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2004.04.006

simulations shows a high correlation with empirical evidence.

The processing of an acoustic signal by the peripheral auditory system can be summarized as follows. A sound signal is directed to the ear canal by the pinna (outer ear). The eardrum responds to the pressure wave by deflecting. This deflection causes the three small bones of the inner ear to move, producing a similar movement in the oval window of the cochlea. This vibration starts a travelling wave in the fluid of the cochlea. Up to this point, the system is well characterized by a linear transfer function, but beyond this point, the system is highly nonlinear. The travelling wave produces a peak displacement at some point along the cochlea that is a function of frequency and outer hair cell (OHC) undamping. OHCs are motile members that precisely modulate the basilar membrane; the basilar membrane is tonotopically mapped. Inner hair cells (IHCs) transduce the mechanical displacement of the basilar membrane to nerve firings. The OHCs undamping enhances the IHCs sensitivity and selectivity [29].

The loss of these hair cells produces symptoms such as elevated thresholds, loss of frequency selectivity, and loss of temporal discrimination [26,23]. The consequences of hair cell damage for auditory discrimination are far ranging, taking entire books to catalogue [26]. Necessary for future hearing-aid algorithms is a quantitative understanding of how IHC and OHC loss affects the processing of the auditory system and how that processing affects perception.

The objective of our research is to restore near-normal firing patterns in the auditory nerve, in spite of hair cell damage. While there is some evidence of reorganization in the mammalian auditory cortex [18] resulting from hair cell damage, there is no present evidence that the basic cortical circuitry does not work. That is, the processing in the brain that is eminently capable of segregation, streaming, and decoding, may still be able to function properly if the incoming signals are parcelled properly.

A normal hearing process can be described as the block diagram in Fig. 1, where an input signal X is transformed by the auditory periphery, H, to produce a neural response Y.

The auditory periphery is treated as a black box in many signal-processing applications. In the hearing

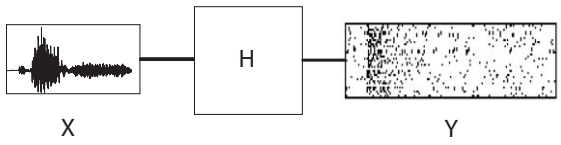

Fig. 1. Block diagram representation of normal hearing system.

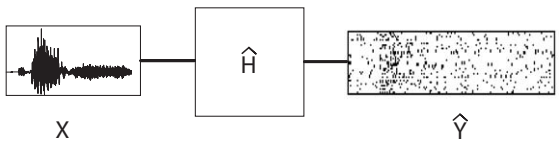

Fig. 2. Block diagram representation of impaired hearing system.

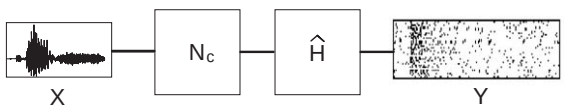

Fig. 3. Block diagram representation of our compensation scheme.

compensation application this approach has severe limitations. The success of the algorithm will be directly proportional to the amount of information about H that one embeds in the design. With the loss of hair cells the functionality of H changes, resulting in a hearing impaired system as shown in Fig. 2. That is, the same input signal produces a distorted neural signal, $\hat{Y}$, when processed by the damaged hearing system $\hat{H}$.

The first step in our method of compensating for hair cell loss is to alter the input signal going into the impaired system such that after the impaired ear there is a normal neural representation; the algorithm to alter the input signal is called the "Neurocompensator" (proposed by Becker and Bruce [4]), here noted as N_c (Fig. 3).

If $\hat{H}$ was invertible the Neurocompensator would be the cascade of the undamaged model and the inverse of the damaged system or $N_c = H\hat{H}^{-1}$, or $H = N_c\hat{H}$. This approach to hearing aid design has been explored by Anderson [1], Chabries et al. [16], and Anderson et al. [2]. Unfortunately, things are not so simple. The auditory system has very important nonlinearities [21], time variances [27] and many to one mappings. The simple fact that a sound can be completely masked by

the presence of a second sound is evidence that the auditory system discards information. This means a perfect inversion is not possible. However, even if H is non-invertible, one may still be able to capture its capabilities sufficiently to approach normal hearing.

The key idea underlying the technique described below is to use models of the intact and damaged auditory system to evaluate the perceptual impact of hearing compensation algorithms offline. In our method the signal is pre-processed by the Neurocompensator, and then fed into the impaired model. The resulting neural representation is compared to the neural representation of the signal after the normal hearing model. Perceptual distortions from sensorineural impairment should be minimized by the Neurocompensator by re-establishing in the impaired auditory system the normal pattern of neural firing. The methodology therefore hinges on a detailed model of the peripheral auditory system.

2. Auditory model

The auditory periphery model used throughout is from Bruce et al. [9], following initial work by Bruce et al. [10], Heinz et al. [20] and Zhang et al. [39]. The system is shown in Fig. 4.

This model describes the function of the auditory system from the middle ear to auditory nerve. For outer ear functioning the head related transfer function from Wiener and Ross [38] is used. The auditory model itself comprises several section, each providing a phenomenological description of a different part of auditory periphery function.

The first section models middle ear filtering. The second section, labeled the "control path," captures the OHCs modulatory functions, and includes a wideband, nonlinear, time varying, band-pass filter followed by an OHC nonlinearity (NL) and low-pass (LP) filter. This section controls the time-varying, nonlinear behavior of the narrowband signal-path basilar membrane (BM) filter. The control-path filter has a wider bandwidth than the signal-path filter to account for wideband nonlinear phenomena such as two-tone rate suppression.

The third section of the model, labeled the "signal path", describes the filter properties and traveling wave delay of the BM (time-varying, narrowband filter), the nonlinear transduction and low-pass filtering of the inner hair cell (IHC NL and LP), spontaneous and driven activity and adaptation in synaptic transmission (synapse model), and spike generation and refractoriness in the auditory nerve (AN). In this model, C_{IHC} and C_{OHC} are scaling constants that control IHC and OHC status, respectively.

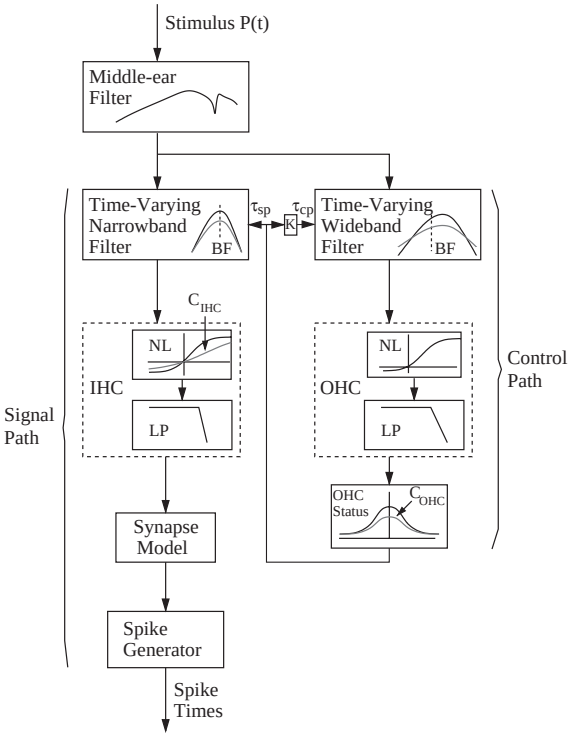

Fig. 4. Block diagram of the computational model of the auditory periphery from the middle ear to the auditory nerve.

The gain functions of linear versions of the time-varying narrowband filter in the signal path, plotted as gain versus frequency deviation Δf from the filter's best frequency (BF) are given in Fig. 5.

The filter is fourth-order and is plotted for five different values of τ_{sp} between τ_{narrow} and τ_{wide}. τ_{sp} is the time-bandwidth control parameter, where larger values correspond to more frequency selectivity, and $\tau_{sp}\varepsilon[\tau_{wide}, \tau_{narrow}]$. $\Delta\tau=\tau_{narrow}-\tau_{wide}$. τ_{narrow} was chosen to produce a 10 dB bandwidth of ~450 Hz, and τ_{wide} was chosen to produce a maximum gain change at BF of ~ 41 dB at 1.7 kHz. This plot can be interpreted as showing the nominal tuning of the filter with normal OHC function at five different sound pressure

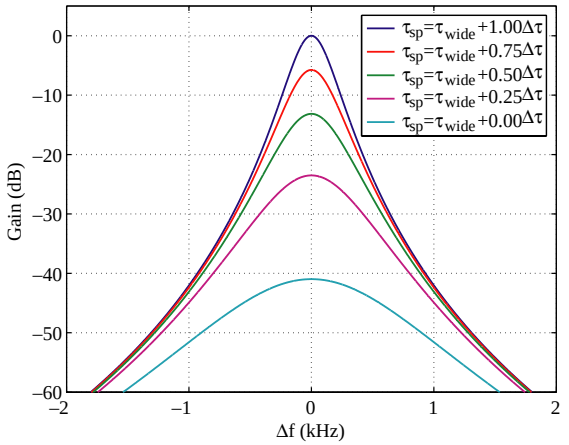

Fig. 5. Filter shaping functions of the time-varying narrow-band filter in the signal path, plotted as gain versus frequency deviation (Δf) from BF. This example is at 1.7 kHz.

levels, or alternatively, as the nominal tuning of the filter for five different degrees of OHC impairment.

The success of the Neurocompensator strategy presented below depends upon the accuracy of the auditory model of the normal and damaged ear. The Bruce et al. [9] model, while being based on cat physiology is thought to correspond very closely with human physiology. This particular model has a long history of development and good fit to a wide range of empirical data. The auditory model can capture a range of phenomena due to hair cell nonlinearities, including loudness-dependent sensitivity and bandwidth modulation (as stimulus intensity increases the output response levels off and frequency-tuning becomes broader), and masking effects such as two-tone suppression. Additionally, the model incorporates critical properties of the auditory nerve response including synchrony capture in the normal and damaged ear and replicates several fundamental phenomena observed in electrophysiological experiments in animal auditory systems subjected to noise-induced hearing loss. For example, with OHC damage, high frequency auditory nerve fibers' tuning curves become asymmetrically broadened toward the lower frequencies and tend to become synchrony locked to lower frequencies.

In theory, the Bruce et al. [9] model is capable of simulating auditory nerve responses in both a normal and damaged human auditory system accurately. The damaged model must be tuned to the parameters of a particular individual's hearing-impairment. This requires estimates of both inner and outer hair cell loss over a range of frequencies. The standard audiological assessment, the audiogram, simply measures the threshold for pure tones at each of a small set of frequencies. An elevation in pure tone threshold cannot differentiate between a reduction in OHC driven gain versus a loss of IHCs tuned to that frequency. In sensorineural hearing disorders, it is generally assumed that a moderate elevation in threshold primarily reflects OHC loss, while a severe elevation reflects an additional IHC loss. Although this pattern is typical in individuals with age-related and noise-induced hearing loss, the exact proportion of IHC to OHC loss may deviate from the typical pattern in some individuals, and also may not hold at all for individuals with less common types of sensorineural damage, e.g. drug-induced. Better methods for estimating, separately, the degree of inner and outer hair cell loss, such as using noise-masked tones [28] are intrinsic to this strategy. Given accurate measurements, the model could be tailored to compensate for many individual patterns of deficits.

3. Damaged models and methods

Initial testing of our algorithm to establish the validity of our model-based approach was accomplished using idealized models of hearing impairment similar to those described in Byrne et al. [12]. The different audiograms are given in Fig. 6A, and the individual contributions of inner and outer hair cell losses to each loss profile are given in Fig. 6B–F.

The impairments of inner and outer hair cells per frequency were calculated so that OHC impairment accounted for around 50–60% of the total threshold shift in dB [28]. The percent IHC loss was then adjusted to explain the remaining threshold shift. Loss profiles 3, 4, and 5 are indicative of presbyacucis or progressive sound-induced hearing loss and are more typical of the normal hearing loss pathology.

The stimulus presented to the normal and hearing impaired auditory model was Gaussian noise shaped to have the same spectrum as the long term average speech spectrum (LTASS, [13], combined data). A 200 ms LTASS input was sampled at 500 kHz and

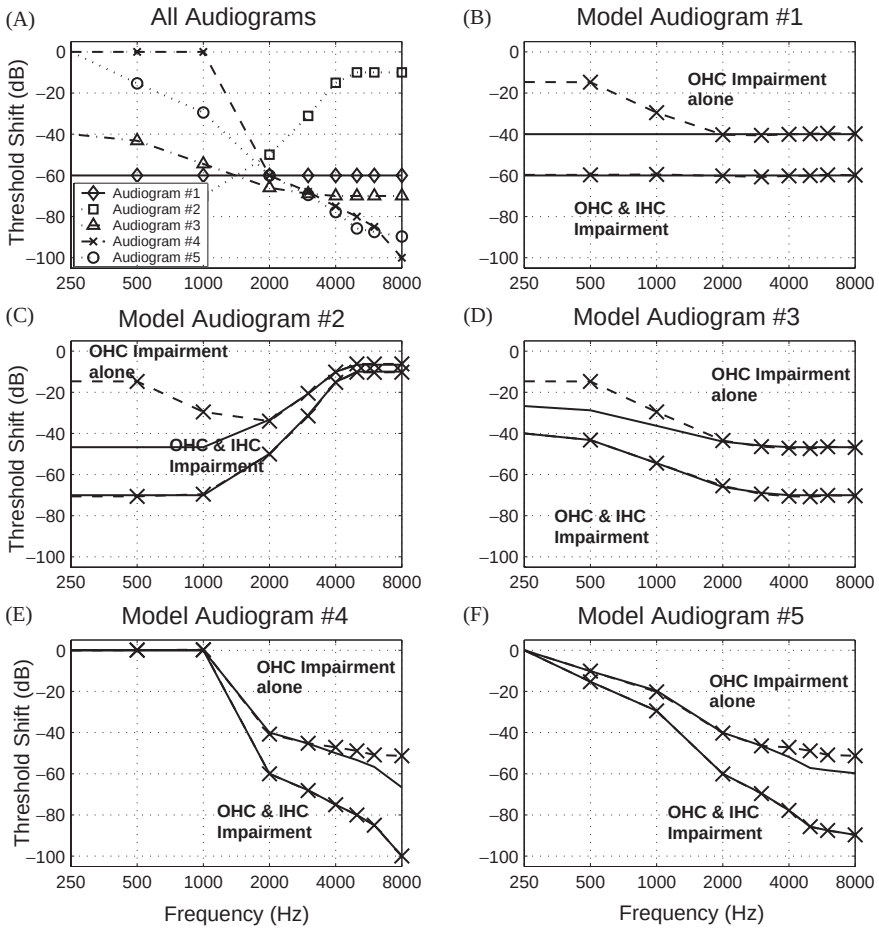

Fig. 6. Hearing threshold shift for five candidate audiograms taken from Byrne et al. [12]. Differences are due to approximating threshold shift with IHC and OHC damage. The dashed and X marked lines are ideal estimated audiograms, while the solid lines are the actual audiogram provided by the Bruce et al. [9] model.

presented at 75 dB SPL into the normal model. The impaired model input would be processed by some test compensation strategy before being input into one of the impaired models, so the power level would fluctuate depending on the compensation strategy used. The high sample frequency is necessary for the Bruce et al. [9] auditory model.

The output of the model was a time series, 230 ms long (the extra time versus the input could be used to judge offset effects), with a 22 050 Hz sample rate, of instantaneous neural spike rates across seven octave bands, starting at 125 Hz and ending at 8000 Hz. The neural best frequencies were chosen to mimic the Byrne et al. [12], (Fig. 2) audiogram data points. A typical output of a normal and impaired auditory model at 250 Hz is given in Fig. 7.

The error function we used was first derived for normal hearing in Bondy et al. [6], it follows the development of the speech transmission index (STI, [3]) but with neural equivalents. It is hoped that encompassing the cochlear processing and impairment will circumvent the problems [19] had with using the articulation index to fit hearing-aids. Our objective function for assessing differences between normal and impaired neural representations is a modified Hebbian rule. Where a Hebbian rule would strictly quantify

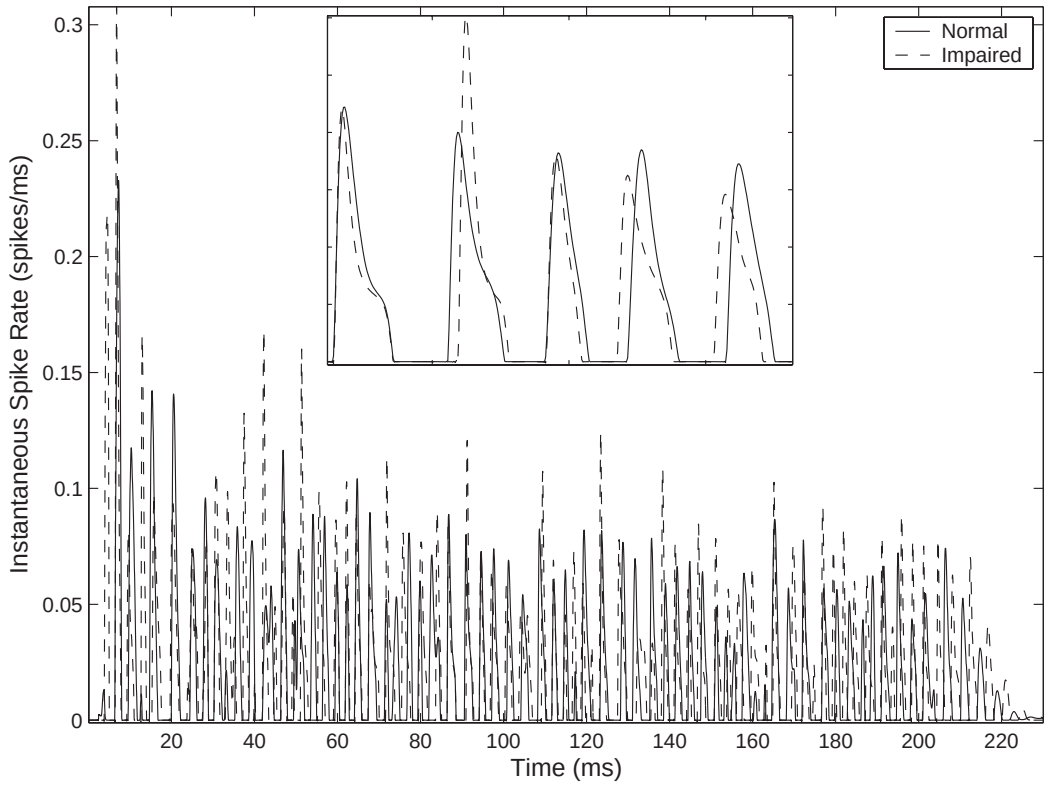

Fig. 7. Time plot of instantaneous spiking rate versus time for the normal auditory model and the impaired auditory model. The impaired model is based on Audiogram #4 from Fig. 6 with the half-gain rule applied (see Section 4) to the input. The inset shows a closeup, where synchrony is very evident, as well as some differences between the Normal and Impaired outputs.

correlate firings between the two models, we use the instantaneous rate, so the information in specific spikes is lost. The error in any band is the absolute value of one minus the correlation between the normal and impaired firing patterns and normalized by the average spike rate from the output of the normal auditory model. This error is calculated for each of the seven frequency bands. The error in the ith frequency band, for the jth impaired condition is:

$$\varepsilon_{ij} = \left| 1 - \frac{\vec{x}_i \vec{y}_{ij}^{\mathrm{T}}}{\vec{x}_i \vec{x}_i^{\mathrm{T}}} \right|, \qquad (1)$$

where $\vec{x}$ is the normal auditory model instantaneous spiking rate vector, and $\vec{y}$ is the impaired auditory models instantaneous spiking rate vector over time. This metric cannot capture transient, or timing information of the auditory model because it cannot be coded through synchrony capture. The LTASS input is a spectrally steady-state signal, so this Hebbian metric can capture distortion in the response. For the following results this simply coincides with a statistically mean processing strategy. This is loosely equivalent to using a signal-to-noise ratio (SNR) metric, since most of the power in an utterance is due to voiced speech, the SNR captures mostly effects of voiced speech. Synchrony capture is very evident in the auditory nerve during voiced speech. Further research is necessary to deal with the auditory systems time-adaptive characteristics.

The individual bands were then summed into a single error value with a weighting function following the STI [34] frequency importance weighting, but derived for the neural representation specifically [6]. The

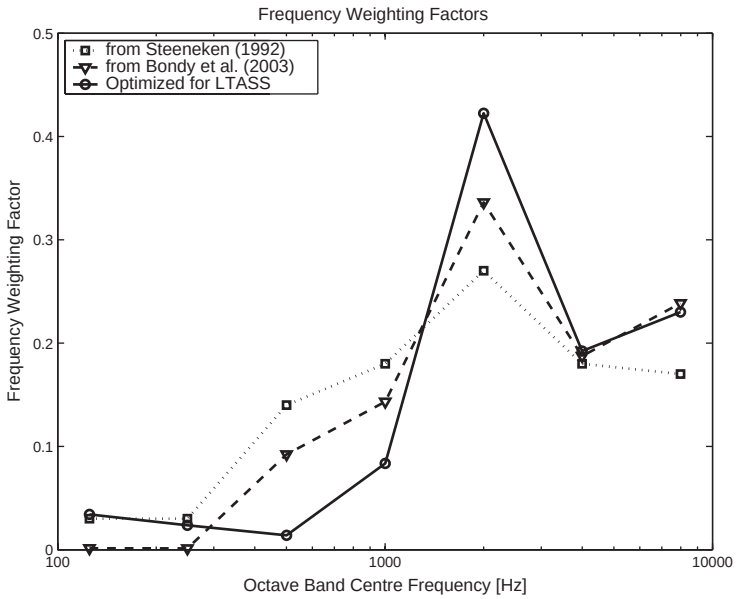

Fig. 8. Frequency weighting factors used in calculating the neural distortion, as well as factors derived from a different stimuli [6] and an acoustic signal counterpart [33]. There are large differences between the importance of different frequencies for different corpuses.

total error is calculated using

$$\text{Error}_j = \sum_{i=1}^{N} \alpha_i \cdot \varepsilon_i, \tag{2}$$

where α_i's are the bands importance weighting functions (shown in Fig. 8), and the ε_i's are calculated through Eq. (1).

Bondy et al. [6] shows that this weighted sum of Hebbian error is a monotonic function of intelligibility. This is important for offline assessment because it gives us a way to judge different hearing-aid processing strategies against one another. The actual error will be a relative indicator between strategies under test, not an intelligibility value. While, there is a rough correlation between this error function and intelligibility, it is best to say that we are assessing the degree that a processing strategy returns a normal auditory representation, and that representation includes things like loudness normalization and synchrony capture.

The process of predicting efficacy of hearing-aid algorithms is to take the LTASS stimulus, pass it through the normal auditory model and then take the same stimuli, preprocess it with the hearing-aid algorithm under test and pass it through an impaired auditory model, whose audiogram loss follows one of the profiles in Fig. 6. Then calculate the neural distortion following Eq. (1). Repeat this for several frequencies and sum this error following Eq. (2). The next section illustrates how this predictive measure closely fits empirical data, and subsequent section extends this to train new processing strategies for hearing-aids.

4. Relation to empirical data

The validity of restoring normal auditory nerve activity is tested by asking the question: would the standard fitting strategies, when applied to the input of the impaired model, result in optimal improvements in neural coding?

Early papers in audiology attempted to describe the amount of gain necessary for comfort and intelligibility. Markle and Zaner [24] give data showing how restoring normal hearing thresholds by setting the gain in each frequency band exactly equal to the threshold loss results in a signal too loud to be comfortable or intelligible. Rather than employing a one-to-one gain to threshold loss, Byrne and Fifield [14] found that a

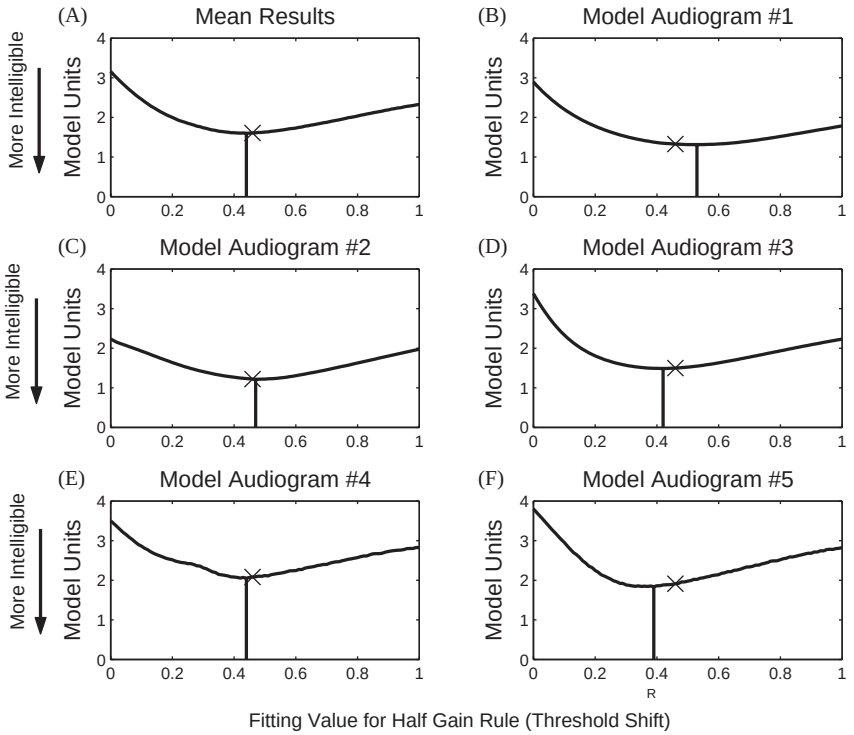

Fig. 9. Neural error function versus gain ratio (R), showing a minimization of differences between the normal and impaired auditory models for a hearing-aid gain ratio. The vertical line is at the ratio (R) which minimizes the error curve, the X is at the value predicted by Byrne and Fifield [14] data (0.46). The mean scheme in A minimizes neural differences between the normal and impaired at a ratio of gain to hearing loss of 0.44 which is very close to the empirical data of 0.46 dB of gain per dB of hearing loss. The Threshold Shift is raised to the power of the ratio since the fitting strategies correspond to a dB:dB gain. Graphs B–F show the error curves for each loss profile.

0.46-to-one gain to threshold shift was optimal. That is, for every 10 dB threshold shift, the gain for optimum intelligibility should be 4.6 dB. We started with modeling the Byrne and Fifield [14] data because it is the basis for the widely used [25] fitting strategies from the National Acoustics Lab of Australia (NAL-1, NAL-R, NAL-RP, NAL-NL1…).

The first experiment modeled the neural representation distortion introduced by setting different gains per dB threshold shift. Multiple ratios, R, are modeled with the gain in dB per dB of loss changing from 1:1, or 1 dB of gain per dB of threshold loss (mimicking Markle and Zaner [24]) to 0:1, or no processing whatsoever.

The LTASS input stimulus and the error calculation is described in Section 3. The experiment is run for all five loss profiles with a sweep of the gain ratio, and the results are shown in Fig. 9, the y-axis is in model units, with larger values representing more distortion in the auditory nerve.

The vertical line drawn at the minimum error point for the average of the five model audiograms is 0.44 dB gain per dB threshold loss. Clearly this result is very close to empirical evidence (0.46). Model audiograms three, four and five are more indicative of typical hearing loss pathology and these have less individual error than the flat audiogram of loss profile one. Another important insight is that the more acute losses need higher gain. This is consistent with empirical data for fitting the profoundly impaired.

Brooks [7] gives a possible theoretical footing for increasing gain at half the hearing threshold degradation by pointing out that the most comfortable level is approximately half way between threshold and the maximum tolerance level. Kates [22] suggests that the half-gain rule is based on the natural compression ratio

Table 1
Frequency shaping gain values for NAL-R

	Frequency (Hz)						
	250	500	1000	2000	3000	4000	6000
k_i (dB)	−17	−8	1	−1	−2	−2	−2

of the active ear. These results show another possible scenario: people are trying to fit their gain response to return normal neural representation.

The next experiment dealt with the prediction of linear fitting strategies. The formula used for the following examples is the NAL-RP [15] formula, introduced as the NAL-R [11]. Without the profound loss additional gain factors NAL-RP is

$$H_{3FA} = (H_{500} + H_{1000} + H_{2000})/3, \quad (3)$$

$$X = 0.15 * H_{3FA}, \quad (4)$$

$$IG_i = X + R * H_i + k_i. \quad (5)$$

Here H_i is the threshold shift measured at frequency i Hz. H_{3FA} is the average threshold shift at 500, 1000 and 2000 Hz, X is a gain factor across all frequencies, and R is the gain in dB for each dB of loss. In the NAL-RP formula, $R = 0.31$. The insertion gain at frequency i, IG_i, is made up of the constant gain factor X, 0.31:1 dB of gain per dB of threshold shift at that frequency, and a gain factor that is dependent upon the frequency, k_i. k_i is described in Table 1.

Following the half-gain rule experiment, the neural error was used to try to predict the constants in the NAL-RP formula. Again using LTASS input at 75 dB SPL, the same input and output frequencies, and the same loss profiles, we checked to see if the gain per dB of threshold shift in Eq. (5) (0.31:1) is predicted. The curve in Fig. 10 is the neural distortion when the multiplier (R) is swept from zero to one, with the fitting strategy having a historical value of 0.31.

The minimum error of the multiplier for the strategy that attempts to restore neural firing patterns (0.34), based on our simulations, closely matches the NAL-RP fitting strategy's multiplier (0.31), derived through empirical evidence. Most of these curves have a lower minimization point with the NAL-RP formula than the half-gain rule as well, with the exception of loss profile #4.

Rankovic [30] reports that people with profound hearing loss found that fitting with a high amount of gain in high loss frequency regions had their intelligibility reduced. Fig. 10 higher tail towards the full 1:1 dB of gain per dB of loss agrees with Rankovic, if intelligibility is a monotonic function of neural distortion.

The final modeling was to predict the NAL-RP frequency weighting factors (k_i). Here, a simultaneous optimization of the seven frequency gain factors, starting at unity gain, was simulated. All initial condition were as before. The gain per dB of threshold shift was the NAL-RP recommended 0.31:1, not 0.34:1 as recommended by the previous experiment, the remaining NAL-RP factors were used. It is hoped that the frequency weightings calculated would closely match NAL-RP's k_i factors. The optimized frequency weightings versus the original weightings are in Fig. 11. This is for loss profile #4, as a profile that is typical of moderate age or noise induced hearing loss.

The calculated frequency mask in Fig. 11 has had a small, flat gain shift (less than a dB, which could coincide with the different optimal gain ratios) applied to it before being plotted to center it on the NAL-RP gain curve, emphasizing the differences in shape.

The basics of the calculated frequency mask and the prescribed mask are clearly similar. The general low frequency attenuation, and the second formant range being emphasized is represented in both. There are differences including a lowering of the gain at the knee point of the audiogram and much lower high frequency gain. The knee point effect could be introduced by some nonlinearity between the normal hearing region and impaired region, or a model effect. The lower high frequency gain could be similar to how NAL-NL1 [12] limits gains in highly damaged regions.

The above set of experiments started out simply and attempted to increase the complexity to test the basic assumption that offline modeling to return neural patterns to a hearing impaired auditory system corresponds to empirical data. This culminated in a metric that shows pronounced similarities to experimental data while being able to optimize multiple parameters. This section attempted to illustrate a connection between traditional, empirically derived hearing aid fitting strategies, and a new quantitative metric based on re-establishing normal neural representations in a hearing impaired individual. The neural error metric

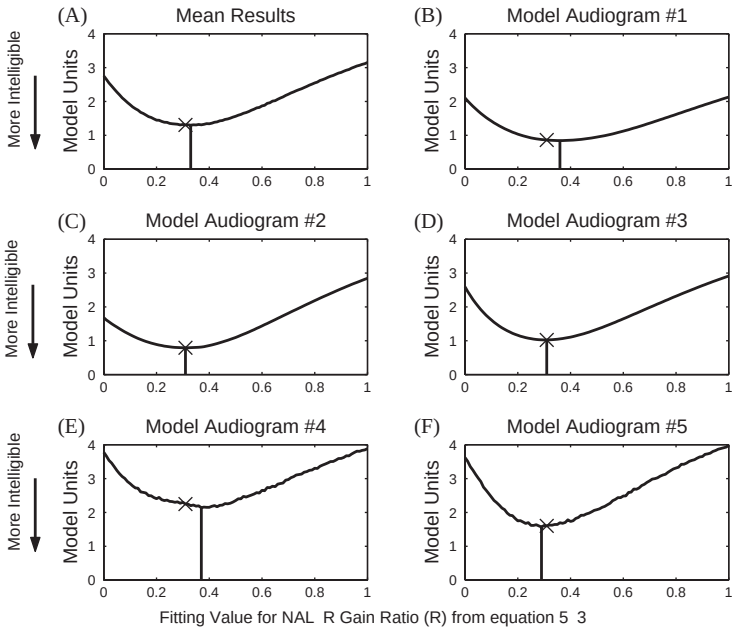

Fig. 10. NAL-RP fitting strategy with the threshold shift versus gain ratio swept to see if the neural distortion error predicts empirical, published results. The X is at $R = 0.31$ and the vertical bar is at the minimum value of the error surface. The mean optimal value predicted by the neural distortion (in A) error is 0.34 dB gain per dB of hearing threshold shift versus the historical data of 0.31. The individual error functions for each model audiogram are given in parts B–F.

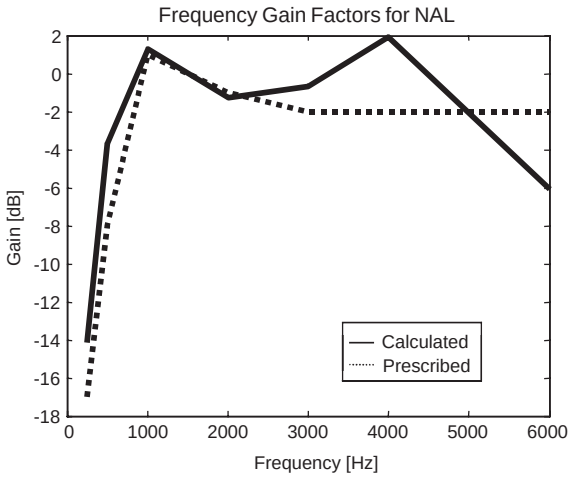

Fig. 11. NAL frequency weightings calculated through neural error metric optimization and what was prescribed in the original NAL-RP strategy.

produced results very similar to pragmatic data, giving credence to the possibility of evaluating many fitting strategies quantitatively, and in corollary: the ability to calculate optimal characteristics in designing hearing-aid algorithms offline.

5. Spectral enhancement from neural compensation

The formation of the metric and validation of the strategy above provides for supervised training of any type of hearing aid algorithm. Since the new compensation strategy relies heavily on neural network type training, and is in essence trying to re-establish normal neural activity, the general processing strategy was coined Neurocompensation. A Neurocompensator is any block whose weights are fitted to an individual's hearing loss through a training sequence that attempts to return the normal neural code. The training sequence is represented in Fig. 12.

The Neurocompensator, N_c, is trained on a set of input signals, supervised by the difference between the output across a set of frequencies of the normal auditory model, H, and the output of the impaired auditory model, H. For each training iteration the

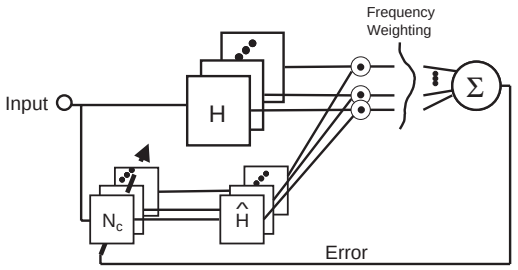

Fig. 12. Block representation of the Neurocompensator training sequence. The dot operator before the frequency weightings corresponds to Eq. (1), the weighting operator corresponds to Eq. (2). The normal and impaired auditory output is a set of the auditory model at different best frequencies, while the Neurocompensator is represented as a different preprocessor at the different frequencies, but this is not necessarily the case. There may be only one Neurocompensator preprocessing block.

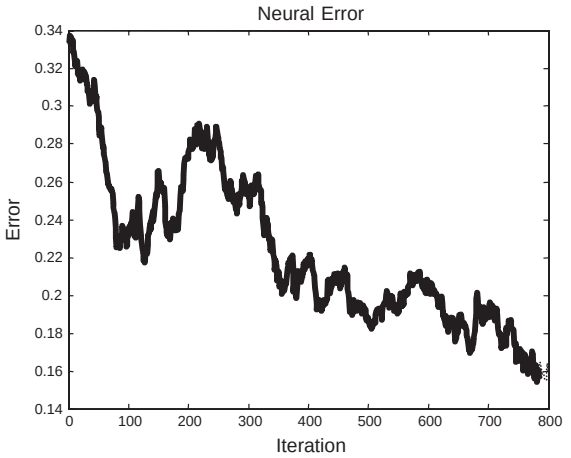

Fig. 13. Training curve for the Neurocompensator for 800 iterations.

Neurocompensator is adjusted by changing weights in its gain function to minimize the error signal. Training with LTASS noise will lead to a Neurocompensator that is optimal in the mean sense. But, more interestingly we are also able to train the Neurocompensator using speech material.

The conceptual compensator block is an attempt at spectral contrast enhancement following Schwartz and Simoncelli [32]. The analytic equation is given in Eq. (6).

$$G_i = \frac{v_i f_i^2}{\sum_j w_j f_j^2 + \sigma}. \tag{6}$$

The gain at a frequency indexed by i, G_i, is a divisive function of the weighted (weighted by v_i) input power, f_i^2, at frequency index i, and the weighted sum (weighted by w_j) of all the frequencies power, f_j^2. σ is a term to ensure that G_i does not go to infinity. The weights, v_i and w_j, are trained in this Neurocompensator. The format of this example will produce a compensator that can apply level dependent gain, but not compression versus level, and ideally will produce some spectral contrast enhancement. The level dependent gain should produce a weighting that will show compression limiting.

The first step in training the Neurocompensator is a pre-processing stage where the time signal is compartmentalized into time-overlapped windowed samples. These windowed samples are filtered into twenty frequency subbands, corresponding to the model bands that will be combined in the error signal, and the power is taken in each band (f_k where $k = [1, 2, \ldots, 20]$). These are the statistics used as the input to the compensator model. A time series per frequency channel is derived, or G_i changes over time.

Each weight, G_i, is applied per time slice to the short-time Fourier transform and the inverse Fourier transform is taken. All the time-slices are assembled by overlapping and adding the processed windowed samples. The resulting time-domain waveform is the input to the damaged model. The input to the normal model can be thought of having G_i equals one over every frequency and every time-slice.

During the training phase, the v_i and w_j, gain coefficients are adapted to minimize the error metric summed over all the time slices. The parameters of the compensator are optimized so that the output of $\hat{H}$ matches the output of H as closely as possible. Once the compensator is trained, the gain coefficients are set and it becomes the final stage of processing in a digital hearing aid, replacing the fitting strategy.

The Alopex algorithm [35,5] was used to train the model weights via the error signal. Alopex is a stochastic optimization algorithm closely related to reinforcement learning and dynamic programming methods. It relies on the correlation between successive positive/negative weight changes and objective function changes from trial to trial to stochastically decide in which direction to move each weight.

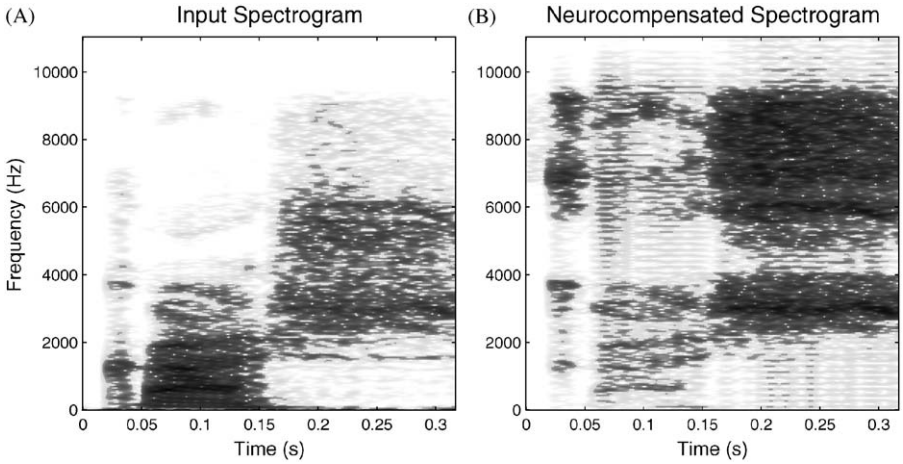

Fig. 14. The unprocessed input spectrogram (A) and the spectrogram of the signal that would be presented to the hearing impaired ear after Neurocompensation (B).

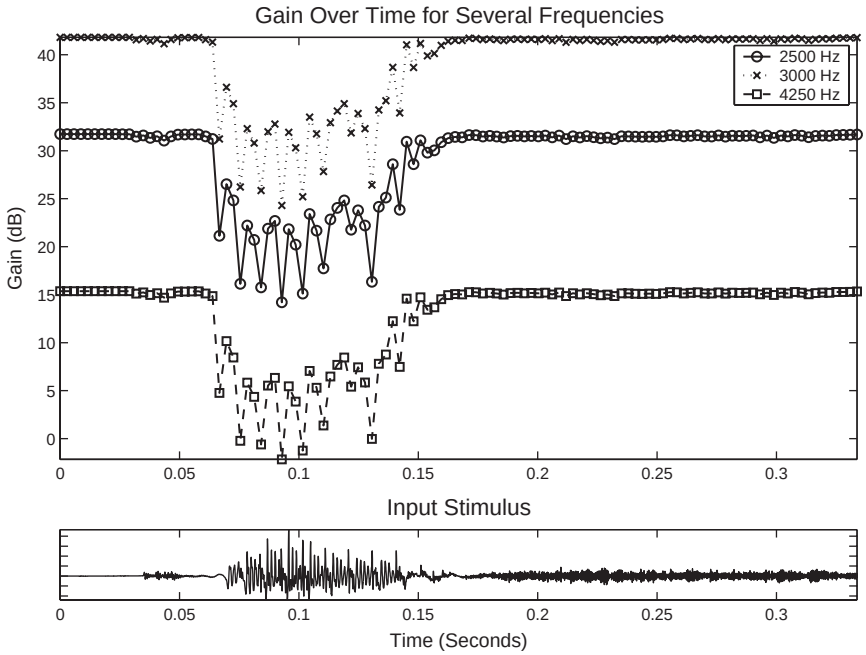

Fig. 15. Gain over time for several frequencies of the trial Neurocompensator.

Initial experiments were conducted with loss profile five. Instead of using LTASS noise, the dutch syllable "kas" from the van Son et al. [37] corpus was chosen. This was because of the difficult to compensate spectrum of the stop /*k*/ and fricative /*S*/ would give a more interesting example for contrast enhancement.

Fig. 13 shows the error signal plotted over iteration. Longer test results show that this error does not go to zero, meaning that better compensation models

are needed, or that full restoration of the neural representation is not possible.

The idea behind the original processing block was that it would provide spectral unmasking, or provide some contrast enhancement for the ear. This is plainly evident when one compares the settled algorithm's spectrogram in Fig. 14B to the input spectrogram in Fig. 14A.

The dynamic range of both spectrograms is 60 dB. The Neurocompensator has 35 dB more energy than the unprocessed input signal. Of special note is the second formant in the signal spectrum for the Neurocompensator. It shows evidence of compelling lateral inhibition that reduces the spectra above and below the formant. This would spread the response to the formant, and formant capture is very evident in the normal auditory system.

Similarly to Kates [22] the weights are dependent on the input stimulus, and should change over time mimicking the cochlea's cycle-to-cycle adaptive behaviour. At present, it is beyond the scope of the objective function to capture time adaptive and nonlinear, stimulus dependent effects. How the v_i and w_j change is a matter of future research.

This type of processing also introduces a gain dependent upon received level. An example of the weighting factors changing over time is given in Fig. 15.

Fig. 15 clearly shows an attempt to aid the transient response, or the '/*k*/' stop and the '/*S*/' fricative, and limit the voiced vowel '/*a*/'. This can be viewed as loudness equalization across time periods, but the present Neurocompensator trial does not have look-ahead or look-behind in time capabilities, so it should not be able to return proper time adaptive auditory processing that is lost with sensorineural hearing impairment. The conclusions give a few brief words on designing time adaptive compensation strategies.

6. Conclusion

Neural code compensation can be used to expedite development to the clinical stage. At present the Bruce et al. [9] model of the auditory periphery seems to be an excellent tool for offline evaluation of hearing-aid algorithms. It is able to reproduce a wide range of empirical results. There should be a few notes of caution, however: this model overpredicts synchrony of damaged fibers, and there is no data on how the model deals with neural coding of transients, or with masking in time. Future extensions to this work include derivation of an error metric that includes temporal information. Then, not only "steady-state" nonlinearities, such as compression ratios and compression thresholds, can be examined but also timing characteristics, such as compression attack and release times (compression ballistics).

For evaluation of future nonlinear hearing-aid algorithms there is a need to fully quantify suprathreshold effects [36] and derive hearing-aid signal processing based on understanding of the impairment. The Neurocompensator has a number of advantages over traditional approaches including the possibility of embedding suprathreshold knowledge in the design procedure.

Traditional hearing aids calculate gain on a frequency-by-frequency basis determined solely by the audiogram, without taking into account masking effects due to cross-frequency/cross-temporal interactions. Such methods work well for restoring the detection of pure tones but fail to correct for many of the masking and interference effects caused by the loss of hair cell nonlinear tuning. Another benefit of the Neurocompensator is the ability to create a fitting profile for individuals with the same audiogram, but different loss types. While more sophisticated measurements are needed to determine the percent impairment to the OHCs and IHCs, the differences in intelligibility between people with the same audiograms may be accounted for, unlike present strategies [17]. We used a simple 60% OHC impairment rule to fit hair cell loss to the audiograms, but specific determination of hair cell damage may alter or improve these results.

At this point, the design of compensation strategies based on restoring normal neural firing patterns seems to be a general strategy that can subsume historical empirical and recent physiological efforts. The idea that hearing-aid algorithm design can be viewed as neural compensation may open the door to novel concepts in assessing hearing impairment and scrutinizing hearing aid algorithms. To make this strategy complete some understanding of the true computational strategy of the auditory system is necessary. While there have been a few attempts at formulating an intelligibility metric derived from the auditory neural code [8], we are still

a long way from understanding the coding strategies employed by given populations of neurons [31]. Thus, the Neurocompensation strategy can be best used for comparative prediction of new hearing-aid algorithms, as well as being a very useful tool in their design. The Neurocompensator proposed here has the capability, in principle, to restore a number of the filtering capabilities intrinsic to the cochlear hair cells.

Acknowledgements

This research was supported by a CRO grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada and is associated with the Blind Source Separation and Application (BLISS) project.

References

[1] D.V. Anderson, Model based development of a hearing aid, Master's Thesis, Brigham Young University, Provo, Utah, April 1994.
[2] D.V. Anderson, R.W. Harris, D. Chabries, Evaluation of a hearing compensation algorithm, IEEE ASSP-95 (May 1995) 3531–3533.
[3] ANSI, ANSI S3.5-1997, Methods for Calculation of the Speech Intelligibility Index, American National Standards Institute, New York, 1997.
[4] S. Becker, I.C. Bruce, Neural coding in the auditory periphery: insights from physiology and modeling lead to a novel hearing compensation algorithm, in: Workshop on Neural Information Coding, Les Houches France, March 2002.
[5] A. Bia, Alopex-b: a new, simpler but yet faster version of the alopex training algorithm (Special Issue on Non-gradient Optimization Methods), Internat. J. Neural Systems (2001) 497–507.
[6] J. Bondy, I.C. Bruce, S. Becker, S. Haykin, Predicting speech intelligibility from a population of neurons, in: S. Thrun, L. Saul, B. Schoelkopf (Eds.), Advances in Neural Information Processing Systems, Vol. 16, MIT Press, Cambridge, MA, 2004.
[7] D. Brooks, Gain requirements of hearing aid users, Scand. Audiol. 2 (1973) 199.
[8] I.C. Bruce, J. Bondy, S. Haykin, S. Becker, A physiologically based predictor of speech intelligibility, in: International Conference on Hearing Aid Research, Lake Tahoe, CA, August 2002.
[9] I.C. Bruce, M.B. Sachs, E.D. Young, An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses, J. Acoust. Soc. Amer. 113 (1) (2003) 369–388.
[10] I.C. Bruce, E. Young, M. Sachs, Modification of an auditory periphery model to describe the effects of acoustic trauma on auditory nerve response, in: 22nd Annual ARO Mid-winter Meeting, St. Petersburg Beach, FL, 1999.
[11] D. Byrne, H. Dillon, The national acoustic laboratories' (nal) new procedure for selecting the gain and frequency response of a hearing aid, Ear Hearing 7 (1986) 257–265.
[12] D. Byrne, H. Dillon, T. Ching, R. Katsch, G. Keisder, Nal-nl1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures, J. Amer. Acad. Audiol. 12 (2001) 37–51.
[13] D. Byrne, H. Dillon, K. Tran, S. Arlinger, K. Wilbraham, R. Cox, B. Hagerman, R. Hetu, J. Kei, C. Lui, J. Kiessling, M. Kotby, N. Nasser, E. LePage, W. Kholy, Y. Nakanishi, H. Oyer, R. Powell, D. Stephens, R. Meredith, T. Sirimanna, G. Tavartkiladze, G. Frolenkov, S. Westerman, C. Ludvidsen, An international comparison of long-term average speech spectra, J. Acoust. Soc. Amer. 96 (1994) 2108–2120.
[14] D. Byrne, D. Fifield, Evaluation of hearing aid fittings for infants, Br. J. Audiol. 8 (1974) 47–54.
[15] D. Byrne, A. Parkinson, P. Newall, Hearing aid gain and frequency response requirements for the severely/profoundly hearing impaired, Ear Hearing 11 (1990) 40–49.
[16] D. Chabries, D. Anderson, T. Stockham, R. Christiansen, Application of a human auditory model to lousness compensation and hearing compensation, IEEE ASSP-95 (May 1995) 3527–3530.
[17] T. Ching, H. Dillon, D. Byrne, Speech recognition of hearing impaired listeners: predictions from audibility and the limited role of high-frequency amplification, J. Acoust. Soc. Amer. 103 (2) (1998) 1128–1140.
[18] J. Eggermont, H. Komiya, Moderate noise trauma in juvenile cats results in profound cortical topographic map changes in adulthood, Hearing Res. 142 (2000) 89–101.
[19] D. Fabry, D. van Tassell, Evaluation of an articulation-index based model for predicting the effects of adaptive frequency response hearing aids, J. Speech Hearing Res. 33 (1990) 676–689.
[20] M. Heinz, X. Zhang, I.C. Bruce, L. Carney, Auditory nerve model for predicting performance limits of normal and impaired listeners, Acoust. Res. Lett. Online 2 (3) (2001) 91–96.
[21] F. Julicher, D. Andor, T. Duke, The physical basis of two-tone interference in hearing, Proc. Natl. Acad. Sci. USA 98 (2001) 9080–9085.
[22] J. Kates, Toward a theory of optimal hearing aid processing, J. Rehab. Res. 30 (1) (1993) 39–48.
[23] M. Liberman, L. Dodds, Single-neuron labelling and chronic cochlear pathology. iii. Stereocilia damage and alterations of threshold tuning curves, Hearing Res. 16 (1984) 55–74.
[24] D. Markle, A. Zaner, The determination of 'gain requirements' of hearing aids: a new method, J. Audiol. Res. 6 (1966) 371–377.
[25] F. Martin, C. Champlin, J. Chambers, Seventh survey of audiometric practices in the United States, J. Amer. Acad. Audiol. 9 (1998) 95–104.
[26] B. Moore, Perceptual Consequences of Cochlear Damage, Oxford University Press, Oxford, 1995.
[27] B. Moore, B. Glasberg, Comparison of frequency selectivity in simultaneous and forward masking for subjects with

unilateral cochlear impairments, J. Acoust. Soc. Amer. 80 (1986) 93–107.
[28] B. Moore, M. Huss, D. Vickers, B. Glasberg, J. Alcantara, A test for the diagnosis of dead regions in the cochlea, Br. J. Audiol. 34 (4) (2000) 205–224.
[29] R. Nobili, F. Mammano, J. Ashmore, How well do we understand the cochlea? Trends Neurosci. 21 (1998) 159–167.
[30] C. Rankovic, An application of the articulation index to hearing aid fitting, J. Speech Hearing Res. 34 (1991) 391–402.
[31] F. Rieke, D. Warland, R. de Ruyter van Steveninck, W. Bialek, Spikes: Exploring the Neural Code, MIT Press, Cambridge, MA, 1997.
[32] O. Schwartz, E. Simoncelli, Natural sound statistics and divisive normalization in the auditory system, in: T. Leen, T. Dietterich, V. Tresp (Eds.), Advances in Neural Information Processing Systems, Vol. 13, MIT Press, Cambridge, MA, 2001, pp. 166–172.
[33] H. Steeneken, On measuring and predicting speech intelligibility, Ph.D. Thesis, University of Amsterdam, 1992.
[34] H. Steeneken, T. Houtgast, A physical method for measuring speech-transmission quality, J. Acoust. Soc. Amer. 67 (1980) 318–326.
[35] K. Unnikrishnan, K. Venugopal, Alopex: a correlation-based learning algorithm for feedforward and recurrent neural networks, Neural Comput. 6 (3) (May 1994) 469–490.
[36] N. van Schijndel, T. Houtgast, J. Festen, Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal hearing and hearing-impaired listeners, J. Acoust. Soc. Amer. 110 (1) (2001) 529–542.
[37] R. van Son, D. Binnenpoorte, H. van den Heuvel, L. Pols, The ifa corpus: a phonemically segmented dutch "open source" speech database, in: Proceedings of the Eurospeech 2001, Aalborg Denmark, September 3–7, 2001.
[38] F. Wiener, D. Ross, The pressure distribution in the auditory canal of a progressive sound field, J. Acoust. Soc. Amer. 18 (2) (1946) 401–408.
[39] X. Zhang, M. Heinz, I.C. Bruce, L. Carney, A phenomenological model for the responses of auditory-nerve fibers: I, Nonlinear tuning with compression and suppression, J. Acoust. Soc. Amer. 109 (2) (2001) 648–670.