Insights into optimal phonemic compression from a computational model of the auditory periphery

Ian C. Bruce[1,2], Faheem Dinath[2], and Timothy J. Zeyl[1]

[1] Department of Electrical & Computer Engineering and [2] School of Biomedical Engineering, McMaster University, Hamilton, ON, L8S 4K1, Canada

ABSTRACT

Phonemic compression schemes for hearing aids have thus far been developed and evaluated based on perceptual criteria such as speech intelligibility, sound comfort, and loudness equalization. Finding compression parameters that optimize all of these perceptual metrics has proved difficult. The goal of this study was to find optimal single-band gain adjustments based on the response of auditory-nerve fibers to speech. Sentences from the TIMIT database were processed by either the NAL-R or the DSL amplification scheme, and deviations from these linear prescriptions were obtained by adjusting the overall gain from 40 dB below to 40 dB above the prescribed gains in 5 dB steps. Neural responses were obtained using the cat auditory-periphery model of Zilany and Bruce (2006, 2007). Sentences were analyzed on a phone by phone basis to find the gain adjustment that minimized the difference in neural response to the amplified phone in the impaired model and the unamplified phone in the normal model. The optimal gain adjustments were found to depend on whether the error metric included the spike timing information of the neural responses (i.e., a time resolution of several microseconds) or just the mean firing rates (i.e., a time resolution of several milliseconds). To optimize the mean firing rates, gain adjustments on the order of +10 dB were required *above* the prescribed linear gains in general. In contrast, gain adjustments on the order of −10 dB or more *below* the prescribed linear gains tended to optimize the responses including spike timing information. Wide dynamic range compression appears to be more beneficial in optimizing the spike timing information than the mean rate information. These results motivate the development of novel nonlinear amplification schemes that simultaneously optimize both spike-timing and mean-rate neural representations.

INTRODUCTION

Early observations in hearing aid fitting showed that the preferred gain at a particular frequency equalled approximately half the hearing threshold shift at the same frequency. This is referred to as the "half-gain" rule (Dillon 2001). Popular linear hearing aid prescriptions, including the NAL-R (National Acoustic Laboratories) and DSL (Desired Sensation Level) prescriptions, are variations from the half-gain rule based on judgments of speech intelligibility, sound comfort, and loudness equalization (Dillon 2001). Compression schemes were later introduced in hearing aids to counteract the effects of abnormal growth of loudness in impaired ears (see Moore 2004 for a recent review). Nonlinear versions of the gain prescriptions were consequently developed, such as NAL-NL1 and DSL[i/o] (Byrne *et al.* 2001; Scollie *et al.* 2005).

Compression characteristics such as compression thresholds, compression ratios, and attack and release times can be adjusted to achieve goals such as avoiding distorted or uncomfortably loud signals, reducing the intensity differences between phonemes or syllables, providing automatic volume control, increasing sound comfort, normalizing loudness, maximizing intelligibility, or reducing background noise (Dillon 2001). However, the compression scheme parameters that

are required to obtain each of these results are often quite different (Dillon 2001). Thus, it appears that standard compression schemes apply suboptimal gain adjustments. The goal of this study was to find optimal single-band gain adjustments around the NAL-R and DSL prescribed gains by using the neural representation of speech rather than perceptual feedback. We focussed on single-band, rather than multi-band, compression to allow for a simple gain adjustment optimization scheme and also to simplify the interpretation of the results. The simulation results: a) provide a physiological explanation of why standard compression schemes are suboptimal, and b) motivate the development of nonlinear amplification algorithms that better compensate for the physiological effects of cochlear impairment.

METHODS

Model

The auditory-periphery model used in this study was the cat auditory nerve (AN) model developed by Zilany and Bruce (2006, 2007), which describes the auditory pathway from the middle ear through to the AN. A schematic of the model is given in Fig. 1.

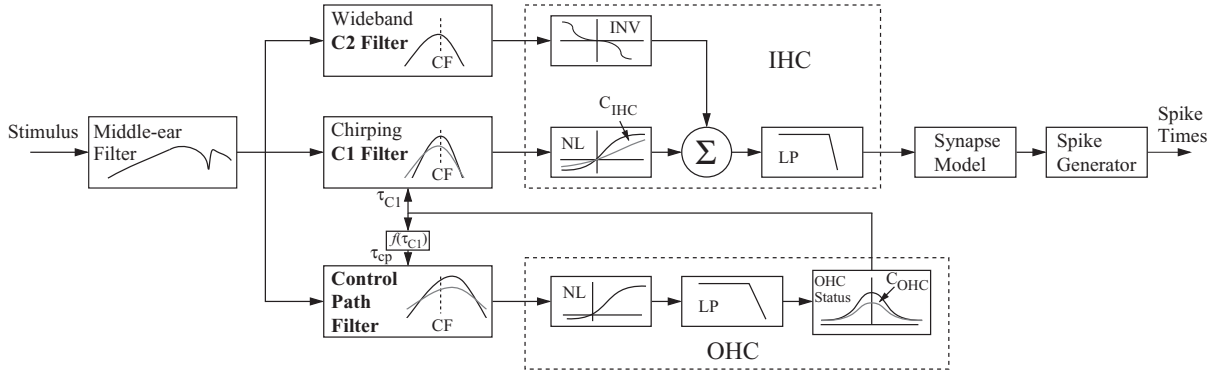

Fig 1: Schematic of the auditory periphery model of Zilany and Bruce (2006, 2007). Reprinted with permission of the Acoustical Society of America © 2006.

Speech waveforms, sampled at 500 kHz and with instantaneous pressures in units of Pascal, are delivered to the model to derive an AN spike train for a fiber with a specific characteristic frequency (CF). In this study, the filtering effect of the outer ear (the real ear unaided gain or REUG) was modelled after a human head-related transfer function described by Wiener and Ross (1946).

Simulating hearing loss and prescribing gain

Model parameters, C_{IHC} and C_{OHC}, which control the level of inner hair cell (IHC) and outer hair cell (OHC) impairment respectively, can be adjusted to provide a desired hearing threshold shift at a specific CF. A C_{IHC} or C_{OHC} of 0 produces full impairment whereas a value of 1 provides normal function. Two hearing loss profiles were simulated in this study, as illustrated in Fig. 2. In each case, it was assumed that two-thirds of the threshold shift at each frequency was attributable to OHC impairment and the remaining third to IHC impairment, and C_{IHC} and C_{OHC} were adjusted to produce these threshold shifts.

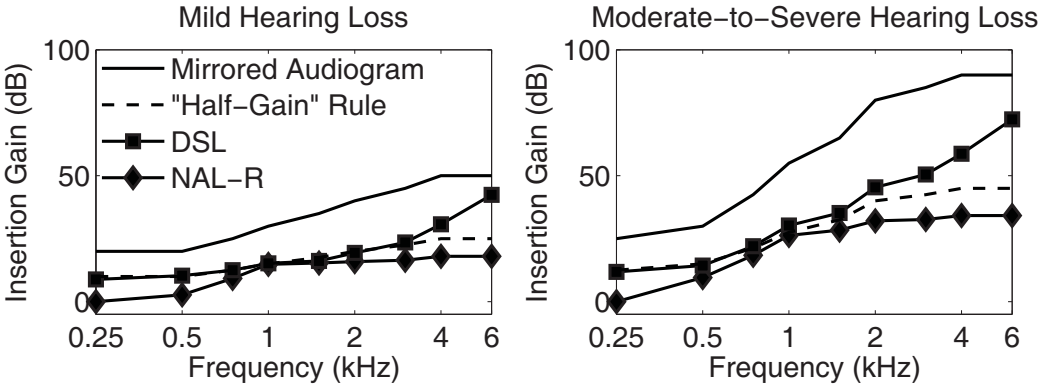

Fig 2: The two example hearing loss profiles and corresponding insertion gains used in this study. Left panel: a mild high-frequency hearing loss; Right panel: a moderate-to-severe high-frequency hearing loss. The mirrored audiogram is shown by the solid line, and its half gain as the dotted line. Insertion gains prescribed by DSL are shown in by the squares and those for NAL-R are shown by the diamonds.

The goal of the NAL-R linear amplification prescription is to maximize speech intelligibility for moderate sound pressure levels (SPLs) by equalizing perceived loudness over the frequency range important for speech (250–8000 Hz). Gain prescribed by the NAL-R formula (Byrne and Dillon 1986) is in terms of insertion gain (IG), that is, the gain provided by the hearing aid above REUG. IGs for the two cases of hearing loss are shown in Fig. 2. In this study, it was assumed that the simulated hearing aid perfectly reproduced the natural gain of the unaided ear, such that the total gain provided was the REUG used for simulations of the normal ear plus the IG prescribed by NAL-R.

The DSL prescription differs from the NAL-R procedure in that it does not try to make speech equally loud, but rather comfortably loud. Although first developed for use in pediatric audiology, ongoing research and modifications have expanded the role of DSL for use with adults (Scollie *et al.* 2005). Gain prescribed by DSL is expressed in terms of the real ear aided gain (REAG), that is, the total gain supplied by the hearing aid. In this study, the REAGs for the DSL simulations were calculated from the table of values on page 243 of Dillon (2001). Shown in Fig. 2 are the DSL IGs for the two cases of hearing loss; the IGs for DSL were calculated by subtracting the model REUG from the DSL-prescribed REAGs.

Stimuli

Speech recordings were taken from the TIMIT database. TIMIT is a corpus consisting of 450 phonetically-compact and 1890 phonetically-diverse read English sentences. Two TIMIT sentences were used in the simulation of the mild hearing loss case and two other sentences were used for the moderate-to-severe case. For good SPL coverage, three different presentation levels (45, 65 and 85 dB SPL) were tested for each sentence.

Analysis of neural responses

The response of the AN to acoustic stimuli is quantified in this study by a "neurogram". A neurogram is similar to a spectrogram, except that it displays the neural response as a function of CF and time. We used 30 CFs spaced logarithmically between 250 and 8000 Hz. The neural response at each CF is composed of responses from 50 AN fibers. In accordance with Liberman (1978), 60% of fibers were chosen to be high spontaneous rate fibers (>18 spikes/s), 20%

medium (0.5 to 18 spikes/s), and 20% low (<0.5 spikes/s). The method for adjusting the spontaneous rate (and the resulting rate-level curve) in the model is described in chapter 5 of Zilany (2007).

The neurogram can include the spike timing information of the neural responses by maintaining a small time bin size (Fig. 3D). In this instance, a bin size of $2\,\mu s$ was utilized, and responses were smoothed by convolving them with a Hamming window 128 samples ($256\,\mu s$) in length. Alternatively, spike timing information can be excluded by computing the moving average of the neural response with a window of several milliseconds to give only the average discharge rate as a function of time (Fig. 3C). In this instance, a bin size of $62.5\,\mu s$ was utilized, and again responses were smoothed by convolving them with a Hamming window 128 samples ($8\,ms$) in length.

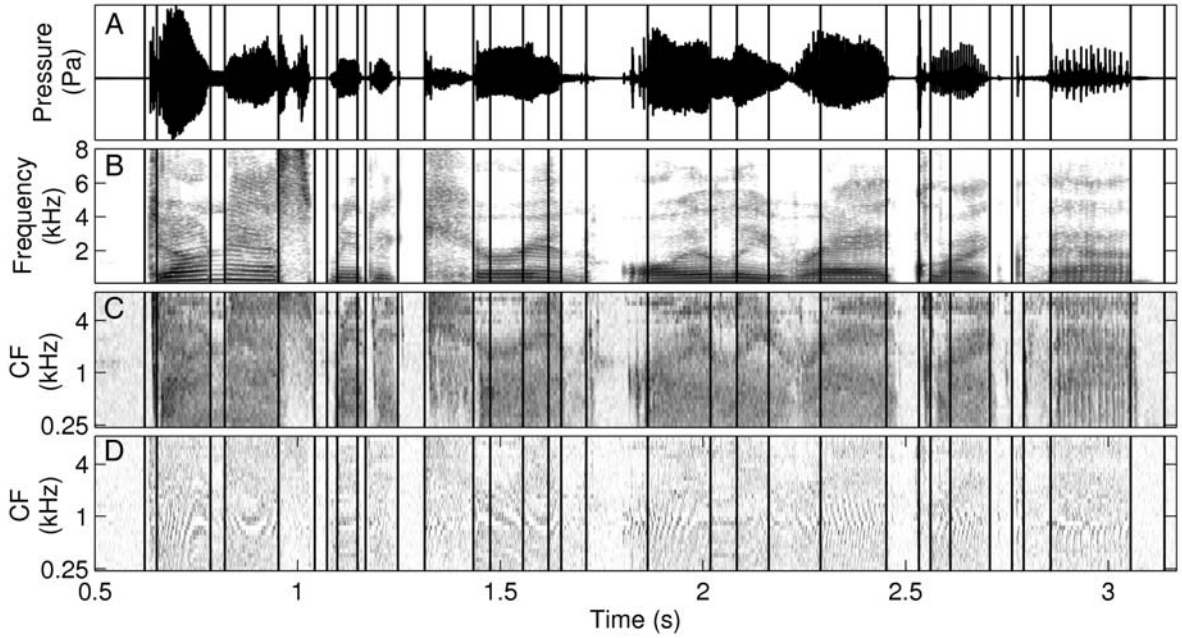

Fig 3: An example sentence from the TIMIT database and the corresponding spectrogram and neurograms. (A) Time-domain pressure waveform; (B) Spectrogram; (C) Neurogram based on the average discharge rate; (D) Neurogram based on the spiking timing information. Phone boundaries are indicated by the thin vertical lines.

Gain optimization strategy

Optimal single-band gain adjustments around the hearing aid prescription gains were obtained though the gain adjustment strategy shown in Fig. 4 below. The gain adjustment strategy compares neural responses to speech sentences on a phone-by-phone basis for the impaired and normal models. To avoid the complicating and confounding effects of attack and release characteristics, the known phone boundaries from the TIMIT sentences are used to apply a constant gain adjustment for the duration of each phone.

The strategy begins by passing the first phone through the normal model to derive the normal neurogram. In the impaired pathway, the phone is passed though a filter implementing either the NAL-R or DSL amplification prescription (see Fig. 2) before a single-band gain adjustment is

applied. Gain adjustments range from –40 to +40 dB in 5-dB increments resulting in 17 uniquely amplified phones. The phones are passed through the impaired model, producing a set of 17 neurograms. The gain adjustment that minimizes the mean absolute error between the normal and impaired neurograms is deemed the optimal gain adjustment for that phone.

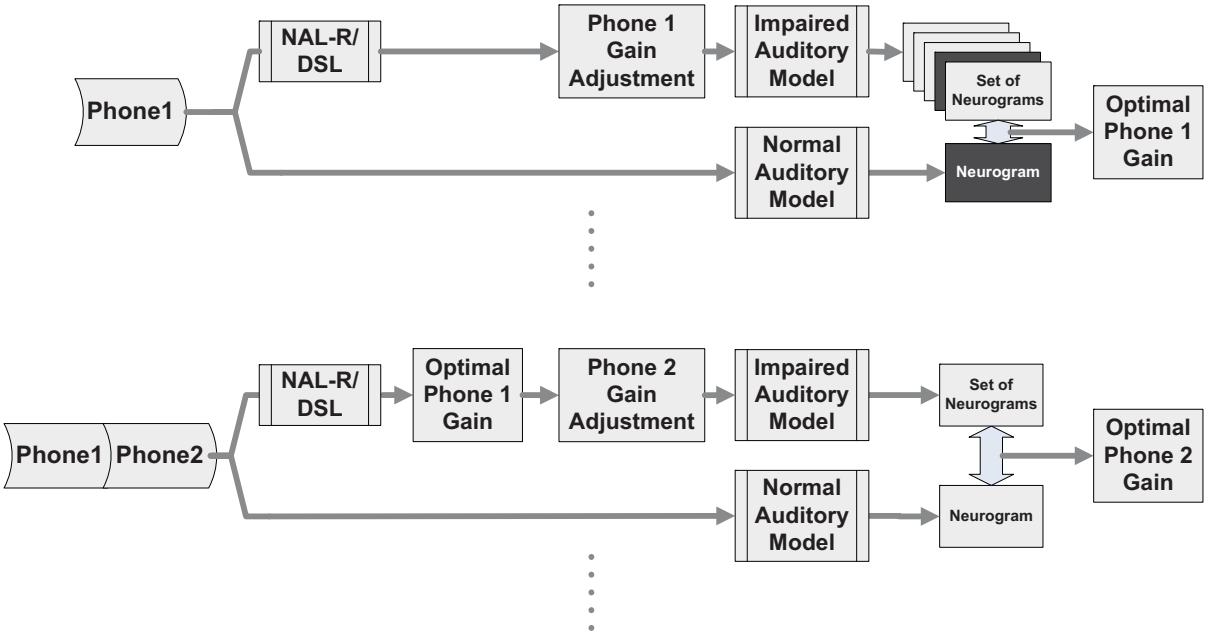

Fig 4: Flow diagram of gain adjustment strategy.

The second and all subsequent phones are analyzed in the same manner as the first, however, due to adaptation in the auditory-periphery model, all prior phones are prepended. The range of gain adjustments is applied only to the current phone and all previous phones are amplified with their optimal gain adjustments.

RESULTS

For each amplification prescription, optimal gain adjustments were found for both the neurograms with spike timing information and the average discharge rate neurograms. Gain-optimization results are shown in Fig. 5 for the case of mild hearing loss and in Fig. 6 for the moderate-to-severe loss. In each figure: the left column shows the results for neurograms based on the average discharge rate and the right column for the spike timing neurograms; the top row shows the results for the DSL amplification prescription and the bottom row for the NAL-R prescription; and each symbol indicates the optimal gain adjustment for an individual phone as a function of the input phone SPL.

For the case of mild hearing loss (Fig. 5), gains *above* the NAL-R and DSL prescribed gains tend to optimize the mean discharge rate at lower SPLs, and only at phone levels around 80 dB SPL is compression required. In contrast, gains *below* the prescribed gains tend to optimize the spike timing information at lower SPLs, and a compression ratio of around 2:1 is indicated for the entire range of phone SPLs. Optimal gain adjustments for the NAL-R prescription are somewhat higher than those for DSL, consistent with NAL-R's generally-lower insertion gains (see Fig. 2).

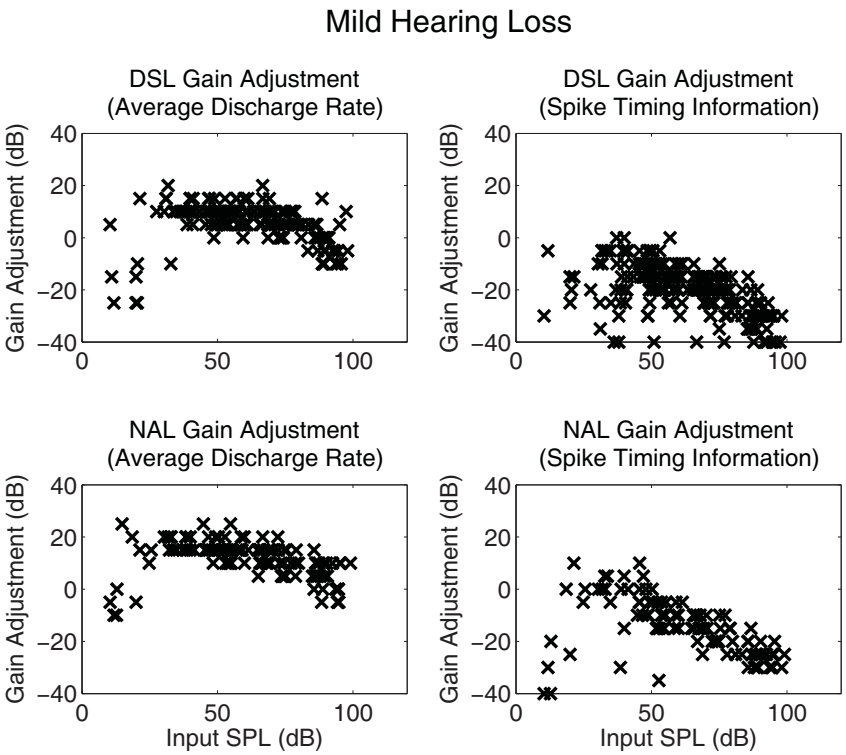

Fig 5: Optimal gain adjustments versus phone input sound pressure for the case of mild hearing loss. Each symbol indicates the optimal gain adjustment for an individual phone.

Similar results were obtained for the case of moderate-to-severe hearing loss (Fig. 6). However, in this case the DSL prescription appears to be biased more towards optimizing the mean discharge rate rather than the spike timing information, whereas the NAL-R tends to retain more of a balance between optimizing the two forms of neural coding.

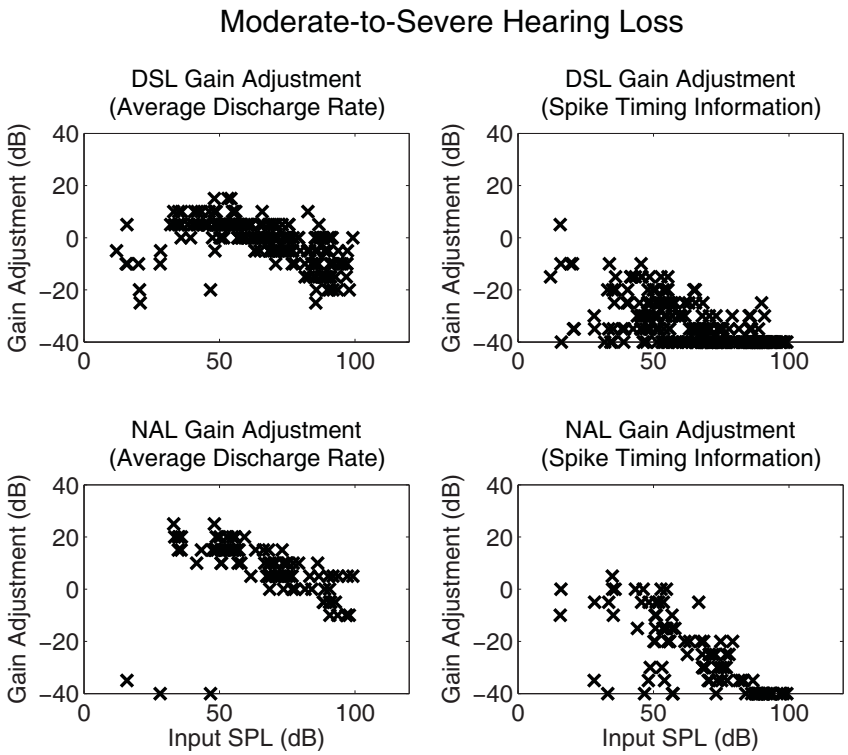

Fig 6: Optimal gain adjustments versus phone input sound pressure for the case of moderate-to-severe hearing loss. Each symbol indicates the optimal gain adjustment for an individual phone.

CONCLUSIONS AND FUTURE WORKS

The results indicate that the NAL-R and DSL amplification schemes tend to find a balance between optimizing the spike timing information and average discharge rate information. The optimal gain adjustments for NAL-R were generally higher than those for DSL, consistent with the lower insertion gains of the NAL-R prescription relative to DSL (see Fig. 2). Wide dynamic range compression appears to be required to optimize the spike timing information more so than the average discharge rate information.

It was found that positive gain adjustments *above* the prescribed linear gains better restored the mean discharge rate representation of speech. This is consistent with the physiological data of Heinz and Young (2004), where they found that on average there is no steepening of AN fiber rate-level curves with hearing impairment. Consequently, their data would argue for hearing aid gains closer to mirroring the audiogram to restore mean discharge rates.

Further investigation of the model responses is required to understand why gain adjustments *below* the prescribed gains better restore the spike timing information in the neural representation of phonemes. Contributing factors could include the spread of synchrony and the change in phase-frequency responses in an impaired ear. Physiological experiments and modelling studies (Miller *et al.* 1997; Zilany and Bruce 2007) have shown that the normal tonotopic representation of vowels is lost in an impaired ear, and instead large populations of AN fibers synchronize to a range of vowel components, particularly the lower-frequency, higher-energy formants. Additionally, Carney (1994) has postulated that the flattening of an AN fiber's phase-frequency response as a result of hearing impairment could be a neural correlate of loudness recruitment.

It appears that linear amplification schemes or standard single-band compression schemes cannot simultaneously optimize both the spike timing information and the average discharge rate information in the neural response to speech. This motivates: a) studies to further understand why the spike timing and mean discharge rate information are optimized at different levels of gain, and b) development of alternative nonlinear amplification strategies to produce simultaneous optimization of both forms of neural coding.

ACKNOWLEDGMENTS

The authors thank M. S. A. Zilany for computer programming and technical assistance. This work was funded by NSERC Discovery Grant 261736 and the Barber-Gennum Chair Endowment (I. C. Bruce) and an NSERC Undergraduate Summer Research Award (T. J. Zeyl).

REFERENCES

Byrne, D. and Dillon, H. (**1986**). "The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid," Ear & Hearing, **7**, 257–265.

Byrne, D., Dillon, H., Ching, T., Katsch, R. and Keidser, G. (**2001**). "The NAL-NL1 procedure for fitting non-linear hearing aids: Characteristics and comparisons with other procedures," J. Am. Acad. Audiol., **12**, 37–51.

Carney, L. H. (**1994)**. "Spatiotemporal encoding of sound level: models for normal encoding and recruitment of loudness," Hear. Res., **76**, 31–44.

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

Heinz, M. G. and Young, E. D. (**2004)**. "Response growth with sound level in auditory-nerve fibers after noise-induced hearing loss," J. Neurophysiol., **91**, 784–795.

Liberman, M. C. (**1978**). "Auditory nerve response from cats raised in a low noise chamber," J. Acoust. Soc. Am., **63**, 442–455.

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**, 3602–3616.

Moore, B. C. J. (**2004**). "Testing the concept of softness imperception: Loudness near threshold for hearing-impaired ears," J. Acoust. Soc. Am., **115**, 3103–3111.

Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., Beaulac, S. and Pumford, J. (**2005**). "The desired sensation level multistage input/output algorithm," Trends in Amplification, **9**, 159–197.

Wiener, F. M. and Ross, D. A. (**1946**). "The pressure distribution in the auditory canal in a progressive sound field," J. Acoust. Soc. Am., **18**, 401–408.

Zilany, M. S. A. (**2007**). *Modeling the Neural Representation of Speech in Normal Hearing and Hearing Impaired Listeners*. Ph.D. Thesis, McMaster University, Hamilton, ON.

Zilany, M. S. A. and Bruce, I. C. (**2006**). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," J. Acoust. Soc. Am., **120**, 1446–1466.

Zilany, M. S. A. and Bruce, I. C. (**2007**). "Representation of the vowel /ε/ in normal and impaired auditory nerve fibers: Model predictions of responses in cats," J. Acoust. Soc. Am., **122**, 402–417.