Modelling the effects of cochlear impairment on the neural representation of speech in the auditory nerve and primary auditory cortex

Ian C. Bruce and Muhammad S. A. Zilany

Department of Electrical & Computer Engineering, McMaster University, Hamilton, ON, L8S 4K1, Canada

Accurate models of normal and impaired neural representations of sound are useful tools in understanding how acoustic stimuli are encoded in the brain, predicting speech intelligibility, and developing and testing speech processing schemes for hearing aids. In this paper we review recent developments in modelling the effects of hair cell impairment on neural responses to speech stimuli in the auditory nerve and primary auditory cortex. Several important cochlear nonlinearities, such as compression and suppression, the shift in tuning with sound pressure level, and the component-1/component-2 transition at very high sound pressure levels, have been incorporated into the latest models of the auditory periphery. These properties of cochlear processing prove to be important not only in forming the normal neural representation of sound but also in determining the degradation of the neural representation in cases of hair cell impairment. We have evaluated these models by using them to predict the effects of presentation level, hearing loss and amplification on speech intelligibility. The models are able to predict both the effects of audibility on speech intelligibility and the "roll over" in speech intelligibility at high presentation levels for normal hearing listeners and for hearing impaired listeners using hearing aids.

INTRODUCTION

Physiological investigations of the effects of cochlear impairment on the neural representation of speech sounds provide important insights into the effects of hearing loss in humans and the action of different hearing-aid amplification schemes in compensating for hearing loss (see Sachs et al. 2002 for a review). However, the practical limitations of physiological experiments with animal models unfortunately restrict the scope of such investigations. This motivates the development of physiologically-accurate computational models of the auditory system, such that greater ranges of acoustic stimuli, hearing loss profiles and hearing aid algorithms can be explored (Bruce et al. 2003; Bruce 2004; Bondy et al. 2004; Zilany and Bruce 2006, 2007a,b). In this paper we a) review the recent development of models of the auditory periphery and cortex and b) illustrate how impairment of outer hair cells (OHCs) and inner hair cells (IHCs) in the peripheral model effects the neural representation of speech sounds in the auditory nerve and the cortex.

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.). ISBN: 87-990013-1-4. Print: Centertryk A/S.

AUDITORY-PERIPHERY MODEL

Figure 1 shows a schematic of the auditory-periphery model of Zilany and Bruce (2006, 2007b). This model aims to reproduce response properties of cat AN fibers in normal and acoustically-traumatized ears. It incorporates features from the series of models developed by Carney and colleagues (Carney 1993; Zhang et al. 2001; Tan and Carney 2003) and from the model of Bruce et al. (2003) and adds a second mode of cochlear filtering and IHC transduction to account for the "component 2" (C2) response properties observed in AN fibers at high sound pressure levels (Kiang 1990; Wong et al. 1998).

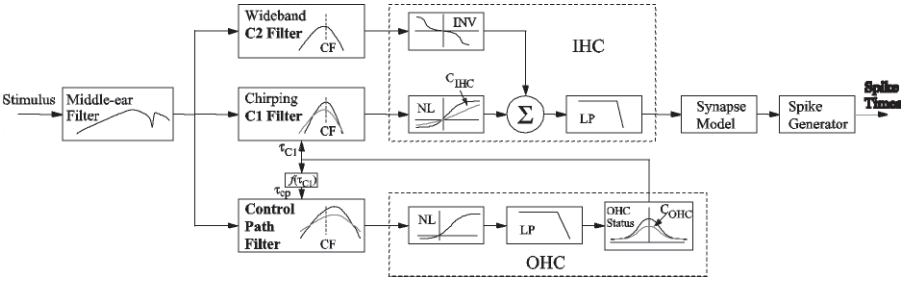

Fig. 1: Schematic of the auditory periphery model of Zilany and Bruce (2006, 2007b). Reprinted with permission of the Acoustical Society of America © 2006.

Damage to the OHCs is modelled as impairment of the cochlear amplifier (CA) control of the primary or "component 1" (C1) mode of cochlear filtering. Consequently, the tip of an AN fiber tuning curve becomes progressively elevated and broadened (illustrated in Fig. 2) and the AN response properties become more linear (i.e., compression and suppression are reduced) with increasing impairment.

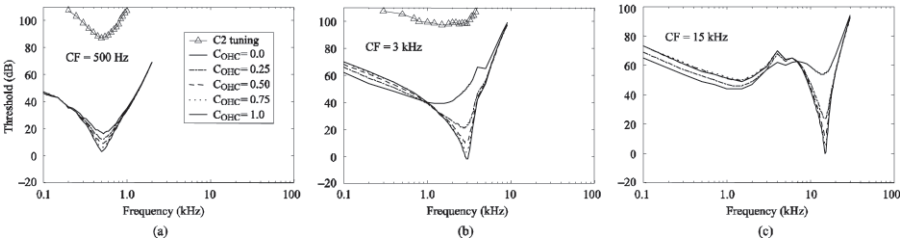

Fig. 2: Tuning curves for model fibers with CFs of (a) 500 Hz, (b) 3 kHz and (c) 15 kHz, for different levels of OHC impairment, ranging from normal function (COHC = 1.0) to complete dysfunction (COHC = 0.0). Also shown are the respective C1/C2 transition-threshold tuning curves. The transition threshold corresponds to a phase shift of 90°. Modified from Fig. 5 of Zilany and Bruce (2006) with permission of the Acoustical Society of America © 2006.

Damage to the IHCs is implemented by scaling down the input to the C1 IHC transduction function alone, because it has been observed that C2 responses are fairly impervious to damage in functioning AN fibers (Liberman and Kiang 1984). IHC impair-

ment leads to an elevation of the tuning curve without broadening. The effects of OHC and IHC impairment on AN fiber rate- and phase-level functions are illustrated in Fig. 3. For the normal model fiber shown in Fig. 3(a), a phase transition is observed at high levels as the dominant response switches from C1 to C2. During the transition, two peaks are observed in the period histogram, a phenomenon referred to as peak splitting (Kiang 1990). With moderate OHC impairment and normal IHC function, a threshold shift in the C1 response is observed in Fig. 3(b), but the strength of the C1 response above threshold is not diminished and peak splitting is preserved. In contrast, in Fig. 3(c) a weakening of the C1 response and elimination of peak splitting is observed in addition to a threshold shift for moderate IHC impairment and normal OHC function. With the elimination of peak splitting, a notch appears in the rate-level function, consistent with the physiological data (Kiang 1990). For substantial OHC and IHC impairment, only a very weak C1 response is observed and the C2 response dominates—see Fig. 3(d).

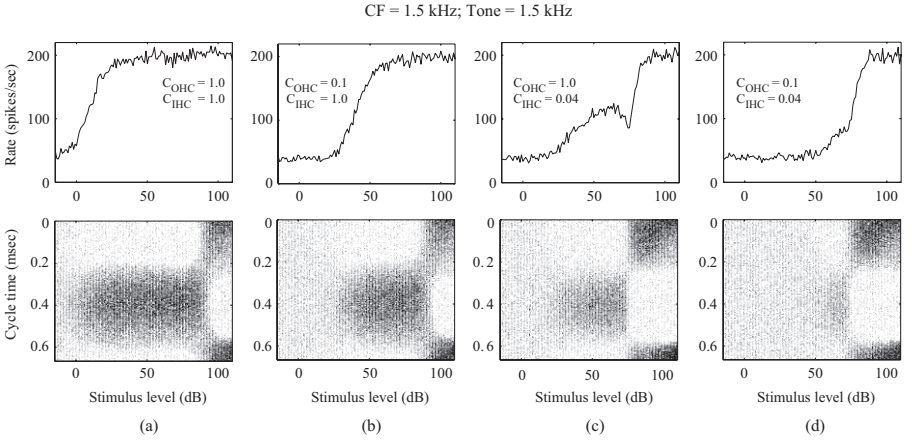

Fig. 3: Spike rate (top row) and phase (bottom) versus stimulus level for different degrees of OHC and IHC impairment. Reprinted from Zilany and Bruce (2006) with permission of the Acoustical Society of America © 2006.

We tested the model against AN fiber data for the synthetic vowel /ɛ/ and showed that it predicts the effects of presentation level and hair cell impairment on the vowel representation (Zilany and Bruce 2007b). In particular, the model provides an explanation as to why the formants of the vowel appear to undergo the C1/C2 transition simultaneously (Wong et al. 1998), which we found rules out several alternative model architectures. Given the model's accuracy in predicting AN data for a range of stimuli including vowels, we were interested to see how the model describes the response of AN fibers to running speech and if it could predict the effects of presentation level and hearing loss on speech intelligibility.

Figure 4 shows the model AN response to the word "doll" taken from the TIMIT recorded speech database. The response is plotted as a "neurogram" of the discharge

rates averaged over 8 ms as a function of time for 128 AN fibers with CFs ranging from 0.18 to 7.04 kHz spaced logarithmically. The output at each CF represents the mean discharge rate of a mix of high-, medium- and low-spontaneous-rate fibers. In the normal model at conversational speech levels (top-left panel of Fig. 4), the formant structure is clearly visible in the AN neurogram. The response of the normal model to speech presented above conversational speech levels (bottom-left panel) exhibits degradation in the representation of the formants because of saturation in discharge rates and spread of excitation. The impaired model shows a degraded formant representation at conversational levels (top-right panel)—the higher-frequency components appear inaudible. Increasing the presentation level (bottom-right panel) produces a stronger response to the lower-frequency speech components but does not restore the representation of the higher-frequency components.

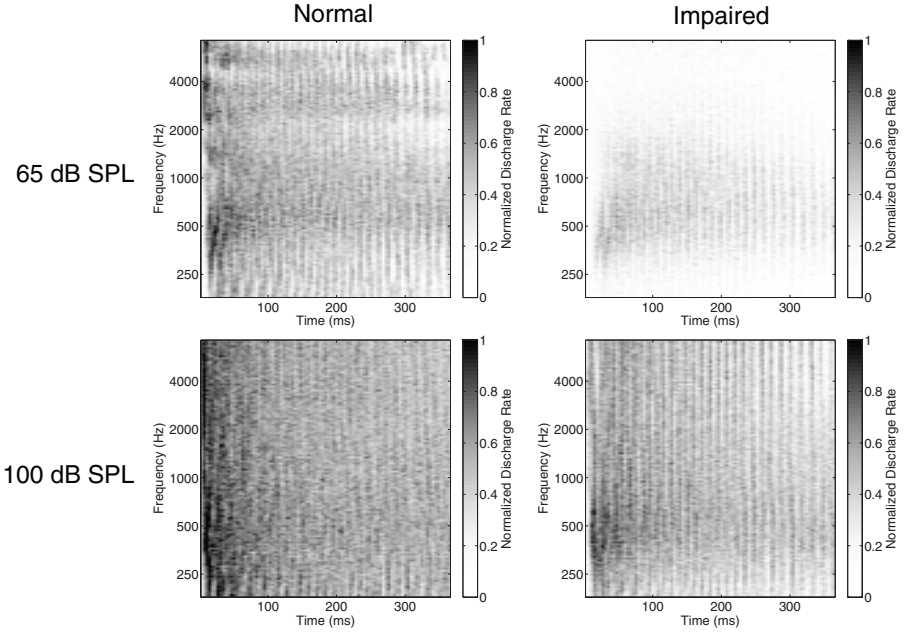

Fig. 4: Model AN neurograms in response to the word "doll". Left column: Normal model response; Right column: Impaired model response, for a gently sloping mild hearing loss; Top row: 65 dB SPL presentation level; Bottom row: 100 dB SPL presentation level.

The AN neurogram results suggest that: a) the AN representation of speech is degraded at presentation levels above those of normal conversational speech, b) cochlear impairment also distorts the AN representation of speech features, and c) louder speech does not compensate for cochlear impairment. In order to quantify these effects to obtain direct predictions of speech intelligibility, we have utilized a cortical model of speech processing.

CORTICAL MODEL

Many cells in the primary auditory cortex (A1) respond preferentially to specific spectro-temporal modulation patterns, such that they can be characterized by their spectro-temporal response fields (STRFs; e.g., Depireux et al. 2001). Elhilali et al. (2003) proposed that an array of cortical spectro-temporal modulation filters may analyse the modulations in the peripheral neural response as a function of CF and time to extract speech features. Elhilali and colleagues' framework for utilizing such a modulation filterbank to derive a speech intelligibility predictor, the spectro-temporal modulation index (STMI), is illustrated in Fig. 5. They found that the STMI is able to predict the effects of background noise, reverberation and phase-jitter on speech intelligibility. However, the auditory-periphery model of Elhilali et al. (2003) is very simple and does not include many of the cochlear nonlinearities that have a strong effect on the AN representation of speech as a function of level and hair-cell impairment (e.g., Miller et al. 1997; Wong et al. 1998). In Zilany and Bruce (2007a), we combined our more accurate auditory-periphery model with a spectro-temporal modulation filterbank based on Elhilali et al. (2003).

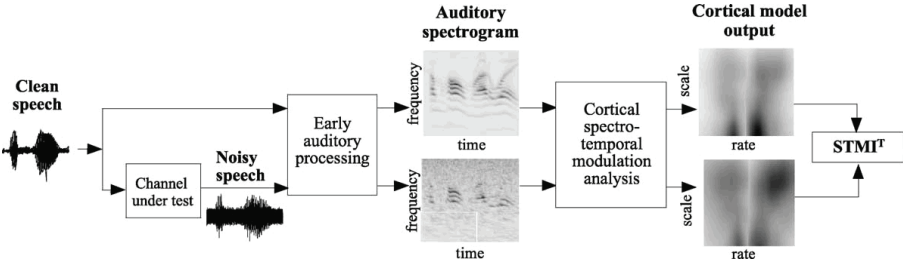

Fig. 5: Schematic of STMI calculation using a model of the auditory periphery ("early auditory processing") and of cortical modulation filters ("cortical spectro-temporal modulation analysis"). The STMI is computed from the difference between a model cortical response template (i.e., response to clean speech) and the response to the corrupted "noisy speech". Reprinted from Elhilali et al. (2003) with permission of Elsevier Ltd. © 2003.

In our modulation filterbank, the temporal rates of the filters range from ±2 to ±32 cyc/sec (Hz), and the scales are in the range from 0.25 to 8 cyc/oct, which covers the range of perceptually important spectro-temporal modulations available in speech for human. The STRFs of the extremes of the modulation filterbank are shown in Fig. 6. Filters with low spectral scale are tuned to broad spectral modulations (bottom row), whereas filters with high spectral scale are responsive to sharp spectral modulations (top row). Likewise, filters with low temporal rate respond preferentially to slow temporal modulations (inner plots), whereas filters with high temporal rate are tuned to rapid temporal modulations (outer plots). Negative rates correspond to an upwards frequency modulation over time in the STRF, while positive rates correspond to a downwards frequency modulation.

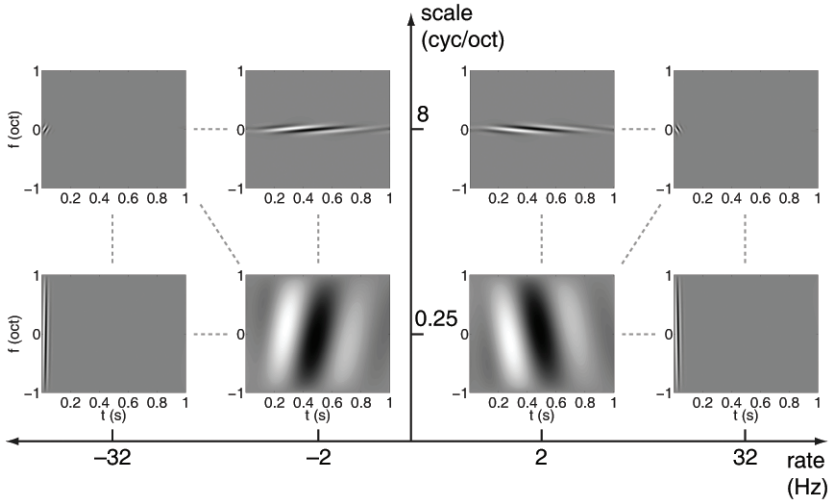

Fig. 6: Example STRFs for the model cortical modulation filters. Black corresponds to time-frequency regions of excitation, while white corresponds to regions of inhibition.

After analyzing the two-dimensional (time and CF) AN neurogram by the modulation filter banks, the cortical output is a four-dimensional (time, CF, rate and scale) complex-valued representation. Figure 7 shows the cortical filterbank output averaged over time and CF for the word "doll". Even when averaged over time and frequency, the cortical output for the normal model at a conversational speech level (top-left panel) shows distinct peaks of spectro-temporal modulations produced by the speech envelope and formant structure. Broad spectral modulations are observed for a range of negative temporal rates (i.e., upwards frequency sweeps over time in the AN neurogram), and downwards frequency sweeps occur at temporal modulation rates of around 4 Hz for ~1 cycle/octave spectral modulations. These distinctive peaks practically disappear for loud speech in the normal model (bottom-left panel). As a result of cochlear impairment, only the low-scale/low-rate modulations are preserved at conversational speech levels (top-right panel), and loud speech wipes out even these modulations (bottom-right panel).

In Zilany and Bruce (2007a), we used the STMI speech intelligibility prediction method depicted in Fig. 5 to obtain direct predictions of speech perception data. In our case, the "clean speech" template was the model cortical response to a sentence presented at 65 dB SPL, and the "corrupted speech" response was produced by background noise, reverberation, deviation in the presentation level from 65 dB SPL, cochlear impairment and/or hearing-aid amplification. The model was able to produce qualitatively good predictions of:

- The rollover in intelligibility at high presentation levels for normal-hearing listeners and aided hearing-impaired listeners;
- The effects of audibility on speech intelligibility for un-aided hearing impaired listeners; and

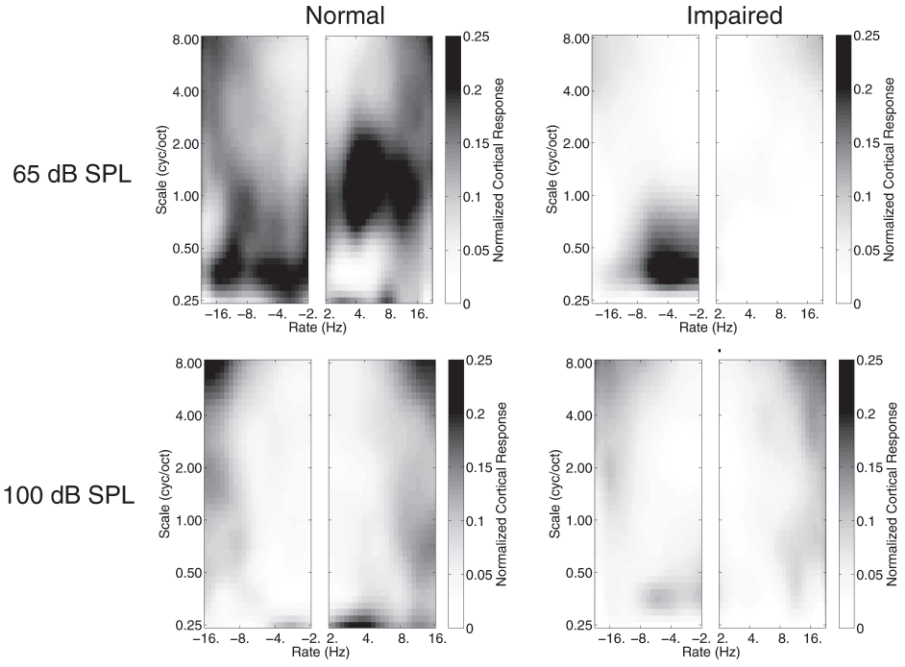

Fig. 7: Model cortical modulation filterbank output for the word "doll". Left column: Normal model response; Right column: Impaired model response, for a gently sloping mild hearing loss; Top row: 65 dB SPL presentation level; Bottom row: 100 dB SPL presentation level.

- The effects of background noise or high-pass/low-pass filtering in these different conditions.

FUTURE DIRECTIONS

The model of Zilany and Bruce (2006, 2007b) is able to describe a wide range of AN fiber data from cats. However, it appears that there are deficiencies in the "synapse" section of the model (Westerman and Smith 1988) that describes the adaptation behaviour of AN responses (Zhang and Carney 2005). The Brachman-Payton synapse model (Payton 1988) appears to explain a somewhat greater proportion of the physiological data, but it is not very computationally efficient and it is difficult to derive suitable model parameters to describe AN fibers with different spontaneous discharge rates (Hewitt and Meddis 1991). Furthermore, Hewitt and Meddis (1991) conducted a comparison of IHC-AN synapse models from the literature and were unable to find a model that was able to explain physiological forward masking data. All of the models that they studied produce exponential or double-exponential adaptation behaviour; we are investigating whether power-law adaptation (Drew and Abbott 2006) may better describe the AN forward masking data, along with the other aspects of AN adaptation.

In addition, it has been suggested that differences in cochlear frequency tuning between cats and humans could have a significant effect on the neural representation of speech in the cat and the human ear (Recio et al. 2002). Modifications to fit human anatomy and physiology have only been investigated with an earlier version of the cat model (Heinz et al. 2001), and the modifications were restricted to changes in the cochlear frequency tuning. It will be of interest to develop a human version of the present model and to investigate how the neural representation of speech may differ in the two species. Substituting a model of the human middle ear (Pascal et al. 1998) for the current cat middle-ear model will be a fairly straight-forward procedure. However, determining how cochlear tuning and compression vary as a function of cochlear position in humans is more difficult; estimates obtained using different psychophysical or otoacoustic emission measurement schemes tend to give somewhat different results (Heinz et al. 2002; Gorga et al. 2007).

With regards to the speech intelligibility predictor, the STMI only takes into account the slower temporal modulations in AN responses and disregards spike-timing information. However, the spike-timing representation of speech appears to be more robust to background noise and hearing impairment than the mean discharge rate representation (Sachs et al. 1983; Miller et al. 1999). In addition, we have found that the gain supplied by a hearing aid to optimally compensate for a hearing loss (according to the AN model) varies depending on whether the spike-timing information is included or just the slower temporal modulations—see the paper by Bruce, Dinath and Zeyl in these proceedings. Furthermore, a recent speech perception study by Lorenzi et al. (2006) also highlights the potential importance of the temporal fine structure of speech for intelligibility in fluctuating background noises. These results motivate the inclusion of spike-timing information in the speech intelligibility metric.

ACKNOWLEDGMENTS

This work was funded by NSERC Discovery Grant 261736 and the Barber-Gennum Chair Endowment.

REFERENCES

Bondy, J., Becker, S., Bruce, I. C., Trainor, L. J., and Haykin, S. (**2004**). "A novel signal-processing strategy for hearing-aid design: Neurocompensation," Signal Process., **84**, 1239–1253.

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

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

Carney, L. H. (**1993**). "A model for the responses of low-frequency auditory-nerve fibers in cat," J. Acoust. Soc. Am., **93**, 401–417.

Depireux, D. A., Simon, J. Z., Klein, D. J., Shamma, S. A. (**2001**). "Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory

cortex," J. Neurophysiol., **85**, 1220–1234.
Drew, P. J., Abbott, L. F. (**2006**). "Models and properties of power-law adaptation in neural systems," J. Neurophysiol., **96**, 826–833.
Elhilali, M., Chi, T., and Shamma, S. A. (**2003**). "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," Speech Comm., **41**, 331–348.
Gorga, M. P., Neely, S. T., Dierking, D. M., Kopun, J., Jolkowski, K., Groenenboom, K., Tan, H., and Stiegemann, B. (**2007**). "Low-frequency and high-frequency cochlear nonlinearity in humans," J. Acoust. Soc. Am., **122**, 1671–1680.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2002**). "Quantifying the implications of nonlinear cochlear tuning for auditory-filter estimates," J. Acoust. Soc. Am., **111**, 996–1011.
Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (**2001**). "Auditory nerve model for predicting performance limits of normal and impaired listeners," Acoust. Res. Lett. Onl., **2**, 91–96.
Hewitt, M. J., and Meddis, R. (**1991**). "An evaluation of eight computer models of mammalian inner hair-cell function," J. Acoust. Soc. Am., **90**, 904–917.
Kiang, N. Y.-S. (**1990**). "Curious oddments of auditory-nerve studies," Hear. Res., **49**, 1–16.
Liberman, M. C., and Kiang, N. Y.-S. (**1984**). "Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate and phase-level functions," Hear. Res., **16**, 75–90.
Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A., **103**, 18866–18869.
Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). "Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers," J. Acoust. Soc. Am., **101**, 3602–3616.
Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999**). "Discriminability of vowel representations in cat auditory-nerve fibers after acoustic trauma," J. Acoust. Soc. Am., **105**, 311–325.
Pascal, J., Bourgeade, A., Lagier, M., and Legros, C. (**1998**). "Linear and nonlinear model of the human middle ear," J. Acoust. Soc. Am., **104**, 1509–1516.
Payton, K. L. (**1988**). "Vowel processing by a model of the auditory periphery: A comparison to eighth-nerve responses," J. Acoust. Soc. Am., 83, 145–162.
Recio, A., Rhode, W. S., Kiefte, M., and Kluender K. R. (**2002**). "Responses to cochlear normalized speech stimuli in the auditory nerve of cat," J. Acoust. Soc. Am., **111**, 2213–2218.
Sachs, M. B., Bruce, I. C., Miller, R. L., and Young, E. D. (**2002**). "Biological basis of hearing-aid design," Ann. Biomed. Eng., **30**, 157–168.
Sachs, M. B., Voigt, H. F., and Young, E. D. (**1983**). "Auditory nerve representation of vowels in background noise," J. Neurophysiol., **50**, 27–45.
Tan, Q., and Carney, L. H. (**2003**). "A phenomenological model for the responses of the auditory-nerve fibers. II. Nonlinear tuning with a frequency glide," J. Acoust. Soc. Am., **114**, 2007–2020.

Westerman, L. A., and Smith, R. L. (**1988**). "A diffusion model of the transient response of the cochlear inner hair cell synapse," J. Acoust. Soc. Am., **83**, 2266–2276.

Wong, J. C., Miller, R. L., Calhoun, B. M., Sachs, M. B., and Young, E. D. (**1998**). "Effects of high sound levels on responses to the vowel /ε/ in cat auditory nerve." Hear. Res., **123**, 61–77.

Zhang, X., and Carney, L. H. (**2005**). "Analysis of models for the synapse between the inner hair cell and the auditory nerve," J. Acoust. Soc. Am., **118**, 154015–53.

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**, 648–670.

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. (**2007a**). "Predictions of speech intelligibility with a model of the normal and impaired auditory-periphery," in Proceedings of 3rd International IEEE EMBS Conference on Neural Engineering, IEEE, Piscatway, NJ, 481–485.

Zilany, M. S. A., and Bruce, I. C. (**2007b**). "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.