Measurement of phonemic degradation in sensorineural hearing loss using a computational model of the auditory periphery

Andrew Hines[†], **Naomi Harte**[*]

Department of Electronic & Electrical Engineering
Sigmedia Group
Trinity College Dublin

E-mail: [†]hinesa@tcd.ie [*]nharte@tcd.ie

Abstract — **A computational model of the auditory periphery enables faster investigation of new signal processing algorithms for hearing aids. This paper presents a study of the degradation of auditory nerve (AN) responses at a phonetic level for a range of sensorineural hearing losses. The AN model of Zilany & Bruce was used to compute responses to a diverse set of phoneme rich sentences from the TIMIT database. The characteristics of both the average discharge rate and spike timing of the responses are discussed. The experiments demonstrate that the model responses are consistent with respect to impairment and inaudible thresholds.**

Keywords — **auditory periphery model, hearing aids, sensorineural hearing loss, phonemic degradation.**

I INTRODUCTION

Hearing loss research has traditionally been based on perceptual criteria, speech intelligibility and threshold levels. The development of computational models of the auditory-periphery has allowed experimentation via simulation to provide quantitative, repeatable results at a more granular level than would be practical with clinical research on human subjects.

Several models have been proposed, integrating physiological data and theories from a large number of studies of the cochlea. The model used in this paper is the cat auditory nerve (AN) model of Zilany and Bruce [1]. The code for the model is shared by the authors and the model responses have been shown to be consistent with a wide range of physiological data from both normal and impaired ears for stimuli presentation levels spanning the dynamic range of hearing[2].

The goal of this study was to analyse the degradation of AN responses at a phoneme level for a range of sensorineural hearing losses, by using the neural representations of speech provided by the model rather than perceptual feedback. This analysis serves a number of objectives. Firstly, it would validate the model's ability to differentiate between phonemes. Having done this, it would allow phoneme groups which have the greatest loss in AN response and largest potentials for restoration to be identified. Finally, it may provide the basis for design of new hearing aid algorithms based on optimal phonetic response restoration.

By presenting a phonetically rich selection of sentences to the AN model, the differences between an unimpaired ear model and three progressively impaired ear models were examined. Unlike prior work where the model's output for individual phonemes [3] or single sentences [4] were examined, this study used a significantly larger test dataset.

Section II introduces the chosen computational model, test dataset used and hearing loss profiles to be examined. Section III presents the methodology employed in gathering the results. Section IV presents the results and section V analyses the results which are subsequently considered with reference to clinical studies of speech intelligibility at a phoneme level. Further work is then proposed based on the results presented.

II BACKGROUND

a) Model

This study used the cat auditory nerve (AN) model developed and validated against physiological data by Zilany and Bruce [2]. The ultimate goal of the model is to predict human speech recognition performance for both normal hearing and hearing impaired listeners [5]. It has recently been used to conduct studies into hearing aid gain prescriptions [3] and optimal phonemic compression schemes[4].

The Zilany and Bruce AN model builds upon several efforts to develop computational models including Deng and Geisler [6], Zhang et al.[7] and Bruce et al.[8]. A schematic diagram of the model is available in Fig. 1 of Zilany and Bruce [2], which illustrates how model responses matched physiological data over a wider dynamic range than previous models by providing two modes of basilar membrane excitation to the inner hair cell rather than one.

The AN model takes speech waveforms, resampled at 100kHz with instantaneous pressures in units of Pascal. These are used to derive an AN spike train for a fibre with a specific characteristic frequency (CF). Running the model at a range of CFs allows neurogram outputs to be generated. These are similar to spectrograms, except displaying the neural response as a function of CF and time.

Two neurogram representations are produced from the AN model output: a spike timing neurogram (fine timing over 10 microseconds); and an average discharge rate (time resolution averaged over several milliseconds). The neurograms allow comparative evaluation of the performance of unimpaired versus impaired auditory nerves.

b) Timit database

The TIMIT corpus of read speech[9] was selected as the speech waveform source. The TIMIT test data has a core portion containing 24 speakers, 2 male and 1 female from each of the 8 dialect regions. Each speaker reads a different set of SX sentences. The SX sentences are phonetically-compact sentences designed to provide a good coverage of pairs of phones, while the SI sentences are phonetically-diverse. Thus the core test material contains 192 sentences, 5 SX and 3 SI for each speaker, each having a distinct text prompt. The core test set maintains a consistent ratio of phoneme occurances as the larger "full test set" (2340 sentences). The speech provided by TIMIT is sampled at 16 kHz.

TIMIT classifies 57 distinct phoneme types and groups them into 6 phoneme groups (Table. 1) and 1 group of "others" (e.g. pauses). The TIMIT corpus of sentences contains phoneme timings for each sentence. These were used in the experiments presented here to analyse neurograms at a phonetic level.

c) Audiograms

Three audiograms representing hearing loss profiles were selected to represent a mild, moderate and profound hearing loss (Fig. 1). The audiograms used match the samples presented by Dillon[10] to illustrate prescription fitting over a wide rage of hearing impairments.

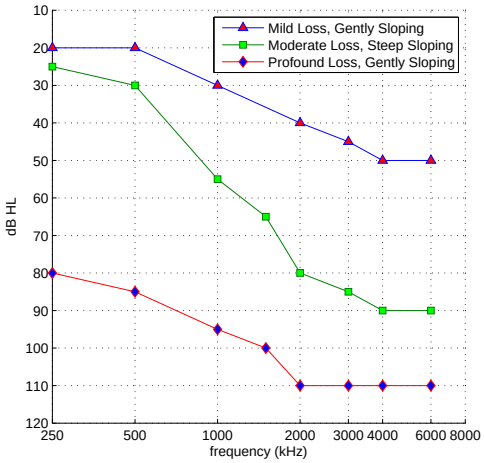

Fig. 1: Hearing Loss Audiograms

III METHOD

a) Collection of Data

For comparative analysis of responses, it was necessary to create and store AN responses for each of the 192 test sentences. The original TIMIT sentence was resampled to the stimulated minimum sample rate for the AN Model (100kHz). For good SPL coverage, the resampled sentence was scaled to 3 presentation levels: a softly spoken level (45 dB SPL), middle 'normal' level (65dB SPL) and a raised voice/shouted level (85 dB SPL). For each presentation level, each sentence was presented to four versions of the AN Model: an unimpaired model, and three increasingly impaired model configurations: mild, moderate and profound. The simulation was carried out with sentences presented free from any form of background noise.

b) Analysis of neural responses

The response of the AN to acoustic stimuli was quantified by the creation of "neurograms". As previously stated, these display the neural response as a function of CF and time. 30 CFs were used, spaced logarithmically between 250 and 8000 Hz. The neural response at each CF was created from the responses of 50 simulated AN fibres. In accordance with Liberman [11] and as used for

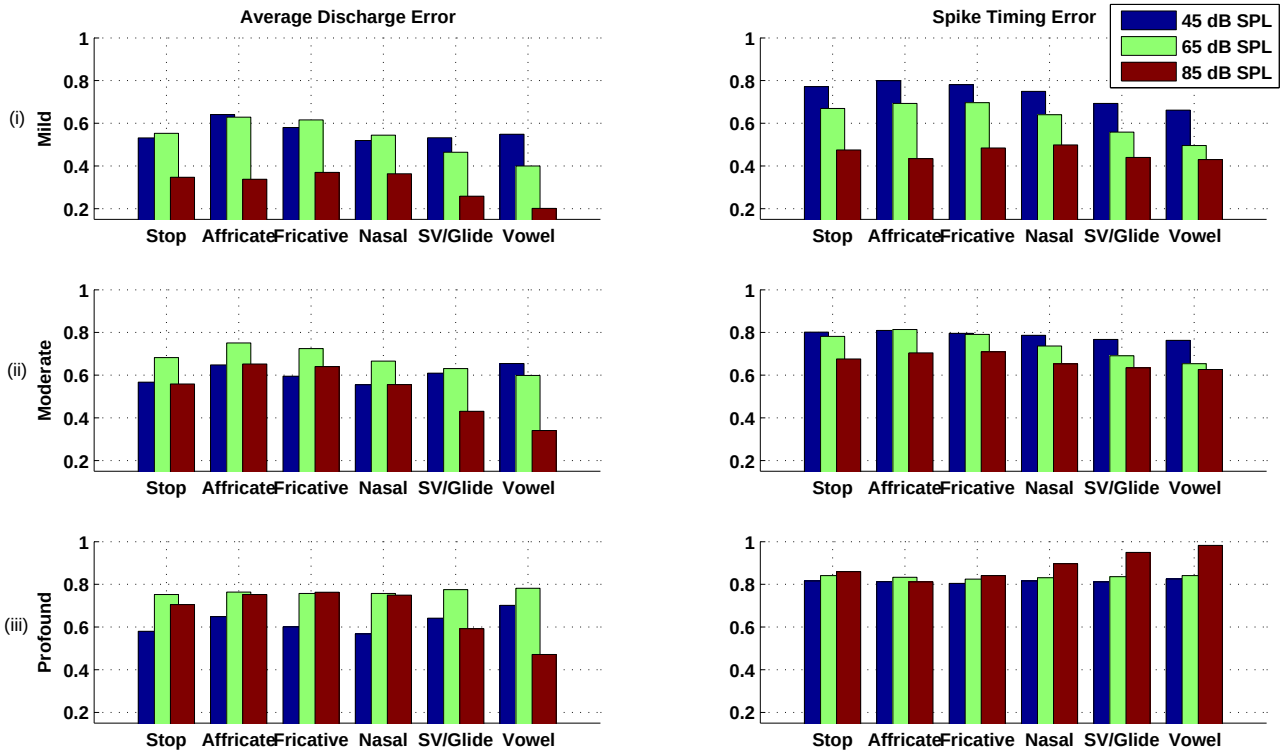

Fig. 2: Error Plot for Mild, Moderate and Profound vs Unimpaired Hearing Loss at 45/65/85 dB SPL.

similar AN Model simulations [4][3], 60% of the fibers were chosen to be high spontaneous rate (>18 spikes/s), 20% medium (0.5 to 18 spikes/s), and 20% low (<0.5 spikes/s). Two neurogram representations were created for analysis, one by maintaining a small time bin size (10μs) which retained granular spike timing information and another with a larger bin size (312.5μs) which gave a moving average discharge rate.

c) Aggregating Phoneme Error Data

The phoneme timing information from TIMIT was used to extract the neurogram information on a per phoneme basis. For each phoneme occurrence, a mean absolute error was calculated between the unimpaired average discharge rate neurogram output and the three impaired models' neurograms. The mean absolute error for a phoneme was divided by the mean of the unimpaired neurogram for that phoneme, to normalise the error with respect to the phoneme sample's input pressure. In effect, the error is then expressed as a fraction of the normal response for the phoneme. This allows for comparisons at different presentation levels and across phoneme types.

This process was repeated using the spike timing neurograms to give two error metrics per phoneme at each hearing loss and presentation level. The errors per phoneme occurrence were collected to find a mean error per phoneme type. These were then sorted into their respective phoneme groupings to find a group mean error.

IV RESULTS

The results are presented in Fig. 2. The three rows (i-iii) represent error measurements for the 3 impaired models (Fig. 1) against the unimpaired model. The first column contains average discharge rate errors and the second contains spike timing errors. Within each bar chart, results are displayed by phoneme group (Table 1) with different coloured bars representing the 3 presentation levels.

Examples of the input signal, input signal spectrogram, and output average discharge rate neurogram and spike timing neurogram for each phoneme group type are presented in Fig. 3. The samples were created against an unimpaired AN model with a presentation level of 85 dB SPL.

Examples of the degradation in output neurograms for each impaired model at a standard presentation level of 85dB SPL are shown for vowels (Fig. 4) and fricatives (Fig. 5).

V DISCUSSION

a) Presentation Level and Audiogram Choices

The chosen presentation levels (45,65,85 dB SPL) were all expected to pose some difficultly to the mildly impaired model with the higher presenta-

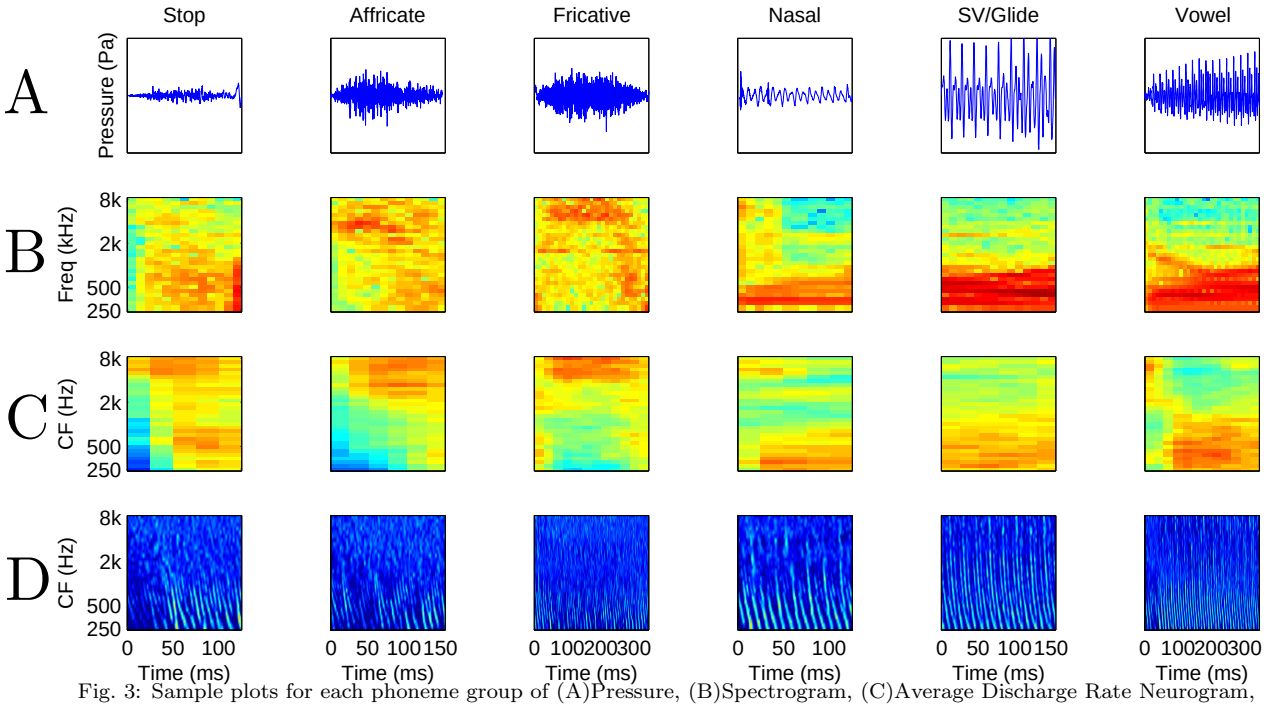

Fig. 3: Sample plots for each phoneme group of (A)Pressure, (B)Spectrogram, (C)Average Discharge Rate Neurogram, (D)Spike Timing Neurogram

Phoneme Group	Phonemes
Stops	b d g p t k dx q
Affricates	jh ch
Fricatives	s sh z zh f th v dh
Nasals	m n ng em en eng nx
SV/Glides	l r w y hh hv el
Vowels	iy ih eh ey ae aa aw ay ah ao oy ow uh uw ux er ax ix axr ax-h

Table 1: TIMIT phoneme groups

tion level near the upper limit of comfortable loudness for both the unimpaired and impaired audiograms.

It was expected that the low presentation level would saturate error readings for some phoneme groups where the important frequencies required were super-threshold. The example neurograms for a vowel (fig. 4) and a fricative (fig. 5) show how the information degrades in the fricative faster than in the vowel in both the average discharge rate and spike timing neurograms.

The error calculations looked at an unimpaired AN Model against three increasingly impaired models. The three audiograms were chosen to give a spread of results. The mild, gently sloping audiogram represents a common sensorineural hearing loss profile of an elderly person. The moderate, steep sloping audiogram gives a contrast between mild loss at lower frequencies and significant loss at higher frequencies. The profound, gently sloping hearing loss gives a profile that should be super-threshold for the first 2 presentation levels and should only be getting limited, low frequency stimulation for the 85 dB SPL presentation level.

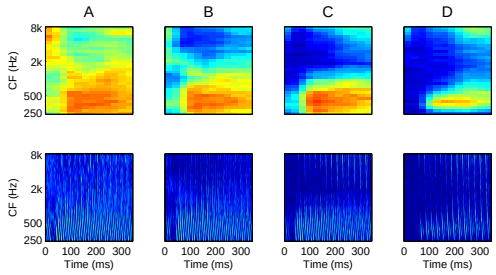

Fig. 4: Sample vowel's neurograms (average discharge rate above, spike timing below) for 4 Audiograms @ 85dB SPL: (A) Unimpaired (B) Mild (C) Moderate (D) Profound

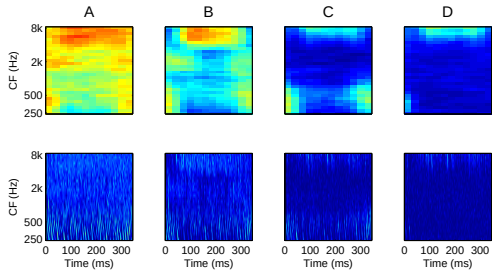

Fig. 5: Sample fricative neurograms (average discharge rate above, spike timing below) for 4 Audiograms @ 85dB SPL: (A) Unimpaired (B) Mild (C) Moderate (D) Profound (85db SPL)

b) Error Analysis

The errors for the moderate and profound losses at 45 dB SPL in Fig. 2 show that both the av-

erage discharge rate and spike timing graphs are very similar and appear to have saturated. The profound error graphs at 65 dB SPL exhibits the same characteristics. This is consistent with the fact that the low presentation level was only supra-threshold for the mild model and partially so for the moderately impaired model.

Examining the mild loss errors at 45 dB SPL it can be seen that in both the average discharge rate and spike timing errors, the predominantly high frequency consonant sounds (stop, affricate, fricative) are comparable in error levels to the moderate and profound losses. This is expected as the audiogram for the mild loss has a threshold of 40+ dB HL for frequencies greater than 2kHz. However, for the nasal, semi-vowel/glide and vowels, the lower frequencies are within the hearing threshold levels and give errors less than the saturated levels. It can be seen that the error magnitudes for mild loss at 45 dB SPL follow the same pattern as in the saturated plots for moderate and profound losses but at a lower level.

At 65 dB SPL, the profound loss is still below threshold for the entire audiogram range. The moderate loss is below threshold from 1kHz. This can be seen in the saturated affricates and fricatives matching the error levels for the profound loss.

For the mild hearing loss, the spike timings for the nasals, SV/glide and vowels follow a similar pattern at 65 and 85 dB SPL with vowels performing best in each case.

While the lower presentation levels and profound loss are useful in validating the model and indicating error saturation points, the 85 dB SPL presentation level gives the most interesting data on the differences between phoneme groups.

At 85 dB SPL, the moderate loss profile is loosing some higher frequencies. The mild loss is active at all frequency ranges and shows that the vowels are performing better than the fricatives. The mild hearing loss vowel error reduces as the presentation level increases suggesting the AN response is benefiting from the higher presentation level. The SV/Glides exhibit similar error patterns.

c) Average Discharge Rate vs. Spike Timing Errors

The average discharge rate and spike timing neurograms represent quite different information.

Examining the mild hearing loss at 85 dB SPL it can be seen that vowel average discharge rate errors are low but the spike timing errors are still comparable to those of fricatives giving greater indication of loss of synchrony in the AN response.

Vowel error for the profound hearing loss at 85 dB SPL is significantly down in the average discharge error reading as information being presented for F1 and F2 in the 250-500 Hz range is partially supra-threshold. However, the spike timing errors remain saturated as this error will capture fine timing errors.

As can be seen in the example illustrations (Fig. 3) glides have a lower frequency than vowels. Vowels consist of a number of formants- 1st 150-850 Hz, 2nd 500-2500 Hz, 3rd 1500-3500. It is generally accepted that the first 2 formants of vowels are the most critical for intelligibility. Fig. 4 illustrates the vowel formant information loss in both neurograms.

d) Intelligibility

In clinical research carried out by Cole et al. [12] and expanded upon by Burkle[13], the contribution of vowel versus consonant information was investigated using a noise replacement paradigm on sentences from TIMIT. Cole used only normal-hearing subjects. Burkle tested two listener groups, one consisting of young normal-hearing participants (YNH95) and the other group of elderly hearing-impaired participants (EHI95).

The signal level was calibrated to a 95dB SPL level so that the sentences would be reasonably audible for this hearing-impaired group. In the study unaltered TIMIT sentences, sentences in which all of the vowels were replaced by noise (Cin); and sentences in which all of the consonants were replaced by noise (Vin) were tested and word and sentence intelligibility were measured.

The hearing loss average thresholds for the (EHI95) group were .25, .5, 1, 2, and 4 kHz were 29, 32, 37, 48, and 57 dB HL, respectively. This is at comparable level to the mild audiogram used in this study (Fig. 1).

The results of both Cole and Burkle's research found that words were more intelligible with only vowels available compared with only consonants available by a factor of 1.5. This was shown to be consistent in trials of both unimpaired and hearing impaired test groups.

Recategorising the error rates from this study for the mild audiogram into the groupings used by Burkle (Table. 2) allowed the errors for vowels vs consonants to be examined (Fig. 6). It can be seen that the errors in the vowels were lower than the consonants in both average discharge and spike timing.

The error rates for vowels vs consonants for mild HL at 85dB seem to have a corollary with the intelligibility of words seen by Cole et al. and Burkle. It is possible to speculate that language has evolved with intelligibility weighted towards the phoneme groups that degrade more slowly with hearing loss or indeed, that more linguistic importance would be weighted in the carrier frequencies with greater robustness. The consonant/vowel er-

ror ratios in Fig. 6 (1.7 for average discharge rate errors and 1.6 for spike timing) are similar to Cole and Burkle's findings but would require more investigation as to whether there is a measurable linkage.

Phoneme Group	Phonemes
Consonants	b d g p t k dx q jh ch s sh z zh f th v dh m n ng em en eng nx l r w y hh hv el
Vowels	iy ih eh ey ae aa aw ay ah ao oy ow uh uw ux er ax ix axr ax-h

Table 2: Burkle vowel/consonant groups

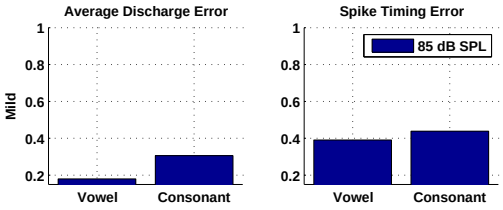

Fig. 6: Mild HL errors at 85 dB SPL using Burkle vowel/consonant groups

	Full	Vin	Cin
YNH95	99	65.1	51.6
EHI95	93.8	40.2	20.0
Combined	96.4	52.65	35.8

Table 3: Burkle Results - Percentage of words identified correctly in each condition

VI CONCLUSIONS AND FUTURE WORK

This study differed from previous studies using the auditory periphery model in that it used a large set of sentences covering 8 dialects and 24 different readers including male and female. This validated the models ability to deal consistently with variable accents, voice pitches and presentation levels.

The results showed that for a wide range of phoneme inputs, the model predicted errors that corresponded well to the phoneme group frequency characteristics. Having validated the models ability to discriminate error rates on a phonemic basis, further tests would yield more information about phonemic differences.

It would be useful to run further audiograms at a mild level and high presentation level to examine the differences in phoneme error where the full frequency range is within the impaired audiogram thresholds. It would also help to compare flat audiograms with 5dB and 85dB HL across all frequencies to allow a better understanding of the minimum and maximum error saturation points for each phoneme type.

The choice of error measure also warrants further investigation as a correlation measure may yield more informative results than the mean absolute error measurement chosen for this study.

Further work, expanding to include classification by visemes would allow analysis of the potential phoneme improvements available by the provision of visual cues.

References

[1] M. S. A. Zilany and I. C. Bruce. Representation of the vowel /E/ in normal and impaired auditory nerve fibers: Model predictions of responses in cats. *J. Acoust. Soc. Am.*, 122(1):402–417, July 2007.

[2] M. S. A. Zilany and I. C. Bruce. Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *J. Acoust. Soc. Am.*, 120(3):1446–1466, Sept 2006.

[3] F. Dinath and I. C. Bruce. Hearing aid gain prescriptions balance restoration of auditory nerve mean-rate and spike-timing representations of speech. *Proceedings of 30th International IEEE Engineering in Medicine and Biology Conference, IEEE, Piscataway, NJ*, pages 1793–1796, 2008.

[4] I.C. Bruce, F. Dinath, and T. J. Zeyl. Insights into optimal phonemic compression from a computational model of the auditory periphery. *Auditory Signal Processing in Hearing-Impaired Listeners, Int. Symposium on Audiological and Auditory Research (ISAAR)*, pages 73–81, 2007.

[5] M. S. A. Zilany. Modeling the neural representation of speech in normal hearing and hearing impaired listeners. *PhD Thesis, McMaster University, Hamilton, ON.*, 2007.

[6] L. Deng and C. D. Geisler. A composite auditory model for processing speech sounds. *J. Acoust. Soc. Am.*, 82:2001–2012, 1987.

[7] X. Zhang, Heinz, M. G., I. C. Bruce, and L. H. Carney. A phenomenological model for the responses of auditory-nerve fibers. i. non-linear tuning with compression and suppression. *J. Acoust. Soc. Am.*, 109:648–670, 2001.

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

[9] U.S. Dept. Commerce DARPA. The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. *NIST Speech Disc 1-1.1*, 1990.

[10] H. Dillon. Hearing Aids. *New York: Thieme Medical Publishers*, 2001.

[11] M.C. Liberman. Auditory nerve response from cats raised in a low noise chamber. *J. Acoust. Soc. Am.*, 63:442–455, 1978.

[12] R.A. Cole, Yan Yonghong, B. Mak, M. Fanty, and T. Bailey. The contribution of consonants versus vowels to word recognition in fluent speech. *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1996 (ICASSP-96)*, 2:853 – 856, 1996.

[13] T. Zachary Burkle. Contribution of consonant versus vowel information to sentence intelligibility by normal and hearing-impaired listeners. *Masters Thesis, Indiana University, Bloomingdale, IN*, 2004.