Modeling intelligibility of hearing-aid compression circuits

Jeff Bondy, Ian C. Bruce, Rong Dong, Suzanna Becker and Simon Haykin
McMaster University, Hamilton Ontario, L8S 4S4

Abstract—**The active filtering effect in the inner ear is disrupted with sensorineural hearing impairment. This causes a loss of frequency selectivity and dynamic range. Compression is often used in hearing-aids in an attempt to re-establish the normal dynamic range of the cochlear response. While some studies show increased speech intelligibility with artificial noise sources for compressive hearing-aids, most show little (< 1 dB versus linear aids) or no advantage in competing speech. In this paper we explore a quantitative model to explain the empirical performance of compressive hearing-aids in competing speech. By combining an accurate cochlear model with a model of higher auditory feature analysis based on spectral-temporal clustering of onsets, we provide an explanation for the failure of hearing-aid compression algorithms to increase intelligibility. Our proposed spectral-temporal intelligibility model suggests that increasing intelligibility for a hearing impaired person in competing speech requires both spectral and temporal suppression.**

Index Terms—**Adaptive signal processing, compression, clustering, intelligibility metrics.**

I. Introduction

The rationale behind including some sort of compressive preprocessing in a hearing-aid is the fact that an auditory system loses dynamic range due to sensorineural impairment [1]. Most researchers now agree that this loss is due to the destruction of hair bundles on outer hair cells (OHCs) in the cochlea. OHCs mechanically modulate the traveling wave of acoustic energy along the basilar membrane. This modulation acts as a nonlinear amplification at a particular frequency, and is also responsible for the suppression or contrast enhancement characteristics of a normal ear. Presumably, to restore normal hearing to a sensorineural impaired individual, there must then be some sort of compression in a hearing-aid.

Compression circuits in hearing-aids are characterized by time, intensity and frequency parameters. Individual parameters are selected based on reasons such as loudness normalization, discomfort avoidance, or dynamic range compression. Dillon gives an overview and tutorial on the competing rationales and characteristics [2]. The degrees of freedom available to a hearing aid circuit designer make it infeasible to perform empirical intelligibility testing across all the possible parameters. Also, these studies look at what can be done to alleviate the symptoms of sensorineural hearing impairment, but do not really address the core problem.

Jeff Bondy is with McMaster University. phone: 905-525-9140; fax: 905-521-2922; e-mail: jeff@soma.crl.mcmaster.ca.

The true problem that needs to be modeled is how the compressive non-linearity of the cochlear amplifier, disturbed by sensorineural hearing loss, can be restored by signal processing in a hearing-aid. There is a complicated set of signal processing that is taking place in the cochlea that ultimately affects intelligibility. Quantitative evaluation of compression circuits in hearing-aids, reduces the burden on empirical testing.

Quantitative analysis must also predict why there is such a large discrepancy between the hearing impaired and normal hearing person's ability to unmask competing speech. Understanding this disparity is key to building optimal compression circuits. Reference [3] shows a SNR advantage between 12-15 dB for normal hearing people over hearing impaired people in identifying syllables in competing speech. Over time, testing methodologies have been refined, but results still show an enormous discrepancy between normal hearing and hearing impaired people's ability to understand speech against contending speech. Table 1 gives an overview of normal versus hearing impaired peoples ability to recognize target speech with a masking speaker.

TABLE I
Intelligibility in Speech and Speech-Like Noise

Study	Description	SRT (Normal/Impaired)
Duquesnoy, 1983 [4]	20 elderly subjects with ski-slope high frequency loss; freefield; Competing @ 55 dBA	-17.6/-5.3
Festen & Plomp, 1990 [5]	20 mixed age and losses; monaural earphones; Competing @ 80 dBA	-11.4/-1.1
Hygge et al., 1992 [6]	24 mixed age; freefield, binaural; Competing Speech.	-9.2/7.0 * SNR
Peters et al., 1998 [7]	10 elderly subjects with ski-slope high frequency loss; monaural earphones; Competing @ 65 dBA	-11.9/0.8

To underscore Table 1, in noise with a long term average speech spectrum (LTASS) the difference in SRTs between normal and impaired hearing individuals is only 2-5 dB [8].

0-7803-8104-1/03/$17.00 ©2003 IEEE

It seems that to allow a sensorineural impaired person the ability to operate in the classical cocktail party in a way that approaches a normal hearing person, auditory compression must be understood in the competing speech regime. This will intertwine the counterbalanced processes of compression and suppression.

II. METHOD

We preprocess an acoustic signal with a compression algorithm, and then model the resultant representation by the auditory nerve activity, taking into account cochlear processing effects including compression and suppression, as well as the effects of sensorineural hearing loss. We further process this auditory representation by calculating regions of onset of activation, and clustering the onset data across time and frequency. This spectro-temporal fusion reveals a very different pattern between normal and impaired auditory representations. We can thereby map the distortion between a normal and impaired hearing representation in this domain, to obtain a novel model of intelligibility. In this way we hope to produce a quantitative model to predict hearing aid performance.

There have been several previous attempts to produce a quantitative model to assess hearing aid performance or for hearing-aid circuit design. Fabry and van Tassel [9] used the articulation index, Kates [10] used a fairly simple compressing/suppressing model, and Anderson [11] used an invertible auditory model. None of these attempts represent temporal information. The basic advantage of our method over these models is the introduction of timing information into the distortion metric. From the introductory discussion it is seen that the temporal modulations in competing speech are important in unmasking target speech in normal hearing people but are not accessible to hearing impaired people. We aim to show that most temporal information is lost with sensorineural impairment, and that present hearing-aid processing strategies do not address this.

A. Empirical Data

Our model follows the process and data in Moore, Peters and Stone [12]. They carried out SRT tests on elderly hearing-impaired people with ski-slope, high-frequency loss with simulated linear, WDRC and multiband (2, 4, 8) compression hearing-aids. The subjects were fitted using the "Cambridge" formula [13] for the linear condition. The compression ratio (CR) and threshold (CT) were determined by applying the following two constraints:

1. The gain in each channel for a 65 dB SPL, speech shaped, input noise is the same as in the linear condition.
2. The gain in each channel makes a 45 dB SPL speech signal in 65 dB SPL noise above the impaired hearing threshold.

The second constraint could not be held in all conditions. Our example loss profile was the average loss profile of the 18 subjects from the study. The hearing loss in dB SPL, compression ratio and compression threshold per channel are listed in Table 2. The attack and release times are typical of fast compression: both were 8.2 ms. The output SRTs in competing speech for the unaided, linear, and eight-channel compression were reported as 0.5, -2.0, and -2.9 dB respectively [12].

TABLE 2
LOSS PROFILE AND PARAMETERS FOR A COMPRESSION CIRCUIT

Frequency	Hearing Loss	CR	CT
250	28	1.7	22.3
500	31	1.1	24.6
1000	38	1.3	16.1
2000	50	1.7	9.5
3000	59	2.4	7
4000	64	2.9	7
5000	66	2.9	7
6000	68	2.9	7

While Moore Peters and Stone used several different noise types, we focused on competing speech because of the large differences in intelligibility between stimulus types at the same SNR. We used the same HINT sentences [14], but recorded for multiple talkers [15]. To compile the statistics we used twenty of the test sentences from the HINT corpus.

B. Auditory Model

The auditory periphery model used throughout was taken from Bruce et al. [16]. This model comprises several sections, each providing a phenomenological description of a different part of auditory periphery function.

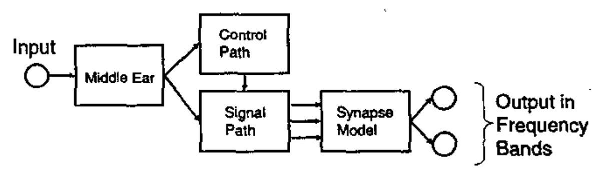

Fig. 1 Phenomenological auditory model [16].

The first section models outer and middle ear filtering up to the cochlea. 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 synapse model transforms the acoustic pressure wave after the nonlinear filtering of the control and signal path into spontaneous rate with adaptation. This is the instantaneous spontaneous rate used throughout this paper.

In our model, there was some damage to the inner and outer hair cells, so there is not ideal functioning in the cochlea. From the threshold losses detailed in Table 2, impairments of inner and outer hair cells were calculated so that OHC impairment accounted for around 50-60% of the total threshold shift at a frequency, in dB [17]. The percent IHC loss was then adjusted to explain the remaining threshold shift. Fig. 2 is an example of normal and damaged auditory responses.

Effects of sensorineural impairment such as spreading in time and frequency are evident in the lack of separation between the lines representing formant frequencies. What is not obvious in this representation is how intelligibility is affected.

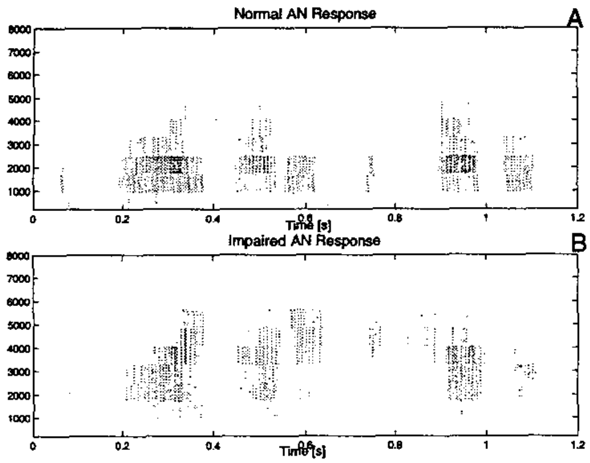

Fig. 2 Auditory representation for the sentence "The boy got into trouble". Part A is from a normal hearing auditory model, B is from a sensorineural impaired auditory model. There is noticeable smearing in time and frequency.

C. Perceptual Grouping of Onset Stimuli

A representation of the acoustic waveform allowing grouping of onset cues was chosen as a way of identifying acoustic events that are perceptually relevant and may be the source of the intelligibility difference between normal hearing and hearing impaired people in competing speech. Onset characteristics of the auditory representation were calculated with a difference of exponentials filter, $h_1[n]$, in each frequency band

$$h_1[n] = \frac{n}{\alpha_1^2}\exp^{-n/\alpha_1} - \frac{n}{\alpha_2^2}\exp^{-n/\alpha_2}. \quad (1)$$

α_1 and α_2 were selected to pass frequencies from 4 to 32 Hz. These frequencies contribute most to intelligibility, with a signal's fine temporal structure only adding a small amount to intelligibility [18].

This onset data was then integrated over a typical acoustic event time window, $h_2[n]$, which had a 6 dB cutoff at 125 Hz. This integrator had a similar form to $h_1[n]$,

$$h_2(t) = \frac{t}{\alpha_3^2}\exp^{-t/\alpha_3}. \quad (2)$$

For a sample rate of 11025 Hz α_1 was 0.06, α_2 was 0.10, and α_3 was 0.001. An adaptive threshold and refraction operation was then applied. The threshold value was determined to produce some percentage of active events in the discretized time-frequency grid when the refractory period is 1 ms. We chose one tenth of a percent, but anything suitably sparse clustered in our experiments, and gave similar results (anything less then 1% clustered in our experience). This produced a discrete event map such as the one given in Fig. 3.

Fig. 3 clearly shows that important timing information is carried across multiple frequencies. To calculate perceptual relevance we applied a clustering algorithm using a hard-decision rule for class membership based on a Gaussian probability distribution assumption. Taking the thresholded

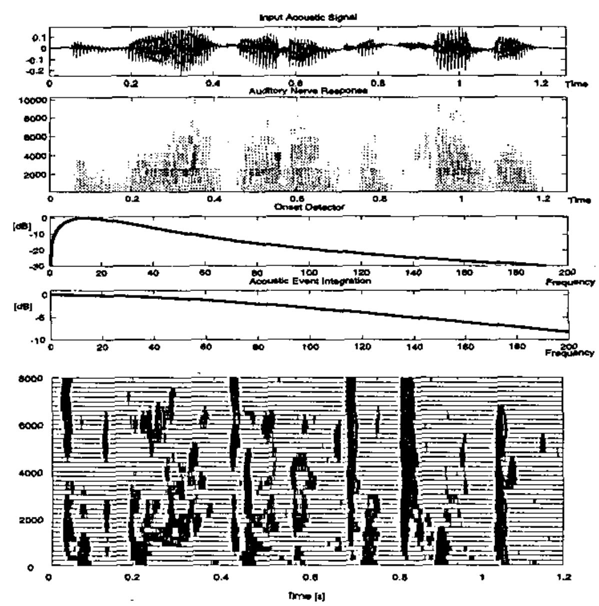

Fig. 3 The input signal is represented on the auditory nerve. Then low frequency auditory information is extracted with an onset detector. The onset events are then integrated over a speech cue period. With thresholding and a refraction window, discrete events are finally mapped.

information from Fig. 3, and making each event, k, a two dimensional sample in time (subscript t) and frequency (subscript f), $\bar{z}_k = \{z_{tk}, z_{fk}\}$, the whole set of acoustic events is represented as **Z**. Starting with a limited number of possible classes, J, we run an iterative clustering algorithm, with death for small clusters. A sample was assigned to class j when

$$\pi_j P(j \mid \bar{z}_k) > \pi_i P(i \mid \bar{z}_k), \quad all\ i \neq j \quad (3)$$

where

$$P(j \mid \bar{z}_k) = \frac{1}{\sqrt{2\pi}\left|\det\left(\Sigma_j\right)\right|}\exp^{-\frac{1}{2}(\bar{z}_k-\mu_j)\Sigma_j^{-1}(\bar{z}_k-\mu_j)^T} \quad (4)$$

π_j, μ_j *and* Σ_j are the prior probability, mean and covariance statistics for class j, respectively. All samples were classified before the prior and statistics are recalculated in a batch mode. Classes with a low prior probability were pruned; in these examples, classes with less then half a percent of all the events were discarded. Classification and statistical updates were iterated until the priors stopped changing between iterations by more then two percent root-mean-square.

The classes were then split in half along the temporal axis and classification was again seeded and performed in the halved datasets to account for time warping or long pauses. This bifurcation helps competition and reduces reliance on initial conditions. The result of this clustering, using the onset data from Fig. 3 and bifurcated once with 50 initial classes is given in Fig. 4.

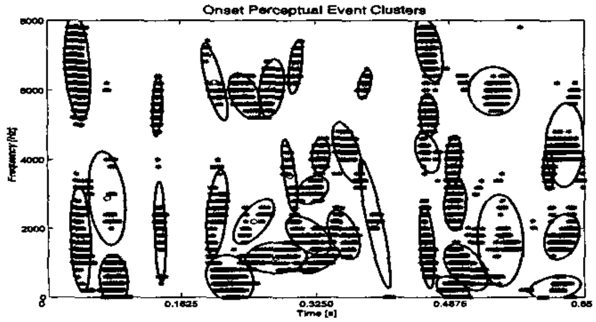

Fig. 4 The normal hearing, perceptual clusters mean and two-standard deviation contours are plotted for the first half of the example sentence.

III. Results

A. Single Speech Stream

The goal of this research was to be able to quantify effects on intelligibility of nonlinear, dynamic algorithms for sensorineural impairment. The question we looked at first was whether our "perceptually relevant" clusters produce distinctly different representations for normal and impaired auditory models. Using the same Cambridge linear fitting strategy, with the simulated steeply sloped hearing loss as detailed in Table 2 we presented the stimuli to the normal and damaged auditory models. The normal model produced the clusters shown in Fig. 4, the damaged model with preprocessing gain calculated by the Cambridge formula produced the clusters shown in Fig 5.

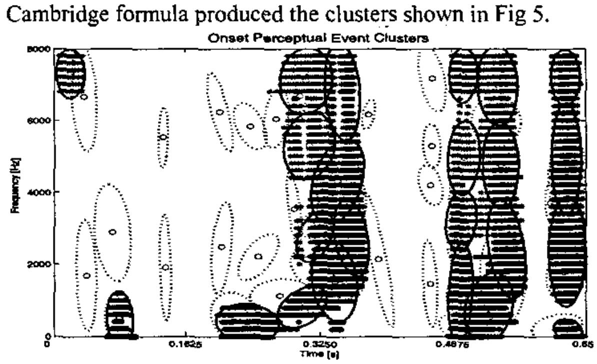

Fig. 5 The sensorineural impaired perceptual clusters mean and two-standard deviation contours are plotted in solid lines. The normal clusters from Fig. 4 are plotted with dotted lines for reference.

In this particular example, in the case of the impaired model, fewer classes are formed, the variances of those classes are greatly enlarged, and entire onset cues for some phones are lost. The dotted lines representing the normal hearing clusters are sometimes far removed from the impaired clusters. These results are indicative of both spectral and temporal spreading. Table 3 highlights the general results.

This general pattern should be indicative of differences between normal and impaired listeners on the order of 3 to 5 dB in SRT. This is the baseline deficit that hearing impaired people face in conditions without any temporally modulated noise. Another test versus empirical data is to judge the difference between linear and the 8-channel compression preprocessing. With the 8-channel compression circuit the impaired results are almost identical to the linear case. In only three of the twenty test sentences did compression produce a "phantom grouping", where a cluster was formed outside of a phone boundary. This is the expected result with speech presented at 65 dB SPL because it will very rarely go under the compression threshold with the windowed energy calculation used here. Compression circuits do not overly change the AN representation of onset cues.

TABLE 3
Differences in Normal and Impaired Perceptual Clustering

Variable	Normal	Impaired Linear	Impaired 8-Channel
σ_t	10 ms	11 ms	12 ms
σ_f	398 Hz	503 Hz	517 Hz
Classes/second	53.8	35.5	32.6

B. Competing Speech Streams

So far we have dealt with speech in noise. An important question is how are these results affected by competing speech streams? Moore, Peters and Stone used a female talker whose long term average speech spectrum was modified to match the male targets as an interfering signal. The time envelope was basically undisturbed. Figure 6 shows the clustering that takes place in a normal hearing model for these data.

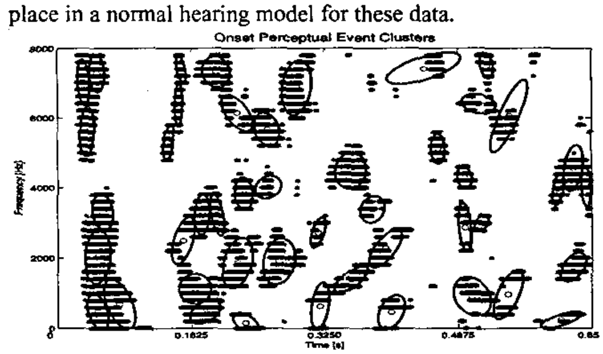

Fig. 6 Normal competing speech clusters for the same input as Figure 4 and 5. There are more classes with smaller variances between them then the normal hearing model clustering without competing speech.

The compression and suppression characteristics of a normal undamaged ear have clearly changed the representation between target speech and target speech with competing speech. The clusters are smaller and more are made. This is not the case in the impaired auditory system's ability to cluster two speech signals as shown in Fig. 7.

Fig. 7 is the grouping that takes place with 8-channel compression. Clearly the auditory system can not make use of the onset characteristics of a speech signal with this type of compression. While the normal ear responds with more specific groupings of acoustic events because of spectral-temporal suppression, the normal compression circuit does nothing to reestablish normal cochlear signal processing.

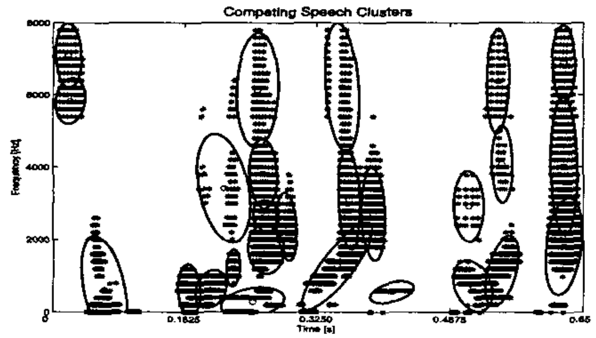

Fig. 7 Impaired competing speech clusters for the same input as Fig. 4 and 5. There are roughly the same number of classes and those class variances remain high.

Table 4 is a comparison of the clustering statistics in competing speech.

TABLE 4
DIFFERENCES IN NORMAL AND IMPAIRED PERCEPTUAL CLUSTERING IN COMPETING SPEECH

Variable	Normal	Impaired Linear	Impaired 8-Channel
σ_t	10 ms	12 ms	12 ms
σ_f	348 Hz	555 Hz	573 Hz
J	70.8	38.4	34.1

Comparing to Table 3 the data that jumps out is the much smaller variances in the normal ear, the larger number of classes, while the statistics for the impaired ear remains remarkably similar. This is conceivably the reason why a normal hearing person has reduced SRT in competing speech versus steady noise (-12 dB versus -4 dB) while a hearing impaired person does not see the same level of advantage (-2 dB versus 1 dB) [7].

We think of intelligibility as the ability to group perceptually relevant acoustic cues while removing other events from different streams. This is different from the articulation index (AI) or speech intelligibility index (SII). They calculate the intelligibility of a speech token based on the summation of signal-to-noise ratio (SNR) in a set of bands. We maintain that a more appropriate measure of the intelligibility of a speech token is the event-to-noise ratio. Here our events will have some spectral-temporal mask that can be used to determine whether the acoustic cue is discriminable. This can test specific phones, while the AI and SII measures have an implicit assumption about the ensemble statistical structure of speech across frequency.

IV. CONCLUSION

We have outlined a way of representing acoustic material that qualitatively predicts intelligibility for a compression circuit in a hearing-aid in competing speech. This representation is affected by time, intensity and frequency parameters. To make it into a useful intelligibility metric it still needs validation against normal conditions, and a mapping between the clustered space and a scalar intelligibility value.

To make the ideal predictor, the clustering space needs to be optimized. The input spectrum is not entirely contiguously represented in the auditory cortex, so the frequency dimension may be "folded" to make different frequencies proximal in the clustering space. There is also the extension to higher dimensionality. One necessary addition would be the phase dimension.

Another possibility to moving towards the ideal hearing-aid predictor would be the addition of attentional information. A prior probability weighting for the expected speech stream could easily be added in the clustering algorithm.

REFERENCES

[1] B.C.J. Moore, Perceptual Consequences of Cochlear Damage. New York: Oxford, 1995.

[2] H. Dillon, "Compression? Yes, but for low or high frequencies, for low or high intensities, and with what response times?" Ear. Hear. **17**, 1996, 287-307.

[3] R.C. Carhart and T.W. Tillman, "Interaction of competing speech signals with hearing losses" Arch. Otolaryngol. **91**, 1970, 273-279

[4] A.J. Duquesnoy, "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons" J. Acoust. Soc. Am. **74**(3), 1983, 739-743.

[5] J.M. Festen, and R. Plomp, "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing", J. Acoust. Soc. Am. **88**, 1990, 1725-1736.

[6] S. Hygge, J. Rönneberg, B. Larsby, and S. Arlinger "Normal hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech and noise backgrounds" J. Speech Hear. Res. **35**, 1992, 208-215.

[7] R.W. Peters, B.C.J. Moore and T. Baer, "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people" J. Acoust. Soc. Am., **103**(1), 1998, 577-587.

[8] B.R. Glasberg and B.C.J. Moore, "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech" Scand. Audiol. Suppl. **32**, 1-25.

[9] D.A. Fabry and D.J. van Tassell "Evaluation of an articulation-index based model for predicting the effects of adaptive frequency response hearing aids" Journal of speech and hearing research, **33**, 1990, 676-689.

[10] J.M. Kates, "Toward a theory of optimal hearing aid processing" J. of Rehab. Res. **30**(1), 1993, 39-48

[11] D.V. Anderson, Model based development of a hearing aid. M.S. Thesis, Brigham Young University, 1994.

[12] B.C.J. Moore, R.W. Peters and M.A. Stone "Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips" J. Acoust. Soc. Am. **105**(1), 1999, 400-411.

[13] B.C.J. Moore and B.R. Glasberg, "Use of a loudness model for hearing-aid fitting. I. Linear hearing aids" Br. J. Audiol. **32**, 1998, 317-335.

[14] M. Nilsson, S.D. Soli and J.A. Sullivan, "Development of the Hearing in Noise Test for measurement of speech reception thresholds in quiet and in noise" J. Acoust. Soc. Am. **95**, 1085-1099.

[15] L. Trainor, R. Sonnadara, K. Wiklund, J. Bondy, S. Gupta, S. Becker, I.C. Bruce, and S. Haykin "Development of a flexible, realistic hearing in noise test environment (R-HINT-E)" to appear in Signal Processing 2003.

[16] 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. Am., **113**(1), 2003, 369-388.

[17] B.C.J. Moore, B.R. Glasberg, and D.A. Vickers "Further evaluation of a model of loudness perception applied to cochlear hearing loss" J. Acoust. Soc. Am. **106**(2), 1999, 898-907.

[18] R. Drullman, J.M. Festen, and R. Plomp, "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670-2680.