A HEARING MODEL TO ESTIMATE MANDARIN SPEECH INTELLIGIBILITY FOR THE HEARING IMPAIRED PATIENTS

Pei-Chun Tsai[1], Shih-Ting Lin[1], Wen-Chung Lee[1], Chung-Chien Hsu[1], Tai-Shih Chi[1] and Chia-Fone Lee[2]

[1]Department of Electrical and Computer Engineering
National Chiao Tung University, Hsinchu, Taiwan 300, R.O.C.
[2]Department of Otolaryngology
Hualien Tzu Chi Hospital, Hualien City, Hualien County, Taiwan 970, R.O.C.

ABSTRACT

A hearing model, which is parameterized by hearing thresholds, degrees of loudness recruitment and reductions of frequency resolution of a hearing-impaired (HI) patient, is proposed in this paper. The model is developed in the filter-bank framework and is flexible for fitting hearing-loss conditions of HI patients. Psychoacoustic experiments were conducted under clean and noisy conditions to validate the model's capability in predicting Mandarin speech intelligibility for HI patients. Statistical analysis on the hearing-test results suggests that the proposed model can predict Mandarin speech intelligibility for HI patients to a certain degree.

***Index Terms*—**hearing model, hearing impaired, loudness recruitment, cochlear frequency selectivity, speech intelligibility

1. INTRODUCTION

Common hearing aids usually focus on compensating the hearing thresholds for HI patients, such as the National Acoustic Laboratories' (NAL) prescription [1]. However, in addition to the hearing threshold, hearing loss also associates with reduced dynamic range (loudness recruitment) [2], reduced temporal resolution for narrow-band fluctuating stimuli [3][4], and reduced frequency selectivity [5][6]. Although DSP algorithms can be developed to compensate hearing-loss factors, their efficacy is hard to assess. The most straightforward evaluation for an algorithm is to play processed speech signals to the hearing impaired. However, each patient suffers from a unique degree of hearing loss and his willingness to participate in extensive psychoacoustic experiments is usually low. These problems become troublesome for developing hearing aids. Therefore, as an alternative approach, a hearing model for the hearing impaired is in great need. The goal is to build a hearing model such that normal people hearing processed speech through the model would have similar sensations to the hearing impaired perceiving normal speech. If we can successfully build such model, we might be able to evaluate hearing-aid algorithms by conducting psychoacoustic experiments on normal-hearing (NH) subjects.

Nejime and Moore integrated previously developed algorithms [7][8] to form a model which simulates hearing-loss factors of hearing threshold elevation, loudness recruitment, and reduced frequency selectivity [9]. Their goal was to verify those hearing-loss factors indeed affect speech intelligibility of NH people. In the first stage of their model, they adopted the concept of smearing the spectrum to simulate the reduced frequency selectivity of the cochlea. The equal loudness curve (ELC) correction was also considered to compensate the gains given by the outer and the middle ear [10]. In the second stage, the speech signal was split into 13 subbands and threshold elevation and loudness recruitment were simulated in each subband [8]. As shown in [11], by matching the loudness curves of two ears of subjects with unilateral hearing loss, one can observe the loudness function between the normal ear and the impaired ear is almost a straight line up to 90~100 dB sound pressure level (SPL), and the slope N of the function determines the degree of loudness recruitment. The larger the slope, the more severe loudness recruitment the patient experiences. This linear function simply corresponds to increasing the instantaneous magnitude of each sub-band signal to power of the value of the slope. It was implemented in [8] with small modifications to avoid generating noise during the exponentiation process. In short, this stage amplifies the magnitude difference of adjacent time-domain samples. Such large variations in loudness are usually perceived and reported by HI patients.

Some characteristics of the above cited model are worth noting. First, the minimum audible level (MAL) in each subband cannot be set independently. In the model, the MAL depends on the degree of the loudness recruitment (the slope N), which is assumed ended at 100 dB SPL. In other words, the MAL is automatically determined by the 100 dB SPL and the recruitment slope N. Second, the frequency smearing was implemented by smoothing the magnitude response of a Fourier spectrum and keeping the phase response intact. However, as reported in [7], keeping the original phase would offset the degree of smearing imposed on the magnitude spectrum such that the desired degree of smearing is hard to reach. Third, the frequency smearing and the loudness recruitment were implemented in the short-term Fourier transform (STFT) and the filter-bank frameworks, respectively [9], but not in a unitary framework. In this paper, we propose a unitary hearing model in the filter-bank framework for the hearing impaired. Our proposed model has parameters of MALs, degrees of loudness recruitment and degrees of frequency smearing. In addition, the degree of frequency smearing is ensured by using white-noise carrier to discard the original fine structures (phases) in each subband.

This paper is organized as follows. In section 2, we describe the proposed model. Experiment results of Mandarin speech intelligibility tests are demonstrated in section 3 to validate the proposed model. Finally, we end in section 4 with conclusions and discussions.

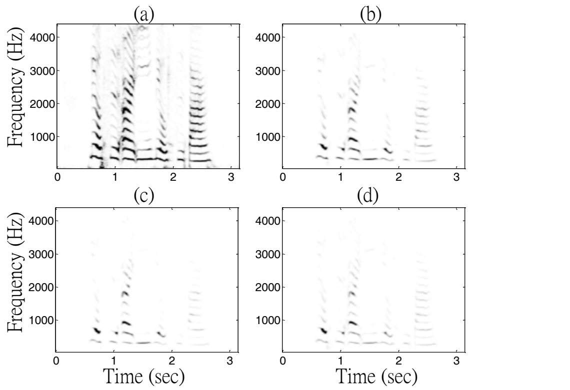

Fig. 1. Magnitude spectrograms with different degrees of loudness recruitment. (a) Original spectrogram; (b)-(d) spectrograms with loudness recruitment factor N = 1.5, 2, and 3.

2. PROPOSED MODEL

In this section, we describe the proposed model which includes the loudness module and the spectral smearing module to simulate the effects of loudness recruitment and reduction of frequency selectivity, respectively. Simple examples are given to show the effects of these modules. At the end, the two modules are integrated within the filter-bank framework.

2.1. Loudness module

We adopt the same assumption as in [9] that NH listeners can hear sound from 0 dB to 100 dB SPL. Instead of assuming the loudness recruitment ends at 100 dB, we include the MAL as an important parameter. Therefore, we modify the magnitude of speech signals sample-by-sample using the following piecewise linear function.

$$\begin{cases} L_p = 0 + (L_u - T), & L_u < T \\ L_p = 0 + N(L_u - T), & T \le L_u < T + 100/N \\ L_p = 100 + [L_u - (T + 100/N)], & T + 100/N \le L_u \end{cases} \quad (1)$$

where L_u (in dB) is the intensity of the original sound; L_p (in dB) is the intensity of the processed sound; *N* represents the degree of loudness recruitment; *T* (in dB) represents the MAL; and 100/*N* represents the range of sound level where loudness recruitment occurs. Note, L_p simulates the perceived intensity by HI patients. According to equation (1), no recruitment occurs when the sound intensity is less than the MAL. When the sound intensity is within the recruitment range, the perceived intensity should be determined by *N* and *T* collectively. When the sound intensity is higher than the upper bound of the recruitment range, the perceived intensity is the surplus plus 100 dB.

Fig. 1 shows the magnitude spectrograms of a sample utterance with different recruitment factor N = 1.5, 2 and 3. The minimum audible level *T* is set to 33.3, 50 and 66.6 dB, respectively, to match the L_u, L_p relation of the recruitment model proposed in [9]. The normalized root mean square error between magnitude spectrograms from our proposed module and Moore's model in [9] is less than 1%. Therefore, our module not only can match Moore's model by choosing an appropriate *T*, it also provides an option to set MALs, which are important hearing parameters of a HI patient.

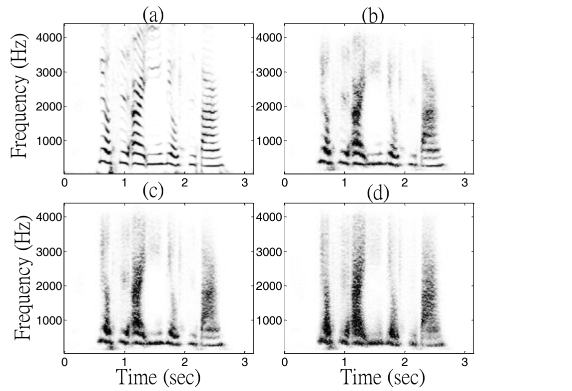

Fig. 2. Magnitude spectrograms after frequency smearing. (a) Original spectrogram; (b)-(d) smeared spectrograms with symmetric broadened factor of 1.5, 3, and 6, respectively.

2.2. Spectral smearing module

Broader bandwidths of cochlear filters degrade the frequency selectivity of the HI patients. The decrease of the frequency resolution can be modeled by smearing the spectrum. Considering cochlear filters at a particular location on the basilar membrane of NH people and of HI patients, frequency components within the impaired (broader) cochlear filter but out of the normal (narrower) cochlear filter would not excite the auditory nerve of NH people. Therefore, we need to modulate those components in a certain way such that NH people can hear those components to experience the sensation of HI patients. To model any shapes of broadened filters for different HI patients, each broadened filter is constructed by a linear combination of all normal filters in our approach. Therefore, the response of a broadened filter can be simulated by modulating output powers of all normal filters with their corresponding gains (i.e., coefficients of the linear combination equation) to the center frequency of that broadened filter.

The 128 cochlea filters in the auditory model [12] were used to simulate cochlear filters of NH people. A linear combination of these normal filters was used to approximate a broadened filter. The linear combination coefficients can be determined by solving the following optimization problem:

$$\begin{cases} \min_{\mathbf{x}} |\mathbf{Ax} - \mathbf{b}|^2, & \text{subject to } 0 \le x_i \le 1 \\ \mathbf{A} \in \mathbb{R}^{\left(\frac{F_s}{2}\right) \times 128},\ \mathbf{x} \in \mathbb{R}^{128 \times 1},\ \mathbf{b} \in \mathbb{R}^{\left(\frac{F_s}{2}\right) \times 1} \end{cases} \quad (2)$$

where F_s is the sampling frequency of the sound and each column of the **A** matrix is the frequency response of a NH cochlear filter from 0 to $F_s/2$ Hz. The vector **x** contains 128 linear combination coefficients to approximate the broadened cochlear filter **b**. In principle, **b** can be in any shape, or be approximated using the rounded-exponential filter with coefficient p (i.e., the roex(p) filter) just like in [7].

After deriving the linear combination coefficients, we need to modulate output envelopes of 128 normal cochlear filters to the broadened cochlear filter by multiplying a proper carrier signal. The processed sound is intended for testing NH subjects, so the carrier signal must be somehow related to the normal cochlear filter. As indicated by [13][14], listening performance of NH subjects to the white noise modulated sounds sets the upper bound for performance of HI patients to normal sounds. Besides, our goal is to predict Mandarin speech intelligibility for HI patients in noisy conditions. Therefore, we used the white noise filtered by the normal cochlea filter as the carrier signal in our module.

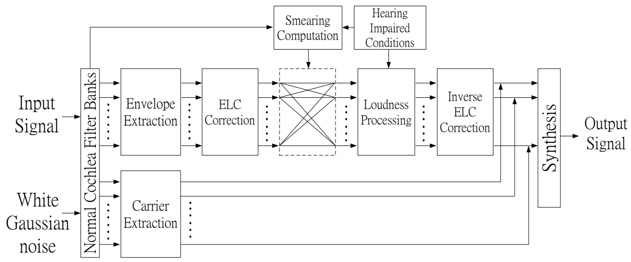

Fig. 3. Block diagram of our proposed model.

Fig. 2 shows an original STFT spectrogram and broadened spectrograms produced by our module with broadened factors (BFs) of 1.5, 3, and 6. The BF is defined as the ratio of the equivalent rectangular bandwidth (ERB) of the broadened cochlear filter to the ERB of the normal cochlear filter. Clearly, the harmonic structure only appears at low frequencies due to the nature of constant-Q cochlear filters. The low-frequency cochlear filters are much narrower than the high-frequency cochlear filters such that individual harmonics can be resolved only at low frequencies but not at high frequencies. As shown in Fig. 2, the spectrogram is more smeared with a higher BF. The harmonics can still be observed with BFs of 1.5 and 3, but almost destroyed with the BF of 6. In addition, the timbre (encoded by spectral profiles) is severely distorted due to the smearing effect. Comparing with examples shown in [7], our proposed model has stronger smearing effect than the model in [9] by using noise carriers.

2.3. Integrated model

Because both loudness and spectral smearing modules are proposed in the filter-bank framework, they are easily integrated into a unitary system as shown in Fig. 3. The ELC correction is also included in our system. Not like in [9] that the ELC correction was implemented in the STFT domain, our system corrects the loudness at center frequencies of 128 NH cochlear filters. The "Hearing Impaired Conditions" block stores each patient's personalized MALs, degrees of loudness recruitment (DLR) and the BFs. The MALs and DLR are used by the loudness module and the BFs are used by the spectral smearing module.

3. MODEL VALIDATION AND DISCUSSIONS

The model is validated by results of psychoacoustic experiments which assess speech intelligibility scores of NH subjects to the processed speech and scores of HI patients to normal speech.

In these experiments, six 25-Mandarin-word lists in [15] were used for intelligibility tests. Each Mandarin monosyllable word consists of three elements: two phonemes (initial consonant and final vowel) and a tone. The intelligibility scores were derived by the ratio of the number of correctly identified elements to the total number of elements. All of the 150 Mandarin speech signals for listening tests were produced from the website (http://stroke-order.learningweb.moe.edu.tw/home.do?rd=72) developed by the Ministry of Education of Taiwan for learning Mandarin. All sounds were downsampled with 16 kHz sampling frequency and normalized to equal power. As in [16], listening tests were conducted under five conditions: clean, in speech-shaped noise (SSN) of 0 and 4 dB SNR, in two-talker speech (TTS) interference of 3 and 7 dB SNR.

3.1. Experiments for HI patients

The MALs at several frequencies of 4 HI patients (P1, P2, P3, and P4) were measured and are listed in Table 1. In addition to MAL parameters, the BFs were also measured at three center frequencies (0.5, 1.0, 2.0 kHz) by following procedures in [10] and are shown in Table 2. Test sounds were first normalized to 65 dB SPL and linearly amplified based on the Cambridge formula [17] with each patient's corresponding MALs at center frequencies of 0.5, 1.0 and 2.0 kHz to ensure they are audible to each patient. Experiments were conducted in Tzu Chi Hospital using an AKG K702 studio headphone in an anechoic chamber. During tests, we found intelligibility scores of patient 4, who is with the least severe hearing loss, were very high in all noisy conditions. Therefore, we reduced 4 dB in all noisy conditions for patient 4 to make his intelligibility scores more distinct. The intelligibility scores of 4 patients under clean and four noisy conditions are shown in Table 3.

Table 1. MALs (in dB) of the better ear of 4 HI patients

	Age	**250 Hz**	**500 Hz**	**1000 Hz**	**2000 Hz**	**4000 Hz**
P1	27	35	40	60	70	70
P2	72	55	55	55	55	55
P3	36	50	50	40	65	90
P4	56	30	35	40	55	45

Table 2. BFs of the better ear of 4 HI patients

	500 Hz	**1000 Hz**	**2000 Hz**
P1	2.85	3.29	3.19
P2	2.6	2.79	2.57
P3	3.2	2.95	4.66
P4	3.41	1.27	3.25

Table 3. The Mandarin intelligibility scores (in %) of 4 HI patients under clean and four noisy conditions

	clean	**SSN 0**	**SSN 4**	**TTS 3**	**TTS 7**
P1	88.22	72	74.67	70	75.33
P2	80.22	53.33	56.67	55.33	60
P3	60.22	61.33	56	49.33	56.67
	clean	**SSN -4**	**SSN 0**	**TTS -1**	**TTS 3**
P4	88.67	71.33	85.33	72	75.33

3.2. Experiments for NH subjects

The second set of experiments was to assess intelligibility scores of NH subjects to processed Mandarin speech signals by our model with all personalized parameters of each patient. Two NH subjects aged between 22 and 26 were recruited for each test condition of each patient's model. The average intelligibility score and corresponding standard deviation are plotted against the original intelligibility score of each HI patient under each test condition in Fig. 4. Results for P1 to P4 are demonstrated in Fig. 4(a) to 4(d), respectively.

As shown in Fig. 4(a), the original intelligibility scores of P1 are higher than the intelligibility scores of NH subjects in clean and all noisy conditions. A possible reason is that P1 was born with hearing loss such that he might have a more robust cognitive function for recognizing Mandarin speech. In addition, Fig 4(d) shows the intelligibility scores from the P4 model are very close to the original intelligibility scores of P4, who is with the least severe hearing loss.

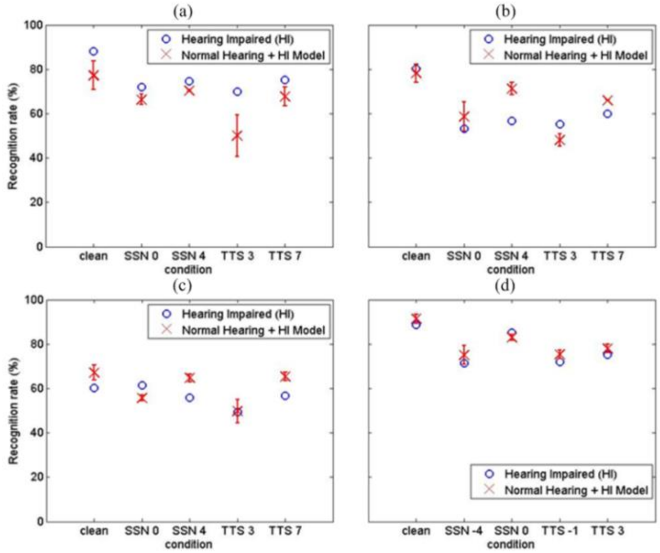

Fig. 4. The mean and the standard deviation of intelligibility scores of NH subjects to processed speech and the intelligibility scores of HI patients to original speech under clean and noisy conditions.

3.3. Statistical analysis on test results

In addition to showing the mean and the standard deviation in Fig. 4 for each test condition, we also conducted analysis of variance (ANOVA) to assess the statistical significance of the predicted scores. First, one-way ANOVA tests using each patient's original intelligibility scores and predicted scores from two NH subjects were carried out for three comparisons: (1) low SNR SSN (SSN-L) versus high SNR SSN (SSN-H), (2) low SNR TTS (TTS-L) versus high SNR TTS (TTS-H), and (3) clean versus SSN-H. The analysis results are shown in Table 4~6. These ANOVA results strengthen the results shown in Fig. 4 and four out of twelve tests turn out to be significant.

Table 4. Results of one-way ANOVA between SSN-L and SSN-H conditions

	SSN-L vs. SSN-H	
P1	$F(1, 4)=1.911$, $p = 0.239$	not significant
P2	$F(1, 4)=2.563$, $p = 0.185$	not significant
P3	$F(1, 4)=1.388$, $p = 0.304$	not significant
P4	$F(1, 4)=18.406$, $p = 0.013$	significant

Table 5. Results of one-way ANOVA between TTS-L and TTS-H conditions

	TTS-L vs. TTS-H	
P1	$F(1, 4)=2.672$, $p =0.177$	not significant
P2	$F(1, 4)=16.259$, $p = 0.016$	significant
P3	$F(1, 4)=12.381$, $p = 0.024$	significant
P4	$F(1, 4)=2.595$, $p = 0.183$	not significant

Table 6. Results of one-way ANOVA between clean and SSN-H conditions

	Clean vs. SSN-H	
P1	$F(1, 4)= 5.536$, $p = 0.078$	not significant
P2	$F(1, 4)=5.444$, $p = 0.080$	not significant
P3	$F(1, 4)=0.570$, $p = 0.492$	not significant
P4	$F(1, 4)=18.275$, $p = 0.013$	significant

These one-way ANOVA test results might not be able to offer meaningful implications due to the small sample size in each test. Hence, we also conducted two-way ANOVA tests by considering more data in each test. For each one-way ANOVA test, data collected from 3 subjects (1 HI and 2 NH subjects) were used. In contrast, the two-way ANOVA tests were carried out by treating "noise type" and "patient" as two independent variables. Table 7 shows the two-way ANOVA results for 4 patients and 3 noise types (clean, SSN-L, and SSN-H). These results demonstrate (1) highly significant differences within noise types [$F(2, 33)=29.688$, $p<0.001$] and within patients [$F(3, 32)=32.293$, $p<0.001$]; (2) no significant correlation between "noise type" and "patient" [$F=1.283$, $p=0.302$]; (3) variance within 3 noise types is comparable to variance within 4 patients [$F=29.688$ vs. $F=32.293$].

Table 8 shows the two-way ANOVA results for 4 patients and 2 noise types (TTS-L, and TTS-H). These results demonstrate (1) highly significant differences between noise types [$F(1, 22)=30.349$, $p<0.001$] and within patients [$F(3, 20)=21.394$, $p<0.001$]; (2) no significant correlation between "noise type" and "patient" [$F=1.840$, $p=0.180$]; (3) variance between two noise types is larger than variance within 4 patients [$F=30.349$ vs. $F=21.394$].

These "highly significant" two-way ANOVA test results indicate that, for each patient and in each test condition, our predicted Mandarin intelligibility scores and patient's original score are highly probably produced by the same underlying behavior.

Table 7. Results of two-way ANOVA for 4 patients and three noise types (clean, SSN-L, and SSN-H)

Variable	**p-value**	
noise	$F=29.688$, $p<0.001$	highly significant
patient	$F=32.293$, $p<0.001$	highly significant
noise*patient	$F=1.283$, $p=0.302$	not significant

Table 8. Results of two-way ANOVA for 4 patients and two noise types (TTS-L, and TTS-H)

Variable	**p-value**	
noise	$F=30.349$, $p<0.001$	highly significant
patient	$F=21.394$, $p<0.001$	highly significant
noise*patient	$F=1.840$, $p=0.180$	not significant

4. CONCLUSION AND FUTURE WORK

In this paper, we propose a hearing model, which can predict Mandarin speech intelligibility for the HI patients to a certain degree. Our model can simulate hearing-loss conditions (minimum audible level, loudness recruitment, and reduction of frequency selectivity) in the unitary filter-bank framework. The model is validated by assessing Mandarin speech intelligibility of four HI patients under clean and four noisy conditions. In the near future, we will test more patients to establish their personalized hearing models. Later, we will use the personalized model to assess speech enhancement algorithms for each HI patient. Hopefully, the assessment can provide insights to speed up the development of speech enhancement algorithms for each HI patient in the future.

5. ACKNOWLEDGEMENTS

We would like to thank research assistant Chuang-Yi Wang for her help in conducting experiments of HI patients. This research is supported by the National Science Council, Taiwan under Grant No NSC 102-2220-E-009-049, NSC 102-2220-E-303-002, and the Biomedical Electronics Translational Research Center, NCTU.

6. REFERENCES

[1] D. Byrne, and H. Dillon, "The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid," *Ear Hear.*, vol. 7, pp. 257-265, 1986.

[2] E. Villchur, "Simulation of the effect of recruitment on loudness relationships in speech," *J. Acoust. Soc. Am.*, vol. 56, pp. 1601-1611, 1974.

[3] R. S. Tyler, A. Q. Summerfield, E. J. Wood, and M. A. Fernandes, "Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners," *J. Acoust. Soc. Am.*, vol. 72, pp. 740-752, 1982.

[4] B. R. Glasberg, and B. C. J. Moore, "Effects of envelope fluctuations on gap detection," *Hear. Res.*, vol. 64, pp. 81-92, 1992.

[5] M. Florentine, S. Buus, B. Scharf, and E. Zwicker, "Frequency selectivity in normally-hearing and hearing-impaired observers," *J. Speech Hear. Res.*, vol. 23, pp. 646-669, 1980.

[6] R. S. Tyler, J. W. Hall, B. R. Glasberg, B. C. J. Moore, and R. D. Patterson, "Auditory filter asymmetry in the hearing impaired," *J. Acoust. Soc. Am.*, vol. 76, pp. 1363-1368, 1984.

[7] T. Baer, and B. C. J. Moore, ''Effects of spectral smearing on the intelligibility of sentences in the presence of noise,'' *J. Acoust. Soc. Am.*, vol. 95, pp. 1229-1241, 1993.

[8] B. C. J. Moore, and B. R. Glasberg, ''Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech,'' *J. Acoust. Soc. Am.*, vol. 94, pp. 2050-2062, 1993.

[9] Y. Nejime and B. C. J. Moore, "Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise," *J. Acoust. Soc. Am.*, vol. 102, pp. 603-615, 1997.

[10] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.*, vol. 47, pp. 103-138, 1990.

[11] B. C. J. Moore, B. R. Glasberg, R. F. Hess, and J .P. Birchall, "Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears," *J. Acoust. Soc. Am*., vol. 77, pp. 1505-1513, 1985.

[12] T. Chi, P. Ru, and S. A. Shamma, "Multi-resolution spectro-temporal analysis of complex sounds," *J. Acoust. Soc. Am.*, vol. 118, pp. 887-906, 2005.

[13] M. F. Dorman, P. C. Loizou, and D. Rainey, "Speech understanding as a function of the number of channels of stimulation for processors using sine-wave and noise-band outputs," *J. Acoust. Soc. Am.*, vol. 102, pp. 2403-2411, 1997.

[14] R. V. Shannon, F.-G. Zeng, and J. Wygonski, "Speech recognition with altered spectral distribution of envelope cues," *J. Acoust. Soc. Am.*, vol. 104, pp. 2467-2476, 1998.

[15] K. S. Tsai, L. H. Tseng, C. J. Wu, and S. T. Young, "Development of a Mandarin Monosyllable Recognition Test," *Ear Hear.*, vol. 30, pp. 90-99, 2009.

[16] J. Chen, T. Baer and B. C. J. Moore, "Effect of spectral change enhancement for the hearing impaired based on parameter values selected with a genetic algorithm," *J. Acoust. Soc. Am.*, vol. 133, pp. 2910-2920, 2013.

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