ARCHIVES OF ACOUSTICS
29, 1, 25–34 (2004)

INTELLIGIBILITY OF SPEECH PROCESSED BY A SPECTRAL CONTRAST ENHANCEMENT PROCEDURE AND A BINAURAL PROCEDURE

A. SĘK, E. SKRODZKA, E. OZIMEK and A. WICHER

Institute of Acoustics, Adam Mickiewicz University,
Umultowska 85, 60-614 Poznań, Poland
oleksek@amu.edu.pl

(*received 10 March 2003; accepted 15 September 2003*)

The study reported was undertaken as an attempt to improve the intelligibility of selected speech signals (numerical test) masked by a speech-shaped noise, by a proposed algorithm of the speech signal processing based on the spectral contrast enhancement and binaural procedure. The spectral contrast enhancement involved the transformation of the speech signal spectrum to achieve an increase in the level of the amplitude between subsequent minima and maxima of individual formants. The binaural procedure involved the determination of the interaural time difference (ITD) and the interaural intensity difference (IID) in order to select and enhance the fragments reaching the listener from the front. The intelligibility of speech processed by the algorithm was compared to that of the unprocessed speech for different spatial separations between the signal and the noise sources and for a few signal to noise ratios. The results have shown that the intelligibility of the speech presented against the noise significantly depends on the angle between the directions of the signal and the noise. The intelligibility of speech is much larger when the noise and the speech reach the observer from different directions, i.e. when their sources are spatially separated. In general, the algorithm improves the intelligibility of numerical tests presented at the background of the noise by a few percent and the greatest improvement has been observed for low signal to noise ratios. The algorithm performance was different for each hearing impaired subject with binaural hearing loss. To get an objective assessment of the algorithm performance it should be optimised and tested on a larger group of subjects for more diverse sound material.

1. Introduction

One of the most common type of hearing loss is the sensorineural hearing loss usually related to impairment of the cochlea or higher stages of the auditory system. In a large part of people suffering from this form of hearing loss it is binaural. Compensation of the sensorineural hearing loss by the enhancement of the signal by simultaneous compression of its dynamic range to the narrowed dynamic range of the auditory system is realised by a system of automatic gain control (AGC). The systems most often used are those of multichannel compressing of the signal in a few independent frequency bands.

The functioning of those systems is determined by a number of parameters, whose selection has not been unambiguously settled. The threshold of the AGC system and its gain as functions of the level of the input signal are determined by the audiogram, but the number of independent frequency bands and the times of growth and decay determining the response time of the system in each frequency band are assumed rather arbitrarily. Different authors suggest considerable variation of these time parameters, from very short ones, of an order of a few milliseconds as e.g. in the system of syllabic compression [16], to hundreds of milliseconds as e.g. in the Cambridge system [7, 14].

This does not mean that the use of a most sophisticated AGC system solves all the problems related to the compensation of the sensorineural hearing loss. Apart from the effect of loudness recruitment, the sensorineural hearing loss includes a broadening of the auditory filters bandwidth [3, 5, 6, 11]. The broadening of the auditory filters means that the frequency selectivity of the auditory system is significantly impaired. The impaired ear is not able to separate out two frequency components in the way the normally hearing ear can do. Moreover, if the auditory filter is broadened when a signal is detected against noise, the broader filter passes much more noise than a filter of normal bandwidth. Consequently, the signal to noise ratio at the output of the broader filter is much lower than that for a normal one and the signal detection is much more difficult.

The most important consequence of the broadening of the auditory filters is a considerable decrease in the intelligibility of speech, especially when presented on a background of noise [13, 17]. Speech intelligibility especially presented against a background noise, is of great practical significance and has been the main measure of effectiveness of the functioning of the hearing aid. A hearing aid equipped with an AGC system ensures the hearing of a signal by enhancing its level to a certain comfortable value, but it cannot ensure intelligibility of the contents of the sound. Its performance meets the demands of the user only when an undisturbed sound is detected. Therefore, recently, much effort is directed towards designing a hearing aid that would improve intelligibility of speech presented against the background noise.

At present, most of the work on the improvement of the hearing aid design, aimed at increasing the speech intelligibility, are concerned with digital equipments. In a digital hearing aid the input analog signal is converted to the digital form processing according to a certain algorithm (to enhance certain characteristics, perform compression, etc.) and converts it again into the analog form. Usually all these operations are performed in the real time. The most important advantage of the digital solution relative to the conventional one is that it permits a wide range of modifications of an acoustic signal depending on the algorithm applied. However, currently available hearing aids offer improved AGC characteristic but most of them do not contain any elements aimed at improving the intelligibility of speech.

In this paper we present an initial testing of a set of procedures aimed at the improvement of the speech intelligibility when the speech signals are presented at the background of a speech-shaped noise.

2. The aim of the study

The main aim of the study was the designing of an algorithm for the improvement of the intelligibility of speech presented against a speech-shaped noise based on the enhancement of the spectral contrast and binaural procedure and testing its performance. Because of the great complexity of the problem of the algorithm, its performance in real time was left for the next stage of the study devoted to the implementation of the proposed algorithm on a DSP processor of a digital hearing aid. The designed procedures were implemented in the Matlab environment.

The study is divided into three stages: recording of numerical tests in an anechoic chamber, processing of the tests with the help of the algorithm proposed for each subject and for each ear (on the basis of the audiograms) and assessment of the intelligibility of the tests processed by the proposed algorithm. The assessment of the algorithm performance was made on the basis of the relationships between the intelligibility of speech and the signal to noise ratio, enhanced spectral contrast and spatial separation of the sources of speech and noise signals determined for subjects with binaural of approximately the same sensorineural hearing losses.

3. The algorithm

If a complex sound signal is passed through a system of overlapping band-pass filters, the spectral contrast (understood as the difference between the maximum and the neighbouring minimum in the spectrum) is reduced [1, 2, 15]. With reference to the auditory filters this phenomenon is known as the spectral smearing [9]. The degree of the spectral smearing (or the decrease in the spectral contrast) is larger for broader filters. If, for instance, a signal representing a vowel is passed through a system of normal auditory filters, at the output of the filters (i.e. in the excitation pattern corresponding to the envelope of the basilicar membrane displacement) the representation of the vowel formants is relatively good [4, 6]. If the filters are twice or three times wider (which is rather often met in subjects with sensorineural hearing loss), the representation of particular formants in the excitation pattern is much poorer. The spectral smearing is directly related to deterioration in the speech intelligibility. The lack of representation of the formants structure in the excitation pattern implies difficulties in signal recognition and hence identification of the sound.

The algorithm proposed is based on the following two procedures: enhancement of the spectral contrast and binaural procedure. The procedure of the spectral contrast enhancement described in detail by OZIMEK *et al.* [8], is based on the assumption that the broadening of the auditory filters is responsible for the decreased dynamic range of the spectrum of the signal. This means that the spectral maxima of such a signal maintain their height but the minima separating them are not so deep as those in the spectrum processed by normal auditory filters; this is the source of the difficulty in the intelligibility of the speech. The aim of the procedure proposed was to enhance the spectral contrast by deepening the spectral minima so as to obtain at the output of the broader auditory filters an excitation pattern of the basilicar membrane similar to that caused by an unprocessed signal in the normal hearing ear. Apart from the enhancement of the spectral contrast, the procedure employed individual adjustment of the speech signal to the dynamic range of the auditory system of the subjects based on the audiogram and the shape of the auditory filters.

The binaural procedure of the speech intelligibility improvement was based on the analysis of the interaural time difference (ITD) and interaural intensity difference (IID) between the signals reaching the subject. The procedure, described in detail by SĘK *et al.* [12], involves simultaneous processing of the signals reaching two microphones of the dummy head imitating two hearing aids of a person with binaural hearing loss. If the signals recorded at the two microphones are the same, they are treated as coming from a source directly in front of the head, which usually is the location of the speech sound that people tend to face when listening to. Thus, the signals recorded as the same by the two microphones are enhanced. For a given spectra-temporal portion of the input signal, the binaural procedure determines whether the portion comes from the position in front of the head, from the left or from the right side. If there was a phase difference between the signals reaching the two microphones, which meant that the signal came from, let's say, the left side, only the weaker signal in the right channel was enhanced. A subsequent step of the procedure was to enhance the signals in each of the channels of transmission according to the audiograms of the subjects. The enhancement at this stage was meant to adjust the signal's dynamic range to the dynamic range of the each ear. In the experiment, the spectral contrast enhancement algorithm followed the binaural procedure. The latter was performed independently for the left and right ear and concerned only the signals preliminary classified as coming from the front of the head.

4. The method

4.1. Sound recording

The sound materials used were standard numerical tests [10] presented against a background of the speech-shaped noise. Ten tests of ten numerals each were recorded in an anechoic chamber with the use of a dummy head (Neuman). The subsequent numbers were presented at 4 second intervals; the speech-shaped noise was a continuous signal. The spectral density of the noise was constant up to 1000 Hz, and decreased by 12 dB per octave above this frequency. The dummy head was placed in front of the speech signal at a distance of 3 m from it. The source of the noise was placed in the direction of the speech signal or in the direction of 60° with respect to the line joining the head and the speech source. Five different signal to noise ratios were used: 0, -3, -6, -9 and -12 dB, obtained by appropriate attenuation of the speech signal. The level of the speech-shaped noise was constant and equal to 70 dB SPL.

The signals of the speech and noise were generated and recorded by the Tucker Davis Technology System 3, at the sampling rate of 24 kHz and 24-bit resolution. One of the channels of the DA converter of the system, TDT-RP2, generated the speech signals recorder earlier on a hard disk, while the second channel generated the speech-shaped noise in the real time. The signals were fed to the digitally controlled attenuators, TDT-PA5, in order to adjust their levels. Then, the signals were supplied to the power amplifiers Pionier A300 and ZG80 loudspeakers. Two AD channels TDT-RP2 recorded the signals reaching the left and the right ear of the dummy head. Prior to the AD conversion the signals from each ear of the dummy head were amplified by 25 dB by the microphone amplifiers TDT-MA2 in order to use the full dynamics of the converters, and stored on a hard disk.

The algorithm proposed processed the material recorded and the subjects with binaural hearing loss assessed its intelligibility.

4.2. Preparation of the tests and the experiment scenery

As mention above, the recordings were done using a continuous speech-shaped noise. However, to avoid the loudness adaptation or fatigue while listening, as well as to make the calculation much faster, parts of the recorded signals were selected. The selected parts started 200 ms before and lasted up to 200 ms after each word. The subjects were exposed to numerical tests in which individual words were separated by a 4 second silence, in which they were asked to give their response. The subjects were asked to listen to 5 randomly selected numerical tests, which means that for five applied signal to noise ratios and two positions of the source of noise, the intelligibility of speech was assessed on the basis of 50 words.

The subjects with a binaural hearing loss were asked to listen to three types of signals: (1) numerical tests with the compensation of the hearing loss according to the audiogram for each ear of each subject, (2) numerical tests with the compensation according to the audiograms and processed by the binaural procedure, (3) numerical tests with the compensation according to the audiograms processed by the binaural procedure and spectral contrast enhancement algorithm. The total level of the sound material was chosen on the basis of the preliminary exposure to ensure a comfortable hearing in each ear. The experiments were performed in sound attenuated booths and the subjects were asked to give answers on a special form.

4.3. The subjects

The subjects were three persons aged 25–50, diagnosed with a binaural hearing loss; they were paid for their services. The audiograms drawn separately for each ear revealed a medium-size hearing loss in the range up to 6 kHz. The curves describing the bone and air conduction were consistent within the range of 5 dB. Apart from a number of diagnostic tests prior to the experiments, the subjects took part in training sessions to learn what is expected of them. The experiment lasted not longer than 2 hours a day. The tests were presented in 15-minute cycles separated by 5-minute intervals.

5. Results and analysis

The results obtained for the subjects with a binaural hearing loss are presented in Figs. 1 and 2. The diagrams show the percent of the correct answers as a function of the signal to noise ratio, the mean values and standard deviations. The three panels give the results corresponding to the above-described three types of tests: (1) with compensation on the basis of the audiograms for each ear, (2) with compensation and after binaural procedure processing, (3) with compensation, binaural procedure and the spectral contrast enhancement algorithm.

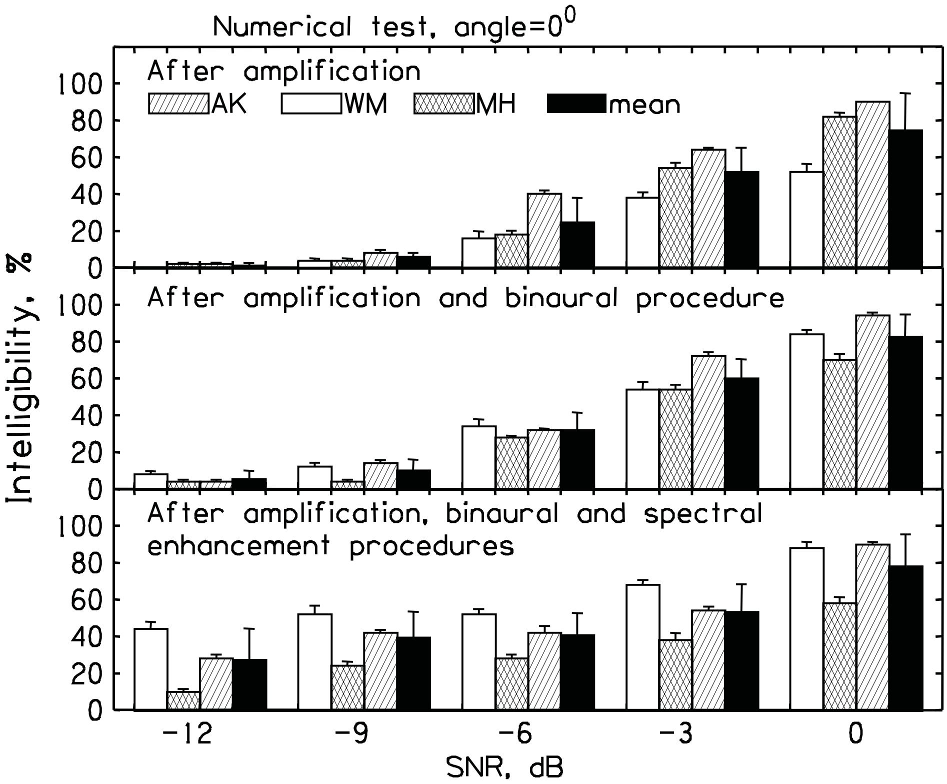

Fig. 1. The intelligibility of speech (numerical tests) for different signal to noise ratios and for the sources of speech and noise aligned in the same directions.

Figure 1 presents the results obtained for the sources of speech and noise arranged in the same direction, i.e. for the angle of 0°, while Fig. 2 shows the same for the source of speech positioned in front of the subject (0°) and the source of noise in the direction of an angle of 60° to the line joining the source of speech and the head. For both arrangements, the intelligibility of the numerical tests was significantly influenced by the signal to noise ratio. The influence was more significant when the sources of the speech and the noise were at different directions. A comparison of the mean intelligibility of the

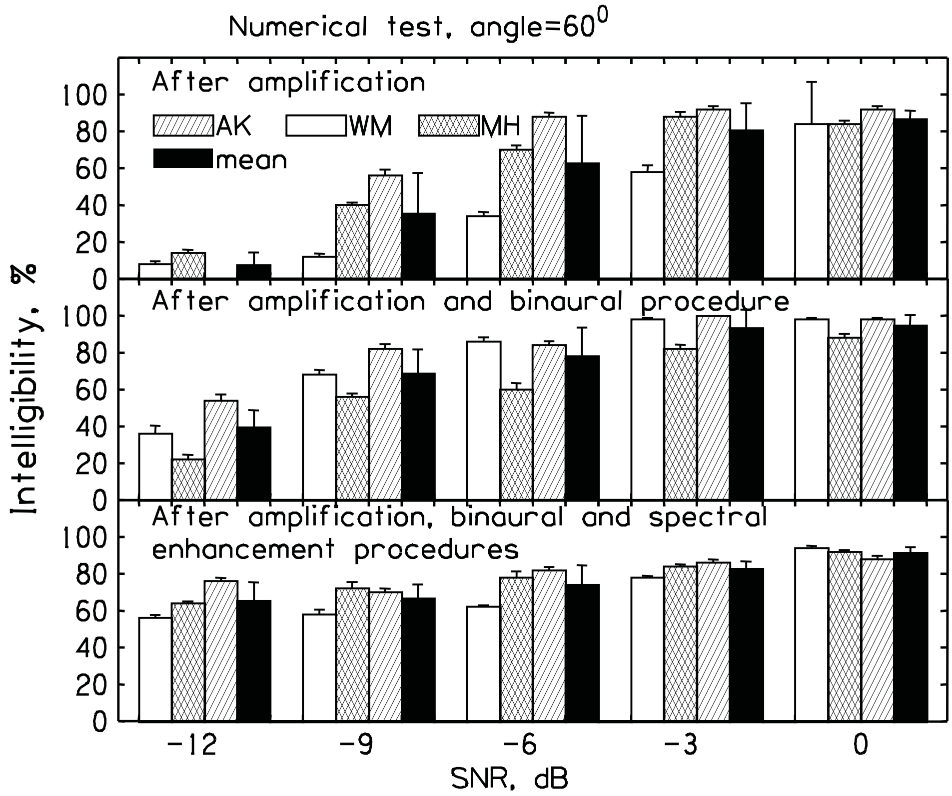

Fig. 2. The intelligibility of speech (numerical tests) for different signal to noise ratios and for the sources of speech and noise spatially separated by the angle of 60°.

three types of tests used, i.e.: (1) with compensation on the basis of the audiograms for each ear, (2) with compensation and after the binaural procedure processing, (3) with compensation, binaural procedure and the spectral contrast enhancement algorithm, has shown that the speech intelligibility was considerably higher for the processed tests (2) and (3), especially when the speech and the noise came from different directions. The increase in the intelligibility was highly individualised and most significant for the subject WM, while it was not so profound for the other subjects. The standard deviations of the mean intelligibility averaged for all subjects are rather high, so the observed increase in the speech intelligibility falls within the range of the error of the measurement. The processing applied in the test (3) was most effective for low signal to noise ratios, whereas the processing applied in test (2) brought the greatest advantage to the subjects for low and medium signal to noise ratios.

The results presented in Figs. 1 and 2 were subjected to the analysis of variance ANOVA, taking into account the following factors: the signal to noise ratio, the spatial separation of the sources of speech and noise and the tests processing applied. As expected, the effect of the signal to noise ratio was highly statistically significant [$F(4,8) = 446.56$, $p < 0.0001$], which confirmed a significant increase in the speech intelligibility with increasing signal to noise ratio. The effect of the spatial separation of the sources of speech and noise was also statistically significant [$F(1,2) = 93.99$, $p = 0.010$]. However, the effect of the binaural procedure was not statistically significant [$F(1,2) = 4.25$, $p = 0.102$]. The mean values characterising the speech intelligibility obtained for each signal to noise ratio and each spatial arrangement of the sources of speech and noise were higher for the tests subjected to the spectral enhancement procedure. However the significant intersubject scatter of the results was responsible for the statistical insignificance of this procedure. The mean values characterising the speech intelligibility, averaged over all signal to noise ratios, for the sources of speech and noise arranged in the same direction and for the three types of tests 1, 2 and 3, were 32%, 38% and 48%, respectively. For the spatial separation of the sources of speech and noise by 60°, the mean values describing the speech intelligibility for the three types of tests were 55%, 74% and 78%, respectively. The effects of the signal to noise ratio and spatial separation of the sources of speech and noise were statistically significant.

6. Conclusions

The results of the study have indicated a marked dependence of the speech intelligibility on the spatial separation of the sources of speech and noise; when the speech and noise come from different directions, the speech intelligibility was significantly higher (see Figs. 1 and 2). The application of the proposed binaural procedure showed a speech intelligibility improvement by a few percent on average, especially for low signal to noise ratios. On the other hand, the application of the spectral contrast enhancement brought an improvement in the speech intelligibility for low and medium signal to noise ratios exceeding about10%. The mean values characterising the speech intelligibility obtained for different signal to noise ratios and for different procedures of the signal processing used are shown in Table 1.

Table 1. Mean values describing the speech intelligibility for the three types of tests applied and for different signal to noise ratios.

	Signals after amplification only	Signals after amplification and binaural procedure	Signals after amplification binaural and spectral enhancement procedure
SNR = −12 dB	4.33%	21.33%	46.33%
SNR = −9 dB	20.67%	39.33%	54.33%
SNR = −6 dB	44.00%	54.00%	59.67%
SNR = −3 dB	65.33%	76.67%	69.67%
SNR = 0 dB	80.67%	88.33%	85.67%

An improvement in the speech intelligibility as the result of the application of the two procedures (tests (2) and (3)) was observed for two spatial arrangements of the sources of speech and noise (0 and 60°). However, the improvement was better when the speech and the noise came from different directions (separation at 60°). The results revealed a significant intersubject scatter. A few times the application of the procedures brought a decrease in the speech intelligibility. However, irrespective of the scatter of the results, in general, an improvement in the speech intelligibility was observed as the result of the application of the proposed procedures. This is true for the majority of the experimental conditions, which confirms a correct functioning of the algorithm proposed.

The conclusions following from the results of our experimental study are as follows:

- The intelligibility of numerical tests presented against a background of noise depends significantly on the spatial arrangement of the sources of speech and noise, and is greater when these sources are spatially separated.
- The tests processing by the binaural procedure have improved their intelligibility only by a few percent, in particular for low signal to noise ratios. The tests processing by the binaural procedure and spectral contrast enhancement has also brought a few percent improvement in the speech intelligibility, also higher for low signal to noise ratios.

As follows from the analysis of the results, a possible implementation of the algorithm in the digital hearing aids would require a further study for a larger test material, greater number of sources and types of noise and for a larger number of subjects.

Acknowledgments

This work was supported by the Polish Committee for Scientific Research, grants no. 8 T11 E 01717 and 4 T11E 014 25.

References

[1] T. Baer, B. C. J. Moore, *Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech*, J. Acoust. Soc. Am., **95**, 2277–2280 (1994).

[2] T. Baer, B. C. J. Moore, *Effects of spectral smearing on the intelligibility of sentences in the presence of noise*, J. Acoust. Soc. Am., **94**, 1229–1241 (1993).

[3] J.R. Dubno, D. D. Dirks, *Filter characteristics and consonant recognition for hearing-impaired listeners*, J. Acoust. Soc. Am., **85**, 1666–1675 (1989).

[4] A. Faulkner, V. Ball, S. Rosen, B. C. J. Moore, A. J. Fourcin, *Speech pattern hearing aids for the profoundly hearing impaired: Speech perception and auditory abilities*, J. Acoust. Soc. Am., **91**, 2136–2155 (1992).

[5] J. W. Hall, J. H. Grose, *Spectro-temporal analysis and cochlear hearing impairment: Effects of frequency selectivity, temporal resolution, signal frequency and rate of modulation*, J. Acoust. Soc. Am., **85**, 2550–2562 (1989).

[6] B. C. J. Moore, *Perceptual consequences of cochlear damage*, Oxford University Press, 1995.

[7] B. C. J. Moore, J. S. Johnson, T. M. Clark, V. Pluvinage, *Evaluation of a dual-channel full dynamic range compression system for people with sensorineural hearing loss*, Ear Hear., **13**, 349–370 (1992).

[8] E. Ozimek, A. Sęk, A. Wicher, E. Skrodzka, J. Konieczny, *Identification of spectrally enhanced Polish vowels by hearing-impaired subjects*, Acta Acustica, submitted (2003).

[9] R.D. Patterson, B. C. J. Moore, *Auditory filters and excitation patterns as representations of frequency resolution*, [in:] Frequency selectivity in hearing, B. C. J. Moore [Ed.], Academic, London 1986.

[10] A. Pruszewicz, G. Surmanowicz–Demenko, T. Wika, L. Richter, *Polski test liczbowy i słowny do badania słuchu oraz testy do treningu słuchowego*, Marke-Med, Poznań 1999.

[11] A. Sęk, B. C. J. Moore, *The critical modulation frequency and its relationship to auditory filtering at low frequencies*, J. Acoust. Soc. Am., **95**, 2606–2615 (1994).

[12] A. Sęk, E. Ozimek, E. Skrodzka, A. Wicher, *Intelligibility of speech presented against a speech-shaped noise for normal and hearing-impaired subjects*, submitted (2003).

[13] P. G. Stelmachowicz, D. E. Lewis, W. J. Kelly, W. Jesteadt, *Speech perception in low-pass filtered noise for normal and hearing-impaired listeners*, J. Speech Hear. Res., **33**, 290–297 (1990).

[14] M. A. Stone, *Spectral enhancement for the hearing impaired*, University of Cambridge, 1995.

[15] M. ter Keurs, J. M. Festen, R. Plomp, *Effect of spectral envelope smearing on speech reception. II*, J. Acoust. Soc. Am., **93**, 1547–1552 (1993).

[16] J. Verschuure, T. T. Prinsen, W. A. Dreschler, *The effect of syllabic compression and frequency shaping on speech intelligibility in hearing impaired people*, Ear Hear., **15**, 13–21 (1994).

[17] P. M. Zurek, C. Formby, *Frequency-discrimination ability of hearing-impaired listeners*, J. Speech Hear. Res., **24**, 108–112 (1981).