Simulation of increased masking in sensorineural hearing loss for a preliminary evaluation of speech processing schemes

Dakshayani S. Jangamashetti (1), Alice N. Cheeran (2), Pandurangarao N. Kulkarni (3), and Prem C. Pandey (4)

(1) Basaveshwar Engineering College, Bagalkot 587102, Karnataka, India
(2) Veermata Jijabai Technological Institute, Matunga Mumbai 400 019, Maharashtra, India
(3, 4) Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India

PACS: 43.66 Ts and 43.66 Sr

ABSTRACT

Sensorineural loss is characterized by increased hearing threshold, reduction in the dynamic range of hearing and loudness recruitment, and increased temporal and spectral masking, resulting in degraded speech perception. Several techniques including spectral contrast enhancement, multi-band frequency compression, and dichotic binaural presentation have been investigated for reducing the adverse effects of increased masking. Assessment of speech processing techniques and optimization of processing parameters involves listening tests on hearing-impaired listeners. These tests are time consuming and may cause a fatigue, particularly in elderly subjects. A simulation of hearing loss, by processing the speech signal through a model of the loss characteristics, is useful in conducting the listening tests on normal-hearing subjects, for a preliminary evaluation of the schemes and particularly for selecting the processing parameters. The present study used addition of broad-band noise, band-limited to speech frequency range, at a specific SNR with respect to short-time (10 ms) energy of the signal. Different levels of loss were simulated by varying the SNR. In this simulation, no noise gets added during silence segments. Listening tests to assess the loss simulation were conducted using three types of test material: vowel-consonant-vowel (VCV) utterances, phonetically balanced word lists, and modified rhyme test. Recognition score from subject responses was used as a measure of speech intelligibility and response time was used as a measure of load on the perception process. For all the three test materials, decrease in the recognition scores and increase in response times for normal-hearing subjects showed the same pattern as the corresponding results for subjects with moderate-to-severe sensorineural loss. A relative information transmission analysis of the stimulus-response confusion matrices for VCV utterances showed that the simulated loss did not affect reception of voicing and nasality features and it had maximum adverse effect on the reception of place and duration features, indicating that the addition of broadband noise with constant SNR with respect to short-time signal energy simulated an increased spectral and temporal masking.

INTRODUCTION

Sensorineural hearing loss is characterized by increased hearing thresholds, reduced dynamic range of hearing and loudness recruitment, and increased temporal and spectral masking resulting in a degraded speech perception [1 – 4]. To reduce the effect of spectral and temporal masking, investigations based on spectral contrast enhancement [5 – 7], multi-band frequency compression [8, 9], and dichotic binaural presentation [10 – 13] have been reported. Listening tests to evaluate the schemes for speech processing at various processing conditions are time consuming and tedious and may cause fatigue. Hence it is difficult to test processing schemes with several combinations of processing parameters directly on the hearing-impaired subjects. Also, there are difficulties in having a large number of volunteering subjects with sensorineural hearing loss, willing to participate in the experiments. A simulation of hearing loss, by processing the speech signal through a model of the loss characteristics, is useful in conducting the listening tests on normal-hearing subjects, for a preliminary evaluation of the schemes and particularly for selecting the processing parameters. Different types of simulation [14-18] have been used to characterize the different aspects of impairment. Villchur [14] simulated loudness recruitment by splitting the speech signal into three frequency bands and applying dynamic range expansion with different ratios in each band. The simulation was tested on four subjects with severe acquired unilateral hearing loss. Subjects judged the simulated stimuli presented to the normal ear to be similar to the unprocessed stimuli presented to the impaired ear. Moore and Glasberg [16] split the input signal into thirteen bands and processed the envelope in each band to simulate loudness recruitment.
In [17], the reduced frequency resolution of the auditory system was simulated by smoothing the envelope of the squared short-time fast Fourier transform (FFT) by convolving it with a Gaussian-shaped filter. The effects of reduced frequency selectivity were simulated by spectral smearing, using the overlap-add method [6]. The smearing of the spectra of the stimuli evoked similar response in normal-hearing persons as the broadened auditory filters of the hearing-impaired persons. Nejime and Moore [18] investigated a scheme for simulating the combined effects of elevated threshold, loudness recruitment, and reduced frequency selectivity. Loudness recruitment was simulated by filtering the speech stimuli into a number of frequency bands, and raising the temporal envelope of the waveforms at the output of each filter to a power greater than one. The effect of reduced frequency selectivity was simulated by smearing the short-term power spectrum of the stimuli in such a way that the excitation pattern produced in a normal ear resembled that of an impaired ear to the unprocessed stimuli.

Simulation of sensorineural loss has often been carried out, by employing different types of masking noise. In several studies [16, 21, 22], elevated thresholds were simulated by adding broad-band noise. In a study conducted by Dubno and Schaefer [4], hearing loss was simulated using spectrally shaped broad-band noise and hearing threshold of normal-hearing subjects was matched with hearing-impaired subjects. Although the two results were similar for consonant recognition, the frequency selectivity of hearing impaired listeners was poorer than normal-hearing subjects with simulated hearing loss. The results of frequency resolution, temporal resolution and speech recognition obtained from hearing-impaired persons were used to predict the results on noise-masked normal-hearing listeners [19]. The prediction was accurate for frequency resolution and speech recognition. In a study to determine the minimum spectral contrast required for vowel identification, Leek et al. [20], used broad-band noise to simulate elevated thresholds in the range of 72 – 75 dB in normal-hearing subjects.

Out of the various methods reported for simulation of masking, addition of noise is the simplest and has been shown to simulate elevated thresholds as well as increased masking, with degradation in speech perception being related to SNR. The objective of the present study is to investigate a scheme for simulating increased masking effect in sensorineural hearing loss. The study used addition of broad-band noise, band-limited to speech frequency range, at a specific SNR with respect to short-time (10 ms) energy of the signal. In this simulation, no noise gets added during silence segments. Different levels of loss were simulated by varying the SNR. The effect of simulation was evaluated by conducting listening tests on normal-hearing subjects with simulated masking effect and on hearing-impaired subjects with moderate sensorineural hearing loss.

**LISTENING TESTS**

Speech intelligibility tests are conducted using different types of materials such as words, nonsense syllables, and sentences [23]. To study the perceptual confusion, Miller and Nicely [24] used 16 consonants /p, t, k, f, ð, s, l, b, d, g, v, ð, z, ð, m, and n/ in CV contexts with vowel /a/. Two of the commonly used intelligibility tests at word level are diagnostic rhyme test (DRT) [25] and modified rhyme test (MRT) [26]. Both are used to assess the consonant perception in consonant-vowel-consonant (CVC) context. In DRT, only the initial consonants are tested. Both the initial and the final consonants are tested in MRT. Another test, often used at word level, uses a set of phonetically balanced (PB) words, mostly presented for open set response. Kryter [27] compared the recognition scores obtained by MRT and PB word test by conducting listening tests on eight normal-hearing subjects and speech spectrum shaped noise as the masker. The scores for MRT and PB test with 200-word list were similar, but 25 % lower score was observed for 1000-word PB test. In a multiple choice listening test, the response time provides a measure of the load on the perception process, and a decrease in the response time indicates an improved listening condition [6, 28, 29].

In the present study, listening tests to assess the loss simulation were conducted using three types of test material: vowel-consonant-vowel (VCV) utterances with vowel /a/ and twelve consonants, phonetically balanced (PB) word lists, and 300 monosyllabic words in CVC form for modified rhyme test (MRT). All the tests were conducted with the test material presented at the most comfortable listening level of the individual subject. Recognition score from subject responses was used as a measure of speech intelligibility and response time was used as a measure of load on the perception process.

The PB tests were conducted with three sets of phonetically balanced monosyllabic words, with each set having 50 to 60 words. All the words had approximately the same intensity. The tests were conducted on seven normal-hearing subjects (age: 18 – 28 years) with the masker added at the SNR values of ∞ (no noise), 3, 0, -3, -6, and -9 dB and on 13 hearing-impaired subjects (age: 19 – 59 years) with moderate to severe sensorineural hearing loss [30].

The test material for MRT consisted of 50 sets of monosyllabic words of consonant-vowel-consonant (CVC) form. Each set consisted of six words with a vowel in the middle and either initial or final consonant remaining the same and the other consonant being different. Each of the words was preceded by a carrier phrase “would you write ---- -?”. All the 300 words (i.e. 50 sets × 6 words in each set) were arranged in 6 test lists of 50 words each. The presentation level was set at the most comfortable listening level as selected by the individual listener. The test was conducted on six normal-hearing subjects (age: 35 – 45 years), with the masker added at the SNR values of ∞, 6, 3, 0, -3, -6, -9, -12, and -15 dB. The test was also conducted on 11 subjects (age: 18 – 56 years) with moderate bilateral sensorineural hearing loss.

The VCV tests involved identification of 12 consonants, as shown in Table 1, in VCV context with vowel /a/. Five normal-hearing subjects (age: 20 – 37 years) and five subjects (age: 32 – 61 years) with sensorineural loss participated in these tests [31]. For normal-hearing subjects, masker was added at the SNR values of 6, 3, 0, -3, -6, -9, -12, and -15 dB. The normal-hearing subjects responded for 60 presentations (12 stimuli × 5 repetitions × 9 SNR values) and the hearing-impaired subjects responded for 540 presentations (12 stimuli × 5 repetitions × 9 SNR values) and the hearing-impaired subjects responded for 60 presentations (12 stimuli × 5 repetitions). The performance measures used were response times, recognition score, and relative information transmission for various consonantal features.

**RESULTS**

**Results of perceptually balanced (PB) test**

The PB test results for the response times and the recognition scores, averaged across the seven normal-hearing subjects, are given in Table 2. Response times increased from 2.09 s under no noise to 2.83 s at -9 dB SNR, and the recognition scores decreased from 99.8 % to 23.9 %. For the hearing-impaired subjects, the response time ranged from 2.1 s to 6.6 s.
were -7 and -9 dB, respectively. Masking was not effective in 5. For the place and duration features, the equivalent SNR Table 4), to obtain equivalent SNR values as given in Table 5 shows the recognition score and relative information transmitted for different features. Averaged recognition score decreased from 100% at no-noise condition to 70% at -15 dB SNR condition. The overall information transmitted [24] decreased from 100% at no-noise condition to 70% at -15 dB SNR condition. The decrease in the information transmission was smaller than that in the recognition scores, indicating that the reception errors were not randomly distributed, but may be distributed in accordance with feature groupings. This necessitated a study of information transmission for various consonantal features (as given in Table 1). Values for relative information transmitted are also given in Table 4. The recognition of voicing and nasality features was modestly affected at higher levels of masking. This is in conformity with the fact that these two are the most robust of the consonantal features. Compared to voicing, reception of manner (stop/frication/nasality) was more

Listening tests were also conducted on 5 hearing-impaired subjects having moderate-to-severe sensorineural hearing loss. Table 5 shows the recognition score and relative information transmitted, averaged across the subjects. These scores were matched (by using linear interpolation) with the corresponding average scores for simulated loss (as given in Table 4), to obtain equivalent SNR values as given in Table 5. For the place and duration features, the equivalent SNR were -7 and -9 dB, respectively. Masking was not effective in simulations the effect of hearing loss on reception of voicing and only moderately effective in that of nasality and manner. However, these features are known to be not susceptible to the adverse effects of temporal and spectral masking.

**DISCUSSION**

Objective of the study was to investigate a technique of simulating increased masking in sensorineural hearing loss by adding broad-band noise to the test material, keeping
constant SNR on short-time (10 ms) basis. Investigations involving listening tests on normal-hearing subjects with simulated loss and hearing-impaired subjects with moderate sensorineural hearing loss. Three types of test materials were used: PB words, MRT words, and nonsense VCV syllables with vowel /a/. The speech perception degraded with decrease in SNR for all the test materials. Decrease in SNR also resulted in increased response time indicating an increased load on the speech perception process. Figure 1 shows the recognition scores for the three types of test material (i.e. VCV utterances, PB words, and MRT words) at different SNR values. The scores obtained for PB words were generally lower than those obtained for VCV and MRT. For PB words, average recognition score of hearing-impaired subjects was 63% and it matched with the recognition score for normal-hearing subjects for SNR of -3 dB. With MRT, the average score for the hearing impaired subjects was 61.3% and it matched that for the normal-hearing subjects at -9 dB SNR. For VCV test, the average recognition score of hearing-impaired subjects was 81%, with an equivalent SNR of -9 dB. A relative information transmission analysis of the stimulus-response confusion matrices for VCV utterances showed that the simulated loss did not significantly affect the reception of voicing and nasality features and it had maximum adverse effect on the reception of place and duration features.

CONCLUSION
The results from the listening tests indicate that addition of broad-band noise with constant SNR with respect to short-time signal energy simulated an increased spectral and temporal masking. The simulation may be useful for a preliminary evaluation of speech processing techniques for optimizing the processing parameters before conducting listening tests with the hearing-impaired listeners.

ACKNOWLEDGEMENT
The research is partly supported by a project grant under the National Programme on Perception Engineering, sponsored by the Department of Information Technology, MCIT, Government of India.

REFERENCES
20. M. R. Leek, M. F. Dorman, and Q. Summerfield, “Minimum spectral contrast for vowel identification by...