Paper No. 980

Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

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 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 with vowel /a/ and twelve consonants, phonetically balanced (PB) word lists, and modified rhyme test (MRT). 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 associated with loudness recruitment, increased temporal and spectral masking associated with reduced temporal and spectral resolution, 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 in the normal ears of four subjects with severe acquired unilateral hearing loss. Subjects judged the simulated stimuli presented to the normal ear, to be similar to 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 of 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. Recognition score, response time and relative information transmitted were compared.

LISTENING TESTS

In speech intelligibility test, test materials such as words, nonsense syllables, and sentences are used [23]. To study the perceptual confusion, Miller and Nicely [24] used 16 consonants / p, t, k, f, θ, s, ʃ, b, d, g, v, δ, z, ʒ, 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] conducted experiments to compare the recognition scores obtained by MRT and PB word test. Eight normal-hearing subjects participated in the listening tests. Speech signal was added to random noise with its spectrum shaped to the long term spectrum of the speech, at various SNR values. No difference in the recognition scores was reported between MRT (300 words in six test list) and PB test (200 phonetically balanced word in five test list). However, the difference of 25% was observed (60% for 1000 word PB test and 85% for MRT) when MRT scores were compared with recognition scores of PB test employing 1000 word.

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 for modified rhyme test (MRT) in CVC form. 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 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.

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 12 subjects (age: 17 – 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 hearing loss participated in these tests. For normal-hearing subjects, masker was added at the SNR values of 6, 3, 0, -3, -6, -9, -12, and -15 dB. Thus the normal-hearing subjects responded for a total of 540 presentations (12 stimuli × 5 repetitions × 9 SNR values). The total number of presentations for hearing-impaired subjects was 60 (12 stimuli × 5 repetitions). The performance measures used were response times, recognition score, and relative information transmission for various consonantal features.

Table 1. Feature groupings of the 12 consonants in VCV utterances

Features	Consonant groups
Voicing(2)	Unvoiced: / *p t k s f* /
	Voiced: / *b d g m n z v* /
Place(3)	Front: / *p b m f v* /
	Middle: / *t d n s z* /
	Back: / *k g* /
Manner(3)	Oral stop: / *p b t d k g* /
	Fricative: / *s z f v* /
	Nasals: / *m n* /
Nasality(2)	Oral: / *p b t d k g s z f v* /
	Nasal: /*m n* /
Frication(2)	Stop: / *p b t d k g m n* /
	Fricative: / *s z f v* /
Duration(2)	Short: / *p b t d k g m n f v* /
	Long: / *s z* /.

RESULTS

Results of 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 with an average of 3.05 s, and the recognition score ranged from 20.6 % to 90.1 % with an average of 62.7 %.

Table 2. PB test results: response time (RT, s) and recognition score (RS, %), averaged across 7 normal-hearing subjects.

	SNR (dB)					
	∞	3	0	-3	-6	-9
RT	2.09	2.16	2.29	2.36	2.66	2.83
RS	99.8	83.6	78.8	66.3	39.9	23.9

Results of MRT

The results of MRT, averaged across the six normal-hearing subjects, are given in Table 3. The mean response time and the mean recognition score changed from 2.64 s and 97.1 % at no noise to 3.45 s and 45.3 % at -15 dB SNR, respectively. For the hearing-impaired subjects, response times ranged from 3.57 s to 4.10 s, with an average of 3.80 s. The average recognition score was 61.3 %, matching with the recognition score of normal-hearing subjects at -9 dB SNR.

Results of tests with VCV utterances

Table 4 gives response times, percentage recognition scores, and relative information transmitted, averaged across the five normal-hearing subjects for different SNR values. Average response time increased from 1.89 s under no-noise condition to 2.61 s at -15 dB SNR, indicating that, addition of noise increased load on perception process in receiving the auditory message.

Table 3. MRT results: response times (RT, s) and recognition score (RS, %), averaged across the six normal-hearing subjects.

	SNR (dB)								
	∞	6	3	0	-3	-6	-9	-12	-15
RT	2.64	2.83	3.09	3.25	3.35	3.38	3.40	3.44	3.45
RS	97.1	92.8	90.3	83.6	75.7	69.5	61.4	54.9	45.3

Table 4. VCV results: response time (RT, s), recognition score (RS, %), relative information transmitted (%) for overall (Ov) and feature groupings: voicing (Vo), place (Pl), manner (Mn), nasality (Na), frication (Fr) and duration (Du), averaged across 5 subjects

	SNR (dB)								
	∞	6	3	0	-3	-6	-9	-12	-15
RT	1.89	2.10	2.24	2.22	2.33	2.24	2.32	2.49	2.61
RS	100	96.0	94.0	93.0	88.7	85.5	81.0	74.0	64.5
Ov	100	96.0	95	93	89	87	83	77	70
Vo	100	99	99	100	99	100	98	95	91
Pl	100	95	91	84	73	62	50	37	29
Mn	100	91	89	92	86	83	76	70	58
Na	100	100	100	100	100	100	100	95	86
Fr	100	85	82	87	77	72	61	52	38
Du	100	95	92	85	77	64	50	35	24

Averaged recognition score decreased from 100% at no-noise condition to 65 % at -15 dB SNR condition. Overall information transmitted 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 and it indicated 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) and the values are also given in Table 4. The recption 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 the most robust of the consonantal features. Compared to voicing, reception of manner (stop/frication/nasality) was more adversely affected, decreasing to 58 % at -15 dB SNR. The decreases for frication, place, and duration features were even larger. At -15 dB SNR, the relative transmission of place and duration features decreased to 29 % and 24 %, respectively. Degradation in perception of these features indicated that the spectral cues for the place feature and temporal cues for the duration feature were masked by the addition of noise, and the severity of masking increased with a decrease in SNR.

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 which are given in able 5. For the place and duration features, the equivalent SNR were -7 and -9 dB, respectively. Making was not effective in simulating 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.

Table 5. VCV test: equivalent SNR giving the same score as the average score for the hearing-impaired (Avg. H.I.) subjects, for recognition score (RS) and relative information transmitted for different features.

	RS (%)	Relative information transmitted (%)						
		Ov	Vo	Pl	Mn	Na	Fri	Du
Avg. H.I.	81	84	85	58	71	94	56	49
Eqt. SNR	-9	-8	< -15	-7	-12	-12	-11	-9

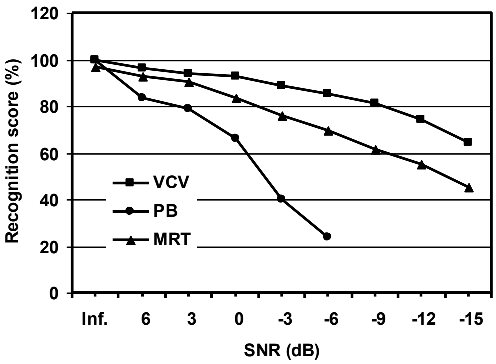

Figure 1. Recognition scores (%) vs. SNR for the three types of tests.

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 involved 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, indicating 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 can be applied 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

1. B. C. J. Moore, *An Introduction to the Psychology of Hearing*. 4th ed. (Academic, London, 1997).
2. J. M. Pickett, *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*, (Allyn Bacon, Boston, Mass., 1999).
3. B. R. Glasberg and B. C. J. Moore, "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," *J. Acoust. Soc. Am*. **79**, 1020 – 1033 (1986).
4. J. R. Dubno and A. B. Schaefer, "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," *J. Acoust. Soc. Am*. **91** Pt.1., 2110–2121 (1992).
5. H. T. Bunnel, "On enhancement of spectral contrast in speach for hearing-impaired listeners," *J. Acoust. Soc. Am.* **88**, 2546-2556 (1990).
6. T. Baer, B. C. J. Moore, and S. Gatehouse, "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times," *J. Rehabil. Res. Dev.* **30**, 49-72 (1993).
7 I. Cohen, "Speech spectral modeling and spectral enhancement based on autoregressive conditional heteroscedasticity models," *Signal Processing*. **86**, 698-709 (2006).
8 K. Yasu, K. Kobayashi, K. Shinohara, M. Hishitani, T. Arai, and Y. Murahara, "Frequency compression of critical band for digital hearing aids," *Proc. China-Japan Joint Conf. Acoustics*. 159 – 162 (2002).
9 P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, D. "Multi-band frequency compression for reducing the effects of spectral masking," *Int. J. Speech Tech*. **10**, 219-227 (2007).
10 P. E. Lyregaard, "Frequency selectivity and speech intelligibility in noise," *Scand. Audiol. Suppl*. **15**, 113 – 122 (1982).
11 T. Lunner, S. Arlinger, J. Hellgren "8-channel digital filter bank for hearing aid use: preliminary results in monaural, diotic, and dichotic modes," *Scand. Audiol. Suppl*. **38**, 75 – 81 (1993).
12 D. S. Chaudhari, and P. C. Pandey, "Dichotic presentation of speech signal with critical band filtering for improving speech perception," *Proc. IEEE ICASSP*, 3601 – 3604 (1998).
13 A. Murase, F. Nakajima, S. Sakamoto, Y. Suzuki, and T. Kawase, T. "Effect and sound localization with dichotic-listening hearing aids," *Proc. 18th Int. Congress Acoust. (ICA)*, Kyoto, Japan, II-1519 – 1522 (2004).
14 E. Villchur, "Simulation of the effect of recruitment on loudness relationships in speech," *J. Acoust. Soc. Am*. 56, 1601-1611 (1974).
15 E. Villchur, "Electronic models to simulate the effect of sensory distortions on speech perception by the deaf," *J. Acoust. Soc. Am*. **62**, 665-674 (1977).

16 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*. **94**, 2050–2062 (1993).
17 M. ter Keurs, J. M. Festen, and R. Plomp, "Effect of spectral envelope smearing on speech reception. I," *J. Acoust. Soc. Am*. **91**, 2872 – 2880 (1992).
18 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*. **102**, 603 – 615 (1997).
19 L. E. Humes, B. Espinoza-Varas, and C. S. Watson, "Modelling sensorineural hearing loss-I. Model and retrospective evaluation," *J. Acoust. Soc. Am.* **83**, 188 – 202 (1988).
20 M. R. Leek, M. F. Dorman, and Q. Summerfield, "Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **81**, 148 – 154 (1987).
21 D. A. Nelson, S. J. Chargo, J. G. Kopun, and R. L. Freyman, "Effect of forward-masked psychophysical tuning curves in quiet and noise," *J. Acoust. Soc. Am*. **88**, 2143 – 2151 (1990).
22 W. Jesteadt. *Modeling Sensorineural Hearing Loss*. (Lawrence Erlbaum Associates, Mahwah, New Jersey, 1997).
23 S. Gordon-Salant and P. J. Fitzgibbons, "Temporal factors and speech recognition performance in young and elderly listeners," *J. Speech Hear. Res*. **36**, 1276–1285 (1993).
24 G. A. Miller, and P. E. Nicely, "An analysis of perceptual confusions among some English consonants," *J. Acoust. Soc. Am.* **72**, 338 – 352 (1955).
25 W. D. Voiers, "Evaluating processed speech using the diagnostic rhyme test," *Speech Tech*. **55**, 30 – 39 (1983).
26 A. S. House, C. E. Williams, M. H. L. Hecker, and K. D. Kryter, "Articulation testing methods: consonantal differentiation with closed-response set," *J. Acoust. Soc. Am*. **37**, 158 – 166 (1965).
27 K. D. Kryter, and E. C. Whitman, "Some comparisons between rhyme and PB-word intelligibility tests," *J. Acoust. Soc. Am*. **37**, 1146 (1965).
28 S. Gatehouse, and J. Gordon, J. "Response times to speech stimuli as measures of benefit from amplification," *Br. J. Audiol*. **24**, 63 – 68 (1990).
29 F. Apoux, O. Crouzet, and C. Lorenzi, "Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners: Effects on identification performance and response times," *Hear. Res*. **153**, 123 – 131 (2001).