Proc. IEEE, ICASSP '98 (Seattle, Wash., U.S.A., 12-15 May, 1998) Paper AE 3.1

DICHOTIC PRESENTATION OF SPEECH SIGNAL WITH CRITICAL BAND FILTERING FOR IMPROVING SPEECH PERCEPTION

Devendra S. Chaudhari[1] and Prem C. Pandey[2]

[1]School of Biomedical Engineering
[2]Department of Electrical Engineering
Indian Institute of Technology, Bombay
Powai, Mumbai 400 076, India

ABSTRACT

Reduction in frequency resolving capacity of the auditory system due to spread of masking of frequency components by neighboring frequency components degrades speech perception in cases of sensorineural hearing impairment. We have carried out experimental evaluation of splitting speech into two signals by using a bank of critical band filters, in order to reduce the effect of spectral masking in the cochlea. The dichotically presented signals are perceptually integrated in the auditory cortex. Listening tests were carried out with vowel-consonant-vowel and consonant-vowel syllables for twelve English consonants on five normal hearing subjects with simulation of sensorineural impairment done by adding white masking noise to the speech signal at various SNRs. Significant improvements in recognition score were obtained under adverse listening condition. Improvement in the reception of speech feature of voicing, place, and manner was observed in information transmission analysis.

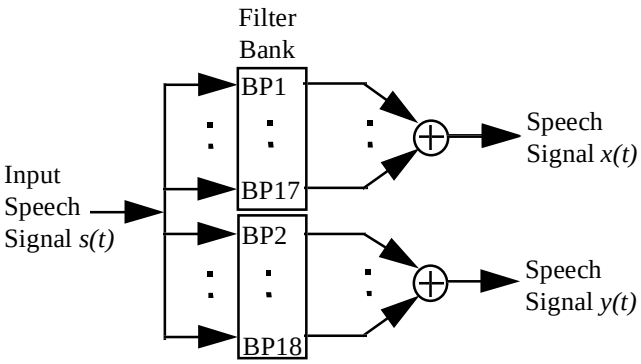
1. INTRODUCTION

The sensorineural impairments are characterized by high frequency hearing loss, increase in the threshold of hearing, compression in dynamic range, severity of temporal masking, and loss of spectral resolution due to spread of masking. The loss of spectral resolution results from masking of frequency components by neighboring components during the first stage of auditory processing along the cochlear partition. A speech processing scheme that splits speech into two signals for presenting to the two ears for reducing this masking effect is likely to improve speech reception in cases of bilateral sensorineural impairment with some residual hearing. Our ability to binaurally receive and perceptually combine signals from two ears for improving speech perception under adverse listening conditions has been well established [7].

The splitting of speech signal into the two channels can be carried out in a number of ways. Lunner et al [5] have used 8-channel constant bandwidth filtering for splitting speech for dichotic presentation and have reported improvements in speech reception during experiments with the hearing impaired subjects. The objective of our investigation is to split the speech in two signals with complementary spectra on the basis of critical band filtering for binaural dichotic presentation as a possible solution to problem of spectral masking. The study was carried out by processing digitized speech, and listening tests were conducted using normal hearing subjects with simulated sensorineural hearing loss.

2. EXPERIMENTAL METHOD

Speech signal was split on the basis of multiple critical band filtering. The critical bandwidths selected are the auditory filter bandwidths reported by Zwicker [11]. The speech was filtered and divided into two parts, as shown in *Figure 1*, in such a way that the odd numbered filter outputs were fed to one ear and even numbered filter outputs were fed to the other ear. The corner frequencies of the band pass filter are given in *Figure 1*. The processing was done by digitally filtering the speech signal, digitized with 12-bit resolution at 10 k Sa/s. The anti-aliasing filter at the input and the two reconstruction filters have same specifications: pass band edge = 4.6 kHz, pass band ripple < 0.3 dB, stop band edge = 5 kHz, stopband attenuation > 40 dB. The input and processed speech signals were spectrographically [8,10] analyzed for verifying the characteristics of the signal processing.



Filter	Pass band	Filter	Pass band
BP1	0.07-0.20	BP2	0.20-0.30
BP3	0.30-0.40	BP4	0.40-0.51
BP5	0.51-0.63	BP6	0.63-0.77
BP7	0.77-0.92	BP8	0.92-1.08
BP9	1.08-1.27	BP10	1.27-1.48
BP11	1.48-1.72	BP12	1.72-2.00
BP13	2.00-2.32	BP14	2.32-2.70
BP15	2.70-3.15	BP16	3.15-3.70
BP17	3.70-4.40	BP18	4.40-5.30

FIGURE 1. Splitting of speech signal using multiple bandpass (BP) filtering. The 3 dB cutoff frequencies of the bands are in kHz.

The testing was done on normal hearing subjects, with simulated sensorineural impairment. On the basis of different studies [1,2,3,4], one can say that broadband noise can be used for simulating various aspects of sensorineural hearing loss in normal hearing subjects for speech reception. We have used Gaussian white noise bandlimited to the band of the speech signal as masking noise at signal-to-noise ratios of ∞, 6, 3, 0, and -3 dB for simulating different levels of the hearing loss. The addition of the noise in each case has been done in such a way that the overall sound level remains unchanged. The sound was presented binaurally at the individual subject's most comfortable listening level which varied, for different subjects, from 75 to 85 dB SPL.

Listening tests were carried out for finding the confusions among the set of twelve consonants /p, b, t, d, k, g, m, n, s, z, f, v/ in the vowel-consonant-vowel (VCV) and consonant-vowel (CV) context with vowel /a/. These tests happen to be repetitive and time consuming, and hence conducted using an automated test administration system with the subject seated in the acoustically isolated chamber. At the end of each session, the confusion matrix, and response time statistics are stored. For each subject, tests were administered for (a) unprocessed speech presented to the left and to the right ears and (b) processed speech dichotically presented to the two ears. For each case, the tests were carried out at five SNR conditions randomized across the test sessions. Subjects were asked to also provide a qualitative assessment of the test stimuli.

3. RESULTS AND DISCUSSION

Listening tests were conducted with four subjects in VCV context and five subjects in CV context. The recognition scores obtained from confusion matrices averaged across the subjects are given in *Table 1*. Paired t-test [9] for testing the significance of differences of averaged scores for the unprocessed versus processed speech were carried out and these are also tabulated along with the scores. The recognition scores for VCV and CV context, averaged across the subjects, are plotted in *Figure 2*.

Under no noise condition, all the subjects have nearly perfect scores with both unprocessed and the processed speech. However, all the subjects showed less response time for processed speech, indicating an improvement in listening condition with processing.

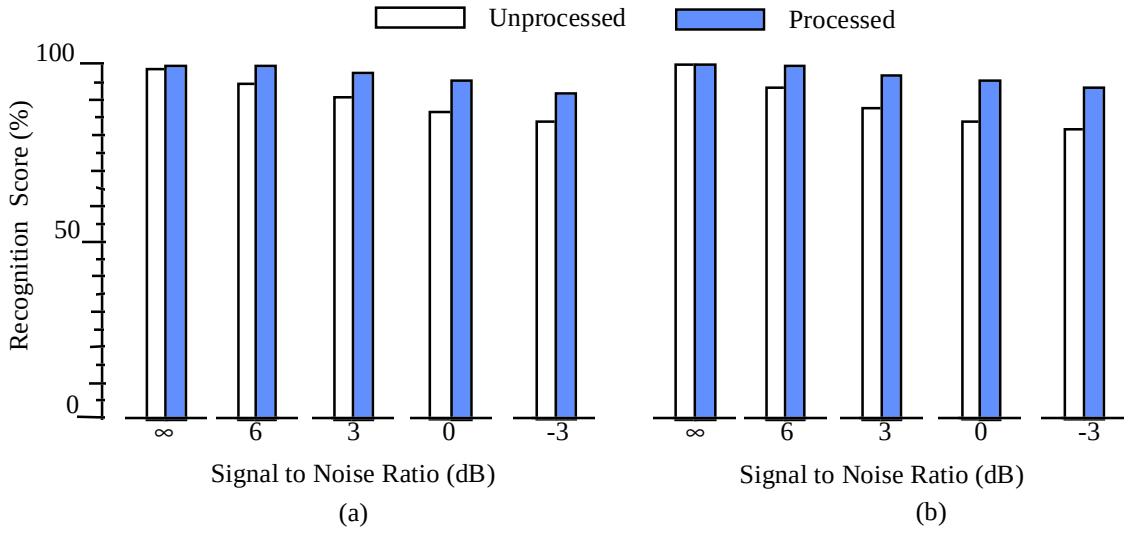


FIGURE 2. Recognition score at different SNRs (a) VCV context and (b) CV context.

For all the subjects, score generally decreases as the masking noise level increases. We further see that the scores for processed speech is higher than that for the unprocessed speech under the same condition of masking noise. It is to be noted that the improvements due to processing are more for higher levels of masking noise, *i.e.*, higher level of sensorineural loss.

For a given masking level, the scores for the unprocessed speech varied across the subjects. Relative improvements in recognition score was calculated as

$$R_s = (S_p - S_u) / S_u$$

where S_p and S_u are recognition scores with processed and unprocessed speech, and these are also given in *Table 1*. For decreasing SNRs, the percentage relative improvements range from 1.1 to 9.9 for VCV context and from 0.1 to 14.3 for CV context, indicating that processing of the speech and dichotic presentation improves recognition scores and improvements are higher under adverse listening condition. A compilation of qualitative assessments, by the subjects, about the set of test stimuli under various listening conditions indicated that the speech quality was better with processing for dichotic presentation.

In order to study the reception of specific consonant features, stimulus-response confusion matrices were subjected to information transmission analysis [6]. Almost all the subjects have shown improvements in relative information transmission in manner, place, and voicing features and overall information transmission. Summary of results for one subject is given in *Table 2*. In case of high SNR conditions, the information transmission is near perfect even with unprocessed speech and improves to 100 % with

TABLE 1. Percentage recognition scores, averaged across the subjects, for the 12-consonant listening tests for unprocessed speech (S_u), and processed speech (S_p), the relative percentage improvement in recognition score (R_s), and the significance level p as obtained from two tailed t-test for S_u and S_p.

SNR	S_u		S_p		R_s	p
	mean	s.d.	mean	s.d.		
(A) VCV context						
∞	98.9	0.8	100.0	0.0	1.1	< 0.1
6 dB	94.8	4.9	99.5	1.1	5.5	< 0.2
3 dB	91.5	7.2	97.9	2.0	7.0	< 0.2
0 dB	86.7	6.7	95.3	2.5	9.9	< 0.1
-3 dB	84.0	6.2	92.1	5.5	9.6	< 0.2
(B) CV context						
∞	99.8	0.2	99.9	0.2	0.1	NS
6 dB	93.5	6.0	99.2	1.4	6.1	< 0.1
3 dB	87.9	11.9	97.3	3.3	10.7	< 0.2
0 dB	84.4	14.0	95.6	4.0	13.3	< 0.2
-3 dB	81.9	14.0	93.6	5.3	14.3	< 0.2

processed speech. With poor SNRs, the information transmission with unprocessed speech decreases, and improvements are seen with the processed speech. The overall improvements are contributed by better reception of all the three features of voicing, manner, and place. However, the improvement is maximum for the place feature. It is to be noted that the place feature is subject to frequency resolving capacity of the auditory processing. Hence it can be inferred that the processing scheme implemented here has reduced the effect of spectral masking.

4. CONCLUSIONS

It was observed from the results for all the subjects that the recognition score generally decreased as the masking noise level increases. Further, we see that under the same condition of masking noise, the score for processed speech is higher than that for the unprocessed speech. The improvements due to processing are more for higher levels of masking noise, i.e. higher levels of simulated sensorineural loss. However, these improvements tend to level, at very high levels of loss. Information transmission analysis of the stimulus-response confusion matrices indicated that improvements in reception of consonants were contributed by better reception of all the three features - voicing, place, and manner with highest improvement for place.

TABLE 2. Percentage relative information transmitted for a typical subject for the set of 12-consonant stimuli in VCV context.

Feature	SNR				
	∞	6	3	0	-3
(A) Unprocessed					
Manner	95	98	96	94	92
Place	96	94	84	73	68
Voicing	100	96	96	97	85
Overall	98	97	94	90	87
(B) Processed					
Manner	100	100	98	98	98
Place	100	100	96	81	84
Voicing	100	100	100	100	91
Overall	100	100	98	94	94

REFERENCES

[1] S. DeGennaro, L. D. Braida, and N. I. Durlach "Study of multi-band syllabic compression with simulated sensorineural hearing loss", *J. Acoust. Soc. Am.*, vol. 69, S16, 1981.

[2] H. Fletcher, "The perception of sound by deafened persons", *J. Acoust. Soc. Am.,* vol. 24, pp. 490-497, 1952.

[3] W. Jesteadt, Ed., *Modeling Sensorineural Hearing Loss,* New Jersey : Lawrence Erlbaum, Mahwah, 1997.

[4] J. P. A. Lochner and J. F. Burger, "Form of the loudness function in the presence of masking noise", *J. Acoust. Soc. Am.*, vol. 33, pp. 1705-1707, 1961.

[5] T. Lunner, S Arlinger, and J. Hellgren, "8-channel digital filter bank for hearing aid use: preliminary results in monaural, diotic and dichotic modes", *Scand. Audiol.*, suppl. 38, pp. 75-81, 1993.

[6] G. A. Miller and P. E. Nicely, "An analysis of perceptual confusions among some English consonants", *J. Acoust. Soc. Am.*, vol. 27 (2), pp. 338-352, 1955.

[7] B. C. J. Moore, *An Introduction to Psychology of Hearing,* New York: Academic, 1982.

[8] L. R. Rabiner and R. W. Schafer, *Digital Processing of the Speech Signals*, Englewood Cliffs, NJ: Printice Hall, 1978.

[9] G. W. Snedecor and W. G. Cochran, *Statistical Methods*, Ames, Iowa: The Iowa State University Press, 1980.

[10] T. G. Thomas, P. C. Pandey, and S. D. Agashe, "A PC-based multiresolution spectrograph", *J. IETE (India)*, vol. 40, pp. 105-108, 1994.

[11] E. Zwicker, "Subdivision of audible frequency range into critical bands (Frequenz-gruppen)", *J. Acoust. Soc. Am.*, vol. 33, p. 248, 1961.