Fifth International Symposium on Signal Processing and its Applications, ISSPA '99, Brisbane, Australia, 22-25 August, 1999
Organised by the Signal Processing Research Centre, QUT, Brisbane, Australia

BINAURAL AUDIO IN MULTIMEDIA SYSTEMS TO IMPROVE AUDITORY PERCEPTION FOR THE HEARING IMPAIRED

D. S. Chaudhari[1] and P. C. Pandey[2]
[1]School of Biomedical Engineering
[2]Department of Electrical Engineering
Indian Institute of Technology, Bombay
Powai, Mumbai 400 076, India
Email: pcpandey@ee.iitb.ernet.in

ABSTRACT

Persons with sensorineural hearing impairments experience difficulty in using the audio part of the multimedia information. Degraded frequency selectivity of the ear, due to increase in spectral masking along the cochlear partition, results in degraded speech perception. Splitting the speech signal by filtering it with a filter bank and adding signals from alternate bands for presenting to the two ears is likely to reduce the effect of spectral masking and thus help in improving speech intelligibility. The dichotically presented signals are perceptually combined in the higher levels of auditory processes. We have carried out experimental evaluation of this scheme, using critical band filters, in off-line and real-time processing of speech signal. The scheme resulted in improving speech quality, response time, recognition scores, and transmission of consonantal features, particularly the place, indicating usefulness of the scheme for better reception of spectral characteristics. The signal processing for improving the auditory perception by persons with moderate sensorineural hearing loss may be integrated as a part of the multimedia system.

1. INTRODUCTION

Sensorineural hearing-impaired persons face difficulty while using audio part of multimedia information. Loss of frequency selectivity due to increase in spectral masking is one of the characteristics of the sensorineural hearing impairment. Researchers have investigated various techniques for improving the speech intelligibility: multiband compression, spectral transposition, speech enhancement using the properties of the "clear" speech, etc. Some of these techniques have been incorporated as a part of digital hearing aids. However, these techniques can't solve the problems in auditory perception due to loss of frequency selectivity.

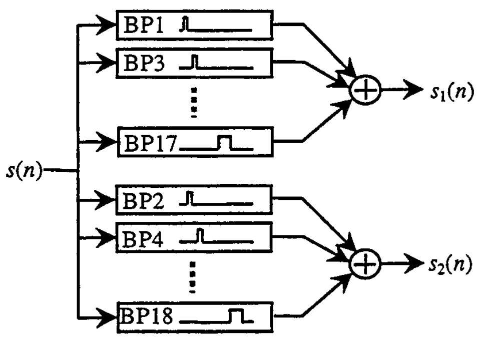
Persons with moderate bilateral hearing loss can exploit the binaural auditory presentation in multimedia system to improve the auditory perception. Binaural listening offers better overall sound quality and intelligibility, more relaxed listening, and it also helps in source localization [7] [4]. Study on spectral splitting of the audio signal for binaural dichotic presentation by employing two sets of complementary comb filters has indicated improvements in speech reception [5]. The choice of the filter characteristics and design had been constrained by considerations of low complexity and power consumption as required for use in wearable digital hearing aids.

We have tested an audio signal-processing scheme for splitting the audio signal for binaural dichotic presentation, using eighteen critical bands corresponding to auditory filters as described by Zwicker [9] and combining the odd and even numbered critical bands. The scheme was implemented with off-line processing of speech signal [1] [2]. Listening tests were carried out for experimental evaluation of the scheme for measuring confusion among the set of twelve English consonants, /p, b, t, d, k, g, m, n, s, z, f, v/ in vowel-consonants-vowel (VCV) and consonant-vowel (CV) contexts, the vowel being /a/ as in father. Five normal hearing subjects (21-40 years), under simulated loss of varying degrees and ten hearing impaired subjects (18-58 years) with 'mild'-to 'very severe' bilateral sensorineural hearing loss were tested. The scheme was found useful in improving speech perception, recognition score, and transmission of speech features, particularly the place feature indicating the effectiveness of the scheme for reducing the effect of spectral masking. On the basis of favorable results in listening tests with off-line processing, the scheme was implemented for real-time processing Listening tests were conducted with six hearing impaired subjects having 'mild'-to-'very severe' bilateral sensorineural loss. As in the case of off-line processing, in this implementation also the scheme was found useful. In this paper, the signal processing schemes used in off-line and real-time are discussed along with the results obtained while implementing the scheme.

2. SIGNAL PROCESSING SCHEMES

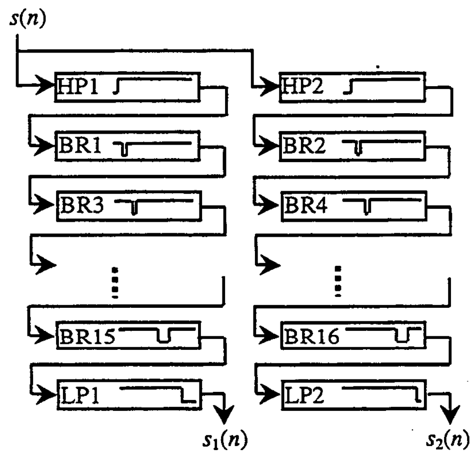
Speech processing for dichotic presentation can be implemented in three different ways. Fig. 1 shows a scheme with two filter banks, each with nine band pass filters. The two output signals are obtained by adding the outputs of the filters in the bank. In this approach using parallel combination of band pass filters, different bands can be added with different gains. However, there is a possibility of notches in the magnitude at crossover frequencies due to phase shifts in the adjacent filters. Fig. 2 shows a scheme in which the splitting of the input signal in two signals is realized by using cascade combination of band reject filters. The phase shifts of individual filters do not affect the magnitude response in this approach. In the scheme shown in Fig. 3, two filters with the desired comb filter response are designed. This may result in overall efficiency of realization.

For off-line processing, the cascade combination of band reject filters as shown in Fig. 2 was selected. All the filters are linear phase FIR filters with 255 coefficients. Such a high filter order was selected in order to obtain



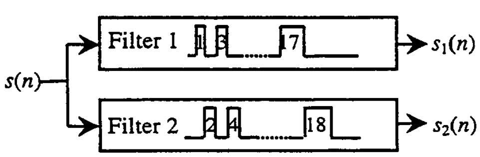
Filter	Freq. (kHz)	Filter	Freq. (kHz)
BP1	-- -0.20	BP2	0.20-0.30
BP3	0.30-0.40	BP4	0.40-0.51
BP5	0.51-0.63	BP6	0.63-0.77
BP7	0.77-0.92	BP8	0.92-1.08
BP9	1.08-1.27	BP10	1.27-1.48
BP11	1.48-1.72	BP12	1.72-2.00
BP13	2.00-2.32	BP14	2.32-2.70
BP15	2.70-3.15	BP16	3.15-3.70
BP17	3.70-4.40	BP18	4.40- --

Figure 1. Splitting of speech signal using two banks of band pass filters. The filter magnitude response is shown in each block (table shows 3-dB cut-off frequencies).



Filter	Freq.(kHz)	Filter	Freq. (kHz)
HP1	0.07	HP2	0.20
BR1	0.20-0.30	BR2	0.30-0.40
BR3	0.40-0.51	BR4	0.51-0.63
BR5	0.63-0.77	BR6	0.77-0.92
BR7	0.92-1.08	BR8	1.08-1.27
BR9	1.27-1.48	BR10	1.48-1.72
BR11	1.72-2.00	BR12	2.00-2.32
BR13	2.32-2.70	BR14	2.70-3.15
BR15	3.15-3.70	BR16	3.70-4.40
LP1	4.40	LP2	5.00

Figure 2. Splitting of speech signal using cascade combination of band reject filters. The filter magnitude response is shown in each block (table shows 3-dB cut-off frequencies).



Band	Passband frequency	Band	Passband frequency
1	-- -0.20	2	0.20-0.30
3	0.30-0.40	4	0.40-0.51
5	0.51-0.63	6	0.63-0.77
7	0.77-0.92	8	0.92-1.08
9	1.08-1.27	10	1.27-1.48
11	1.48-1.72	12	1.72-2.00
13	2.00-2.32	14	2.32-2.70
15	2.70-3.15	16	3.15-3.70
17	3.70-4.40	18	4.40- --

Figure 3. Splitting of speech signal using two comb filters. The filter magnitude response is shown in each block (table shows 3-dB cut-off frequencies).

sharp cut-off, so that there is no significant overlap from the neighboring critical bands. Computational efficiency was not considered to be important since processing of signals was done in off-line mode. The filter coefficients were obtained from the sampled magnitude response using rectangular window [8]. The real-time processing was done using two DSP boards based on 16-bit fixed-point processor, TI/TMS320C50. Each board consists of processor along with an analog interface circuit (AIC) with 14-bit ADC and DAC, and a programmable timer which was used for setting the sampling rate at 10 k samples/s [3].

The real-time processing scheme was implemented in two ways. In the first, PS-CG, the gain of all the filter bands was kept constant (same as in case of off-line processing). The second implementation, PS-AG provided adjustable filter gains, in the range of -3 to +3 dB, as a way of partial matching of the filter response to the frequency characteristics of the individual subject's hearing loss. The pure tone audiogram of the subject was interpolated to obtain the hearing loss up to 5 kHz. The adjustable filter magnitude response in dB, as a function of frequency f, is given as

$$A_a(f) = A_c(f) - 3 + 6\left[\alpha(f) - \alpha_{min}\right] / \left[\alpha_{max} - \alpha_{min}\right]$$

where $A_c(f)$ is the gain in dB for the constant gain implementation, $\alpha(f)$ is the interpolated value of hearing loss in dB, and α_{min} and α_{max} are the minimum and maximum values over the 125 Hz to 5 kHz frequency range. The relationship between filter gain and hearing loss is shown in Fig. 4. The frequency compensation was restricted to ± 3 dB, keeping in view the limited dynamic range provided by 16-bit processor. For both the desired comb filters, the magnitude response was approximated with 128 coefficients using frequency-sampling technique of linear phase FIR filter design. The filter program and coefficients can be loaded into the program RAM on the DSP chip using serial port interface. It is to be noted that no data transfer takes place between the two boards. In the second type of implementation, the frequency response of the comb filter was adjusted for individual subjects.

3. TEST RESULTS

In off-line processing implementation, two sets of listening tests were conducted. In the first set, five normal hearing subjects with simulated hearing impairment participated. The hearing loss was simulated by adding Gaussian white noise limited to band of speech signal as a masker to the speech stimuli with 5 SNR conditions of ∞ (no noise), 6, 3, 0, and -3 dB. The second set of tests was conducted on ten subjects with 'mild'-to-'very severe' bilateral sensorineural hearing loss. Their PTA (pure tone average hearing threshold level, in dB HL, test frequencies: 0.5, 1, and 2 kHz) difference between right and left ear was from 4 to 30 dB. Presentations were done at the comfortable listening level for the subject. Listening tests were carried out for finding the confusions among the set of twelve 12 English consonants in VCV and CV contexts. An automated computerized test administration system was used for these tests.

In both the sets of tests, the qualitative assessment by the subjects indicated a definite preference for processed speech. In the first set of tests, the recognition score generally decreased as the masking noise level increased. Further, it is observed that for a particular level of masking noise, the score for processed speech is significantly higher than that for the unprocessed. The important finding is that 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 simulated loss. Improvements in recognition scores, for -3 dB SNR condition, were highly significant and ranged from 5 to 24 %. In the second set of listening tests, with ten subjects having bilateral hearing loss, most of the subjects indicated highly significant improvement in scores, and these ranged from 4 to 13 %. Information transmission analysis [6] of the stimulus-response confusion matrices, indicate that an overall improvement is contributed by reception of all the six consonantal features (duration, frication, nasality, manner, place, and voicing) with higher improvement for place, voicing, and manner, in almost all normal hearing and in majority of the hearing impaired subjects. In case of normal hearing

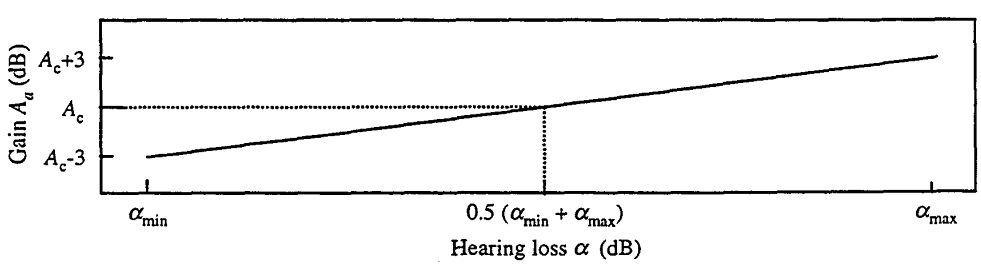


Figure 4. Relationship between filter gain and hearing loss, used for frequency dependent gain compensation in PS-AG implementation.

subjects the percentage relative information transmitted for place feature, for -3 dB SNR condition ranged from 13 to 48. In case of a hearing impaired subjects, it ranged from 2 to 24.

For both the real-time implementations, six subjects with bilateral sensorineural hearing impairment participated. The test material and method were similar to those used for off-line processing. In the first implementation PS-CG, the gain of all the bands was kept constant (same as in case of off-line processing). The second implementation PS-AG, provided adjustable magnitude response gains as a way of partial matching of the filter response to the frequency characteristics of the individual subject's hearing loss.

The qualitative assessment by the subjects indicated a definite preference for PS-CG over unprocessed speech. The improvement in recognition scores for subjects was highly significant and ranged from 8 to 15 %. Information transmission analysis of the stimulus-response confusion matrices indicated that the overall improvement is contributed by reception of all the six consonantal features mentioned above with nearly maximum improvement for place feature, among the features of place, voicing, and manner. The percentage relative information transmitted for place feature for subjects ranged from 6 to 27. In comparison of two implementations, the qualitative assessment of stimuli by the subjects did not show a clear preference for either PS-CG or PS-AG. In recognition scores, three subjects (out of six) showed highly significant advantage of PS-AG over PS-CG. Improvement in the recognition scores was significant and ranged from 1 to 6 %. In information transmission analysis, different features contribute the overall improvement, but there was no distinct contribution by the place feature.

4. CONCLUSIONS

In tests with off-line implementation of the speech processing for dichotic presentation, with both the types of subjects, the overall speech quality improves, recognition scores improve, load on perception process decreases, and effect of spectral masking reduces as indicated by better reception of place feature of consonants.

The tests with the first real-time implementation result in conclusions similar to the ones in case of the off-line processing. In the second implementation with further shaping of filter bands results in a mild advantage. These advantages are significant for subjects having relatively less variation (< 25 dB) in the hearing threshold over the 5 kHz frequency range. It was seen that the place feature did not make a distinct contribution to this. This indicates that the scheme is effective in reducing the effect of spectral masking, because of separation of speech spectra in accordance with critical band filtering. Further shaping of filter bands in accordance with hearing loss characteristics of the individual subject does not contribute to the reduction in the effect of spectral masking.

On the basis of above results it can be concluded that, the speech-processing scheme can be incorporated with multimedia systems without any special hardware modifications for improving speech perception by bilateral sensorineural hearing impaired. Additional improvement in speech perception can be obtained by suitably modifying the filter magnitude response through programming as per the individual's hearing loss. Further investigations are required for assessing the effect of processing on non-speech signals and also on source localization.

5. REFERENCES

[1] Chaudhari D. S. and Pandey P. C. "Dichotic presentation of speech signal with critical band filtering for improving speech perception". *Proceedings of International Conference on Acoustic, Speech, and Signal Processing,* Seattle, Washington, 1998, paper AE3.1.

[2] Chaudhari D. S. and Pandey P. C. "Dichotic presentation of speech signal using critical filter bank for bilateral sensorineural hearing impaired". *Proceedings of 16th International Congress on Acoustics*, Seattle, Washington, 1998, vol. 1, pages 213-214.

[3] Chaudhari D. S. and Pandey P. C. "Real-time speech processing for dichotic presentation for binaural hearing aids". *Proceedings of the National Conference on Biomedical Engineering,* Manipal, Karnataka, India, 1998, paper II-45.

[4] Kollmeier B., Hohmann V., and Peissig J. "Digital signal processing for binaural hearing aids". *Proceedings of International Congress on Acoustics (ICA)*, 1992, Beijing, pages H3-4.

[5] Lunner T., Arlinger S., and Hellgren J. "8-channel digital filter bank for hearing aid use: preliminary results in monaural, diotic and dichotic modes," *Scandinavian Audiology* S38:75-81, 1993.

[6] Miller G. A. and Nicely P. E. "An analysis of perceptual confusions among some English consonants". *Journal of the Acoustical Society of America*, 27(2):338-352, 1955.

[7] Moore B. C. J. *An introduction to the psychology of hearing*. Academic, London, 1997.

[8] Rabiner L.R., and Schafer R.W. *Digital processing of the speech signals.* Prentice Hall, Englewood, NJ, 1978.

[9] Zwicker E. "Subdivision of audible frequency range into critical bands (Frequenz-gruppen)". *Journal of the Acoustical Society of America,* 33:248, 1961.