Critical Filter-based Speech Signal Processing for Persons with Bilateral Sensorineural Hearing Impairment

Dr D S Chaudhari, *Non-member*

Dr P C Pandey, *Non-member*

One of the major factors in degraded speech reception by persons with sensorineural hearing loss is the reduced frequency selectivity along the cochlear partition in the ear. Speech signal processing using a filter bank based on critical bands (corresponding to auditory filters) and presenting signals from alternate bands) in such a way that signals corresponding to odd numbered bands are presented to one ear and even numbered are presented to the other is likely to reduce this effect. This may help in improving speech perception. The processing scheme was implemented using 18 critical bands over a 5 kHz frequency range. For experimental evaluation, the listening test material consisted of nonsense syllables formed with 12 english consonants and vowel /a/ in vowel-consonant-vowel and consonant-vowel contexts. The scheme was tested on ten sensorineural hearing-impaired subjects. It resulted in improvement in perceived speech quality, response time, recognition score and information transmission of consonantal place feature, signifying the usefulness of the scheme for better reception of spectral characteristics. The scheme may be employed in binaural hearing aids for persons with moderate bilateral sensorineural loss.

Keywords: Sensorineural hearing impairment; Critical bands; Spectral masking; Dichotic presentation

INTRODUCTION

Loss of frequency selectivity, due to spread of spectral masking along the cochlear partition of the ear, is one of the characteristics of the sensorineural hearing impairment[1]. The spread of spectral masking results from masking of adjacent frequency components. Use of a filter bank with a number of bandpass filters for splitting speech signal into two signals with complementary spectra, and presenting these signals dichotically to the two ears may reduce the effect of spectral masking, and thus may help in speech perception in cases of bilateral sensorineural hearing impaired. Based on the assumption that degraded frequency selectivity is a result of masking phenomenon at the peripheral level, several schemes have been investigated for processing speech for binaural dichotic presentation. Human ability of perceptually combining the binaurally received signals under adverse listening conditions has been well established earlier[1].

Lyregaard[2] tested a speech-processing scheme of spectral splitting of speech signal by using two complementary comb filters for binaural dichotic presentation. The two filters were realized by adding and subtracting an analog delayed version of the signal. By adjusting the delay, the bandwidth of the alternating pass and stop bands in the comb filter could be adjusted. Experiments were carried for the bandwidths of 200 Hz, 500 Hz and 800 Hz. The listening tests were conducted on three subjects with binaural hearing loss of 50 dB and two normal hearing subjects with dichotic presentation as well as diotic (unfiltered signals to both the ears). Improvements in the scores were not statistically significant for both the normal and the hearing-impaired subjects. Lyregaard suggested that the lack of significant improvement could be attributable to one of the three factors: unsuitable filtering, insufficient listening experience by the subjects, and non- feasibility of binaural fusion of dichotic signals.

Dr D S Chaudhari is with Department of Electronics and Telecommunication, Government College of Engineering, Amravati 444 604 and Dr P C Pandey is with Department of Electrical Engineering, Indian Institute of Technology, Powai, Mumbai 400 076.

This paper was received on October 10, 2000. Written discussion on the paper will be entertained till July 31, 2001.

Lunner *et al*[3], tested the use of digital filter bank for hearing aid use. The filter bank consisted of complementary interpolated linear phase FIR filters to give eight parallel filtered outputs. The outputs were added together with individually adjustable weighting factors in order to obtain a proper fit of the gain frequency response of the hearing aid to the need of the individual hearing aid user. The filters were with equal bandwidths of 700 Hz and 40 dB stopband attenuation. The scheme was implemented in real-time using TI/TMS320C25 digital signal processor based board and a PC. For dichotic presentation alternate bands were added and presented to each ear. In listening tests with subjects having bilateral hearing loss, an improvement of 2 dB in speech-to-noise ratio (SNR) for 50% correct recognition score was obtained for dichotic presentation over diotic.

It is to be noted that in both these schemes, all the bands have same bandwidth. Lyregaard's comb filter realization is straight forward, but it does not have distinct passband and stopband filters and the magnitude response in each band slowly tapers from the maximum to zero. Lunner *et al*'s comb filter has better separation of passband and stopband, but the transitions are not sharp. The present authors have based their comb filter response on the critical bands (auditory filter bandwidth)[1, 4] with emphasis on flat magnitude response in the passbands, sharp transitions and linear phase response. In the implemented scheme speech was split into two signals with complementary spectra, using off-line processing, based on critical bands (corresponding to auditory filters) for binaural dichotic presentation. The scheme was initially tested on five normal hearing subjects in the age group of 21 years to 40 years with varying degree of simulated sensorineural hearing loss[5]. For experimental evaluation the listening tests were carried out for measuring confusions among the set of 12 consonants in vowel-consonant-vowel (VCV) and consonant-vowel (CV) contexts formed with vowel /a/. It was found helpful in improving speech quality, response times, recognition scores and transmission of consonantal features, especially place, indicating the usefulness of scheme for better reception of spectral characteristics. On the basis of these results the scheme was further tested in off-line processing of speech. In this paper the authors present the results of listening tests with ten subjects having 'mild'-to-'very severe' bilateral sensorineural hearing loss.

IMPLEMENTATION

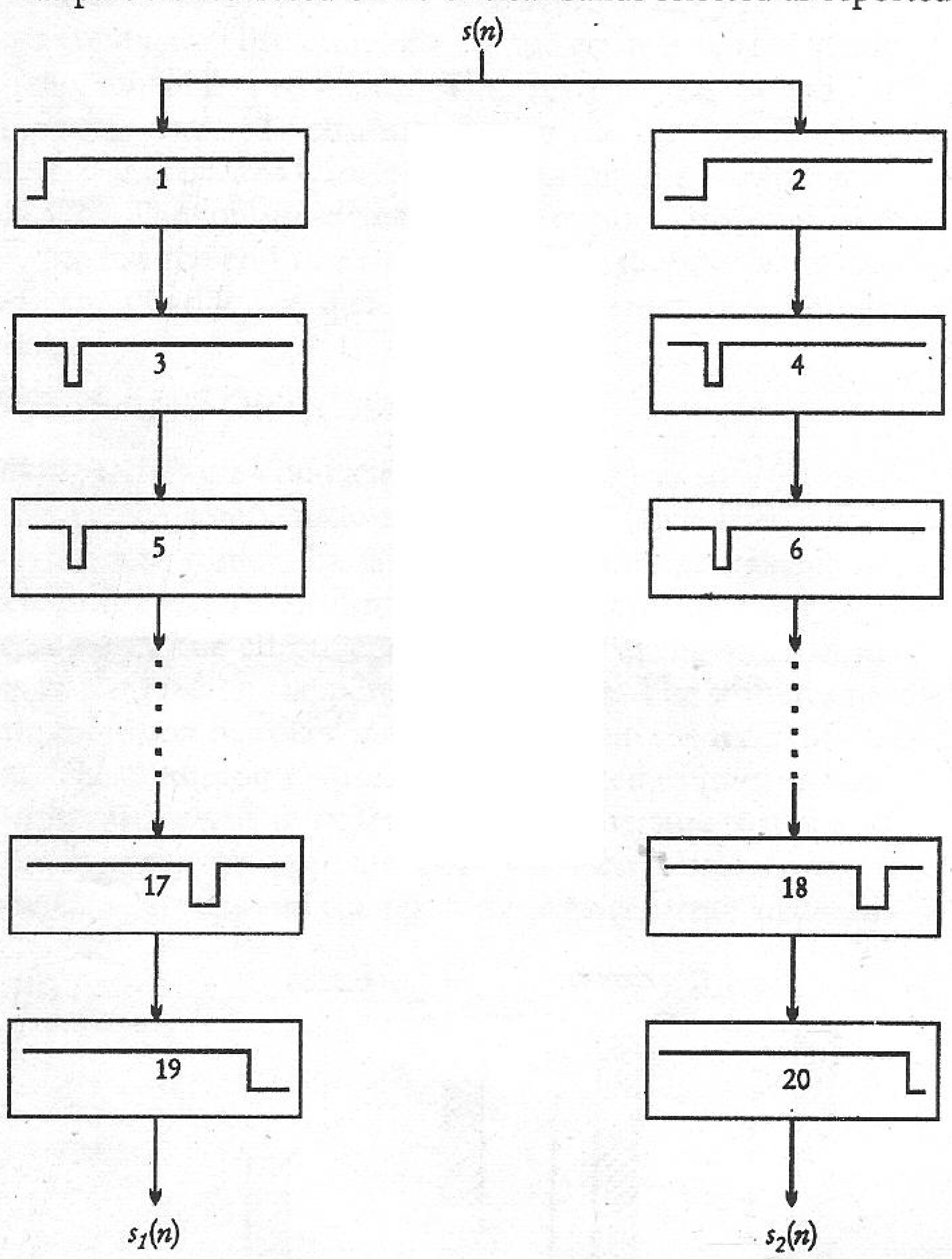
The processing scheme was implemented in which speech signal was partitioned based on 18 critical bands selected as reported by Zwicker[4]. The cascade combination of band reject filters as shown in Figure 1 was selected. All the filters were linear phase FIR filters with 255 coefficients. The filter coefficients were obtained from sampled magnitude response using rectangular window[6,7]. The signal $s_1(n)$ corresponds to odd numbered filters and signal $s_2(n)$ to even numbered.

The magnitude response of each of the two channels was obtained by taking 512-point FFT of their unit sample responses and are as shown in Figure 2. The filters are complementary, linear phase with passbands having bandwidths approximately equal to critical bands or auditory filter bandwidths. The filters have nearly flat magnitude response in the passband having ripples within 2 dB, sharp transitions from the passband to stopband that are less than 55 Hz and large stopband attenuation that is more than 40 dB. The filter responses were spectrographically analysed for verifying the speech signal processing[8].

The test material was acquired using a lowpass filter (cutoff 4.8 kHz) and 16-bit ADC at a rate of 10 k sample. The processing was done off-line by digitally filtering speech with bandwidth equal to critical bands. The input and processed speech signal with said scheme were presented to two ears through DAC, smoothing filter, power amplifier and headphones.

EXPERIMENTAL METHOD

The scheme was tested on ten hearing-impaired subjects (SG: M 27; SJH: F 19; KRN: M 35; DSD: M 19; LGR: M 27;



No	Filter	Freq, kHz	No	Filter	Freq, kHz
1	HP1	0.07	2	HP2	
3	BR1	0.20-0.30	4	BR2	0.20
5	BR3	0.40-0.51	6	BR4	0.30-0.40
7	BR5	0.63-0.77	8	BR6	0.51-0.63
9	BR7	0.92-1.08	10	BR8	0.77-0.92
11	BR9	1.27-1.48	12	BR10	1.08-1.27
13	BR11	1.72-2.00	14	BR12	1.48-1.72
15	BR13	2.32-2.70	16	BR14	2.00-2.32
17	BR15	3.15-3.70	18	BR16	2.70-3.15
19	LP1	4.40	20	LP2	3.70-4.40
					5.00

HP : High Pass; BR : Band Reject; LP : Low Pass

Figure 1 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)

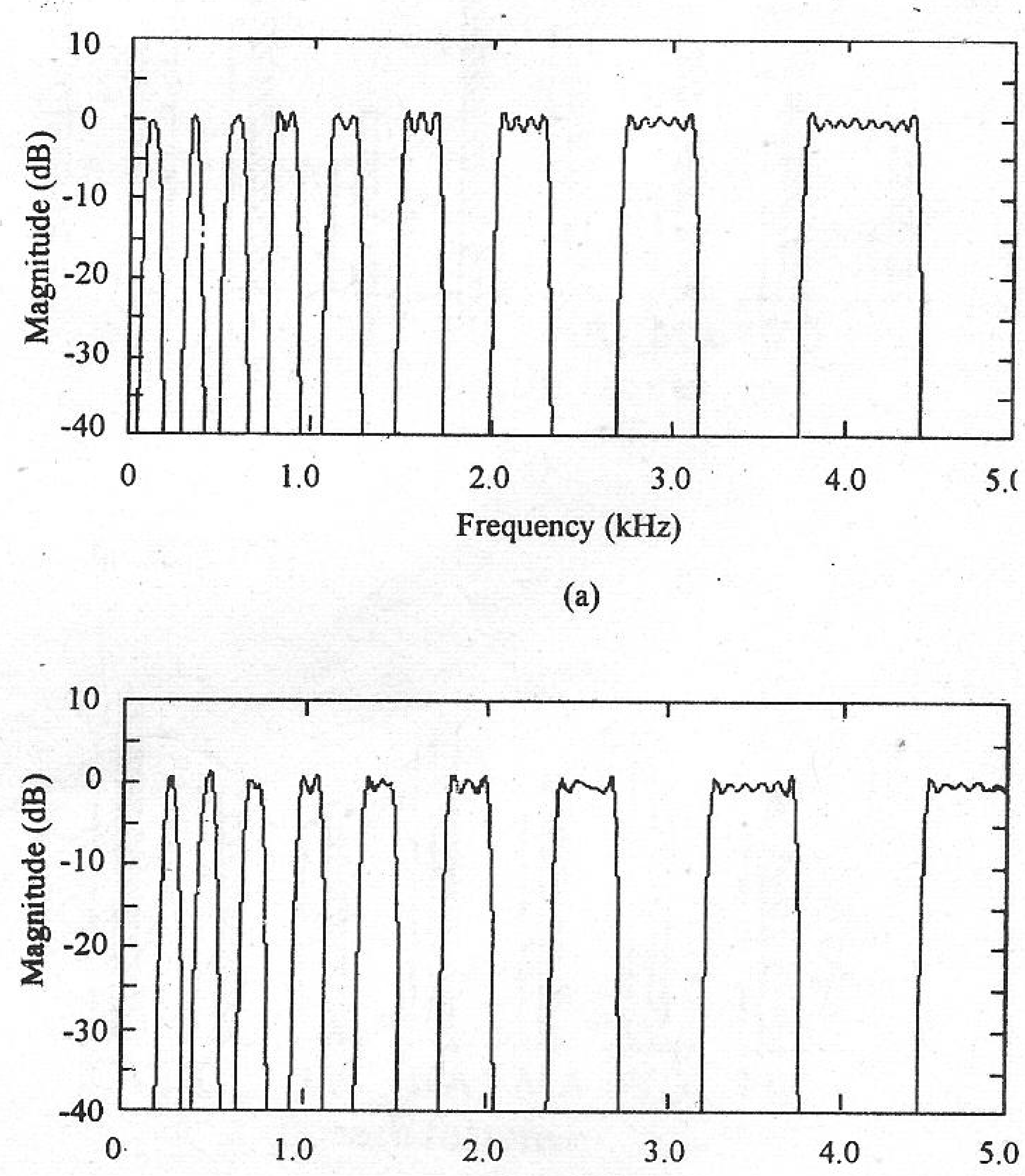


Figure 2 Magnitude response of the filters used in off-line processing of speech signal: (a) left ear (b) right ear

SSN: M 31; KRV: M 49; BAS: M 58; SAV: M 46; and LDM: M 48). The subjects were right handed and familiar with english. The subjects had 'mild'-to- 'very severe' bilateral sensorineural hearing loss and their pure tone average (PTA: hearing threshold levels for frequencies of 0.5 kHz, 1 kHz and 2 kHz) difference between right and left ear was from 0 dB to 30 dB.

The listening tests were carried out for obtaining stimulus-response confusion matrices among the set of twelve consonants */p, b, t, d, k, g, m, n, s, z, f, v/* in VCV and CV contexts formed with vowel */a/*. Due to repetitive nature of the tests, an automated test administration system was used[8].

The tests were administered for (a) unprocessed speech diotically presented and (b) processed speech dichotically presented. Before conducting listening test, the subject was briefed about the experimental procedure. The subject was seated in acoustically isolated chamber during the testing. The test material was presented using a pair of headphones (Telephonics TDH 39P). The confusion matrices and response time statistics was stored at the end of each session. The subjects were also asked to provide the qualitative assessment about the test material.

RESULTS AND DISCUSSION

Listening tests were conducted with ten subjects in VCV and CV contexts. A compilation of subjects' qualitative assessments about the test stimuli for ascertaining the improvement in speech quality was carried out. Averaged response time was used for comparing the effectiveness of the processing scheme, in terms of the load on the perception process. The stimulus-response confusion matrices were used to obtain the recognition scores. The confusion matrices were subjected to information transmission analysis in order to obtain a measure that is not affected by subject's response bias. The twelve stimuli were combined in groups and the resulting matrices were analyzed for reception of the consonantal features of duration, frication, nasality, manner, place and voicing.

Compilation of subjects' qualitative assessment indicated that the speech quality was better with processing for binaural dichotic presentation. In response times for the VCV and CV contexts, most of the hearing-impaired subjects showed decrease in response time due to processing. For most of the subjects, the decrease was statistically significant (as seen by *t*-test). This indicated an improvement in listening condition with processing, resulting in reduction in the load on the perception process. Paired *t*-test, across the subjects showed that the decreases in response time are highly significant ($p < 0.0005$) in both the contexts.

The recognition scores for a subject and averaged across the subjects are plotted in Figure 3. For all the subjects, the scores for processed speech were higher than those for unprocessed speech. The recognition scores were also subjected to *t*-test, for testing the statistical significance of improvements in scores due to processing. Majority of the subjects showed highly significant improvement due to processing ($p < 0.01$). Paired *t*-test, across the subjects for testing the significance of improvement in

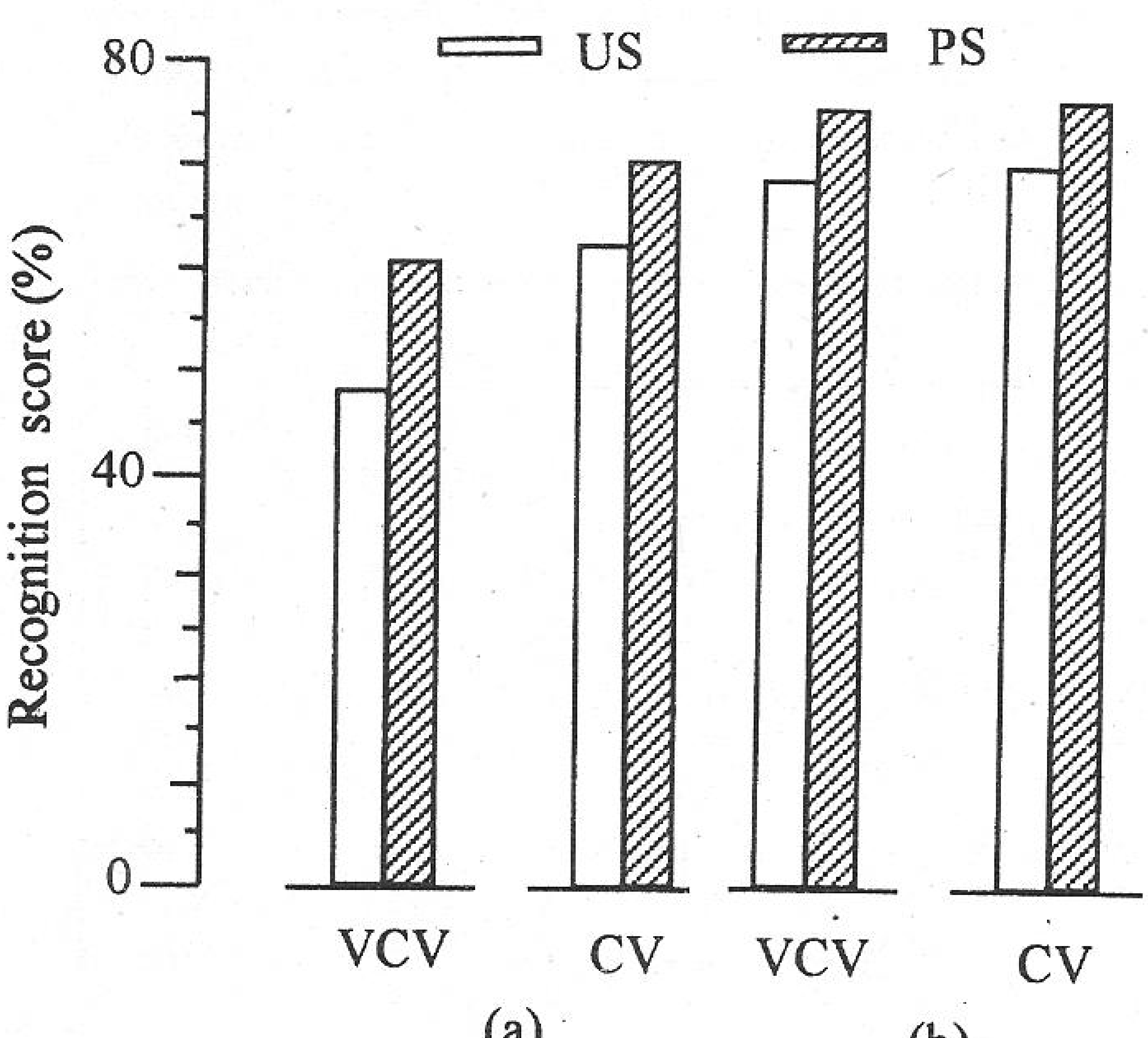


Figure 3 Recognition scores in VCV and CV contexts: (a) for subject SG (b) averaged for the ten subjects. US: unprocessed speech PS: processed speech

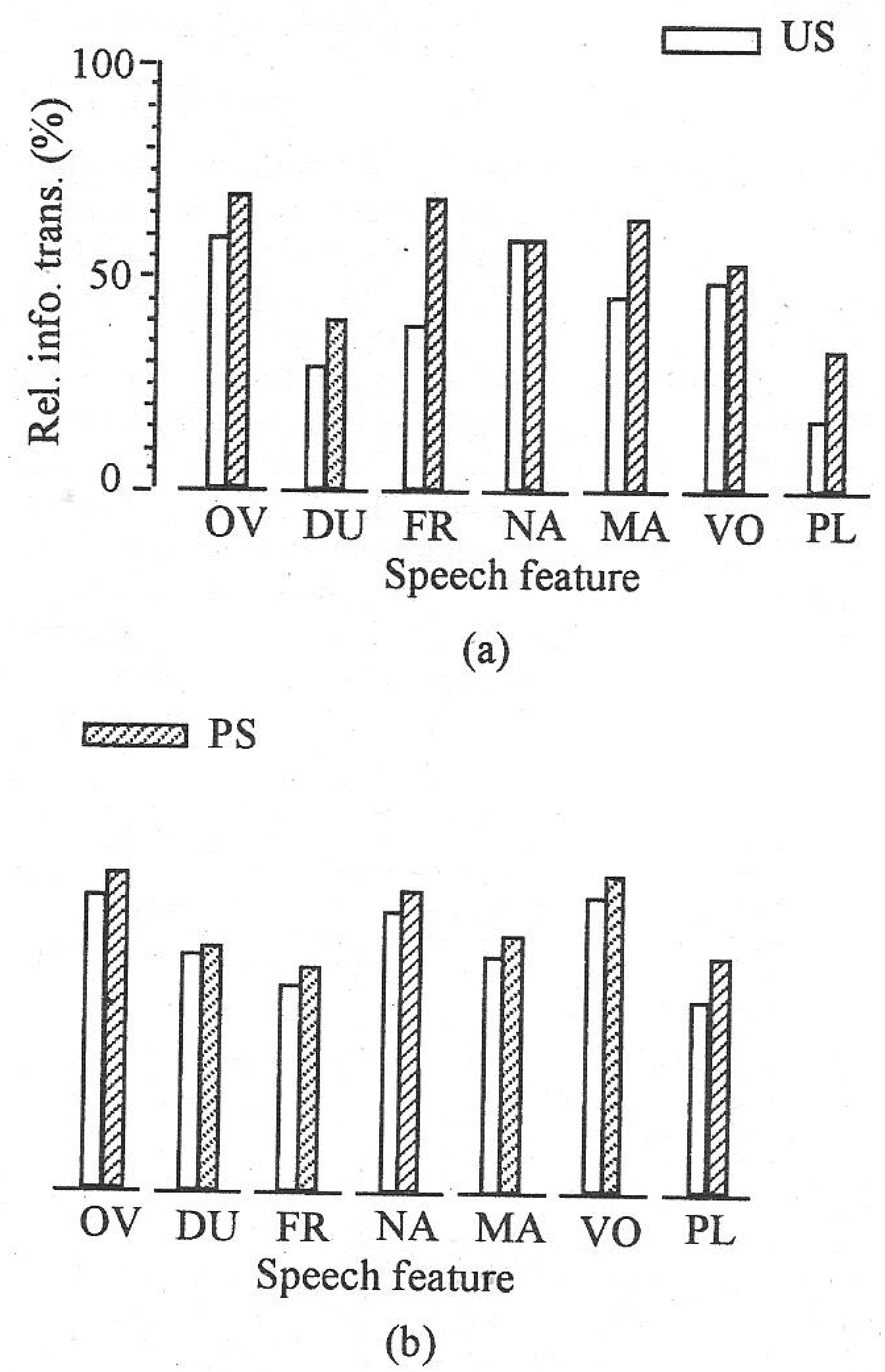


Figure 4 Percentage relative information transmitted, in VCV context (a) for subject SG (b) averaged for the ten subjects. OV: overall, DU: duration, FR: frication, NA: nasality, MA: manner, VO: voicing, PL: place, US: unprocessed speech, PS: processed speech

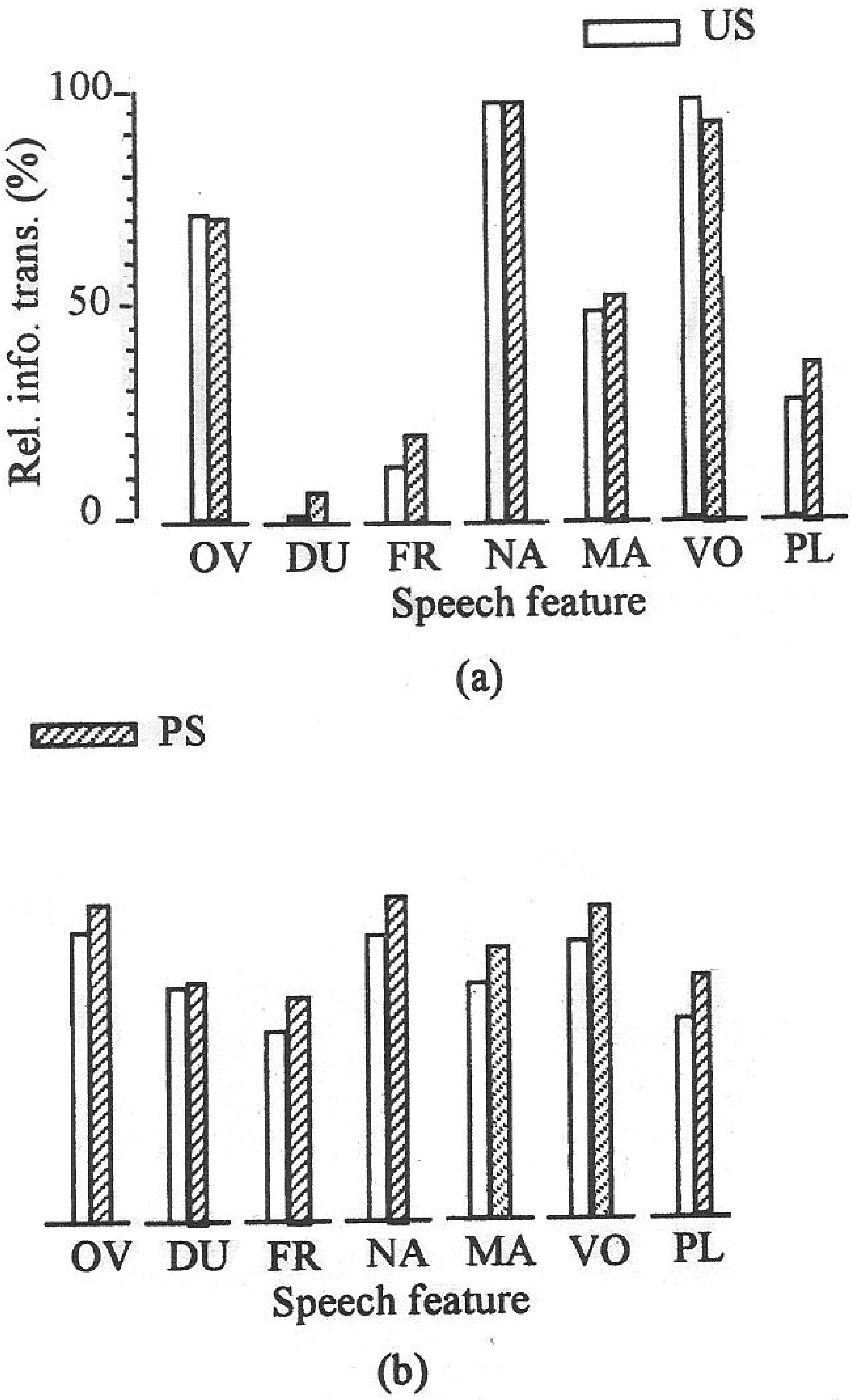


Figure 5 Percentage relative information transmitted, in CV context (a) for subject SG (b) averaged for the ten subjects. OV: overall, DU: duration, FR: frication, NA: nasality, MA: manner, VO: voicing, PL: place. US: unprocessed speech, PS: processed speech

recognition scores due to processing, was carried out and improvements are highly significant ($p < 0.0005$) in both the contexts.

The confusion matrices were subjected to relative information transmission analysis for overall and specific consonantal features. The results for a subject and averaged across the subjects are plotted for the VCV and CV contexts in Figures 4 and 5, respectively. The overall improvements were contributed by better reception of almost all the six features of duration, frication, nasality, manner, place and voicing. It was observed that, among the features of place, voicing and manner, largest number of subjects showed improvement in place feature (six subjects), followed by voicing (three subjects) and manner (one subject) in VCV context. Almost similar pattern of improvement in the reception of feature was observed in CV context also. As the reception of place feature is related to frequency resolving capacity of the auditory process, one can say that the implemented scheme has reduced the effect of spectral masking.

The relative improvements in the relative transmission of information for different features are given in Table 1. All the six features of duration, frication, nasality, manner, place and voicing contributed the overall relative improvements in transmission, with nearly maximum relative improvements for the place feature. Averaged across the ten subjects, the relative improvement for the place feature was 29% and 25% in VCV and CV contexts, respectively.

CONCLUSIONS

A speech processing scheme employing critical bandwidth filters was implemented for splitting speech signal into two signals with complementary spectra. The filters used have low passband ripple (<2 dB), sharp transitions (<55 Hz) and linear phase response. The application of the processing scheme improved the overall speech quality in the listening tests performed on sensorineural hearing-impaired subjects with off-line processing of speech for dichotic presentation. The tests recorded the significant fall in average response time indicating reduction in load on perception process. All the subjects showed remarkable improvement in recognition scores with corresponding enhancement in listening condition. The test also confirms the reduction of the effect of the spectral masking since better reception of consonantal place feature was observed,

Table 1 Relative improvement in relative information transmission of different features

Subject	Relative Improvement, %													
	Context : VCV							Context : CV						
	OV	DU	FR	NA	MA	VO	PL	OV	DU	FR	NA	MA	VO	PL
SG	17	38	74	0	36	10	94	−1	800	57	0	10	−4	34
SJH	−4	25	27	35	27	23	48	11	41	50	19	37	15	33
KRN	6	50	32	1	13	21	9	9	40	50	3	28	28	18
DSD	6	13	3	1	2	7	12	6	−8	8	18	11	19	2
LGR	6	−13	3	28	15	−9	15	14	−4	29	14	23	2	46
SSN	10	16	14	−7	0	19	28	14	13	28	−21	3	29	42
KRV	2	0	−5	17	5	9	24	9	6	−1	42	10	10	10
BAS	5	−4	−3	−11	−8	1	20	8	10	20	18	13	16	18
SAV	14	4	21	−13	−6	20	19	10	−17	7	38	21	14	28
LDM	9	−4	2	22	10	8	18	7	0	21	19	20	4	19
Average	7	12	17	7	9	11	29	9	88	27	15	18	13	25

OV : overall, DU : duration, FR : frication, NA : nasality, MA : manner, VO : voicing, PL : place indicating that the scheme may be used in binaural hearing aids for persons with moderate levels of bilateral sensorineural loss.

REFERENCES

1. B C J Moore. 'An Introduction to the Psychology of Hearing.' *Academic*, London, 1997.

2. P E Lyregaard. 'Frequency Selectivity and Speech Intelligibility in Noise.' *Scandinavian Audiology Journal*. suppl 15, 1982, p 113-122.

3. T Lunner, S. Arlinger and J Hellgren. '8-channel Digital Filter Bank for Hearing Aid use: Preliminary Results in Monaural, Diotic and Dichotic Models.' *Scandinavian Audiology Journal*. suppl 38, 1993, p 75-81.

4. E Zwicker. 'Subdivision of Audible Frequency Range into Critical Bands (Frequenzgruppen).' *Journal of the Acoustical Society of America*, vol 33, 1961, p 248.

5. D S Chaudhari and P C Pandey. '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, paper AE3. 1, 1998.

6. L R Rabiner, R W Schafer. 'Digital Processing of the Speech Signals.' *Prentice Hall*, Englewood Cliffs, New Jersey. 1978.

7, R Kuc. 'Introduction to Digital Signal Processing.' *McGraw-Hill*, New York, 1982.

8. D S Chaudhari. 'Dichotic Presentation for Improving Speech Perception by Persons with Bilateral Sensorineural Hearing Impairment.' Ph D *Thesis*, IIT Bombay, 2000.