Bio Vision 2001 Intl Conf. Biomed. Engg., Bangalore, India, 21-24 Dec. 2001, paper PRN6.

Time Varying Comb Filters to Reduce Spectral and Temporal Masking in Sensorineural Hearing Impairment

Dakshayani S. Jangamashetti, Prem C. Pandey, and Alice N. Cheeran

***Abstract*—The perception of speech gets degraded due to increased spectral and temporal masking in persons with sensorineural hearing loss. A scheme of binaural dichotic presentation to reduce the effect of both spectral and temporal masking simultaneously for persons with bilateral sensorineural hearing loss has been devised and investigated. The scheme uses a pair of time varying comb filters to split the speech signal into two for binaural dichotic presentation. The comb filters used are 256-coefficient linear phase FIR filters with bands in the filter magnitude response corresponding to the auditory critical bands. Each time varying comb filter has sets of pre-calculated coefficients. The coefficients are selected in steps such that a cyclic sweeping of magnitude responses occurs with a time period of 20 ms. At any instant of time, two comb filters with magnitude responses complementary to each other, process the speech signal for presenting to the two ears. The spectral components in the neighboring critical bands that are likely to mask each other are presented to different ears. The investigation was carried out with 2, 4, 8, and 16 filter sets. Listening tests were conducted on normal hearing subjects with simulated hearing loss by adding broad band noise at different signal-to-noise ratios. A closed set of 12 vowel-consonant-vowel syllables were used as test material. The test results have shown that processing resulted in the improvement of recognition scores, response time, and transmission of features particularly place and duration, indicating reduction in the effect of spectral and temporal masking. Improvement due to processing was highly significant for more adverse listening condition.**

***Index Terms*— sensorineural hearing loss, binaural hearing, speech processing for hearing**.

I. Introduction

Sensorineural hearing loss, which occurs due to damage of hair cells in the cochlea or degeneration of auditory nerve fibers or both, cannot be treated medically. The characteristics of this loss are elevated hearing threshold, loudness recruitment (abnormal growth in loudness perception with increase in intensity), reduced frequency and temporal resolution and increased spectral and temporal masking. Reduction in spectral contrasts, act as though the auditory filters are broader than normal [1]. The peaks and valleys of the speech spectrum get smeared affecting the perception of speech. Increased temporal masking results in the increase of forward and backward masking of weak acoustic segments by strong ones, which also affects speech intelligibility. Masking takes place in the peripheral auditory system. Speech perception at higher auditory levels involves the integration of information received from both the ears. Hence splitting of speech signal in a complementary manner and presenting to the two ears can be used to reduce the effect of masking on speech perception.

In a study by Lunner *et. al.* [2], an 8-channel digital filter bank with constant bandwidth of approximately 700 Hz was implemented for splitting speech. For binaural dichotic presentation, alternate bands were combined and they had complementary magnitude response. An improvement of 2 dB in speech-to-noise was observed for dichotic condition with respect to diotic. Further, to obtain combined spectral and temporal splitting a scheme was developed, in which the two comb filters were alternated between the two ears after every 10 ms. No further improvement over spectral splitting was reported, and switching of bands resulted in poor sound quality.

In a recent investigation on binaural dichotic presentations, speech signal was split using a pair of comb filters having complementary magnitude responses [3], [4]. Each of these comb filters had 9 pass bands based on critical auditory pattern described by Zwicker [5]. The bandwidths are constant at 100 Hz for center frequencies below 500 Hz and are 15-17% of the center frequency in the range of 1-5 kHz. The comb filters were linear phase FIR filters with 128 coefficients. The implementation was done in real time on two TI/TMS320C50 DSP processors [6]. Experimental evaluation of the scheme was conducted on bilaterally

Department of Electrical Engineering, IIT Bombay, Powai, Mumbai-400 076, India. E-mail: pcpandey@ee.iitb.ac.in

hearing impaired subjects, which lead to the conclusion that the scheme helped in improving the perception of consonantal features particularly the place feature. Subsequently a scheme of temporal splitting was developed in which inter-aural switching of speech signal was done using fading functions [7], [8]. Off-line evaluation was done on normal subjects with simulated hearing loss, with trapezoidal fading functions of different duty cycles and slopes. This scheme resulted in the improvement of consonantal duration feature.

In the scheme of spectral splitting sensory cells corresponding to alternate bands of the basilar membrane are always stimulated, whereas sensory cells of other bands are always relaxing in both the ears. In temporal splitting all the sensory cells of the ears get relaxed alternately for some time. A scheme with time varying comb filters, in which all the sensory cells of the basilar membrane get periodic relaxation from stimulation is devised and investigated.

II. Implementation

The implementation of combined spectral and temporal splitting was done off-line using a pair of time varying comb filters. The scheme of combined splitting along with the magnitude response of a time varying comb filter is shown in Fig. 1. The digitized input signal is shown as $s(n)$, and $s_1(n)$ and $s_2(n)$ are outputs to the two ears respectively.

Each time varying comb filter constitutes a number of comb filters, depending on the number of shiftings (represented as m). Each of these m comb filters have 9 pass bands corresponding to the auditory critical bands described by Zwicker [5]. At any instant of time, two comb filters which have complementary magnitude response, one each from the pair of time varying comb filters are used to process the speech for binaural dichotic presentation. If the comb filters in the time varying comb filters for the left ear are numbered as in a series as [*1*], [*2*], …, [*m/2*], [*m/2 +1*],…, [*m*], then that for the right ear will be numbered as [*m/2 +1*], [*m/2 +2*],…., [*m*], [*1*], [*2*],…, [*m/2*]. These comb filters have magnitude responses such that the bands of comb filter [*2*] will be slightly shifted along the frequency axis with respect to comb filter [*1*], as shown in Fig. 1(b). In a similar way all the pass bands of each of the comb filter will be a shifted version of the corresponding pass band of the previous one. After m shiftings when the cycle repeats, it looks as though each pass band merge into the next pass band. The complementary pairs are [*1*] and [*m/2 +1*], [*2*] and [*m/2 +2*], …, and [*m/2*] and [*m*] as shown by the positioning of rotating switch in Fig. 1(a).

The 256-coefficient linear phase FIR filters were designed using frequency sampling technique, for obtaining minimum spectral distortion. The transition crossover between any two adjacent bands of a pair used simultaneously, were adjusted to lie between –4 dB and –6 dB with respect to the pass band gain so as to minimize the difference in perceived loudness at the crossovers. For this, the magnitude of samples lying in the transition region were iteratively adjusted [9]. The pass band ripple was constrained to be less than 1 dB. The stop band attenuation was greater than 30 dB. The processing is done with sampling rate = 10 k Sa/s. The comb filters had transition bands of 78 Hz at low frequencies and 117 Hz at higher frequencies. A slowly swept sinusoidal tone when processed with each pair of comb filters that need to be used simultaneously, did not produce any change in perceived loudness.

The pre-calculated set of coefficients were cyclically swept with m shiftings (2, 4, 8, or 16) with a time period of 20 ms. After every time slot of 20/m ms a new set of coefficients for next pair of comb filters takes over. The sweeping of magnitude responses in the time varying comb filters is represented in Fig. 2, for 4 shiftings. It has to be noted that we are realizing filters with swept responses and the number of filters determines the smoothness of the sweeping.

III. Evaluation

The scheme was implemented for off-line processing. Listening tests were conducted on five normal hearing subjects with simulated hearing loss. Different levels of sensorineural hearing loss were simulated by adding broad-band Gaussian noise of different levels. The signal-to-noise ratios (SNR) used were ∞, 6, 3, 0, -3, -6, -9, -12, and -15 dB. In the listening tests the subjects were asked to identify a closed set of 12 English consonants /p, b, t, d, k, g, m, n, s, z, f, v/ in vowel-consonant-vowel context with vowel /*a*/ as in father. A computerized test administration system, consisting of a PC interfaced through RS232C serial port to the subject terminal (VT-220) placed in an acoustically isolated chamber, was used. The speech stimuli were outputted at 10 k samples/s through the two DAC ports of the PCL-208 data acquisition card. The signals are further passed through a smoothing low pass filter and an audio amplifier and are presented through two headphones (Telephonics TDH-39P).

For each of the test conditions, recognition scores were obtained and were stored in the form of stimulus-response confusion matrices. Percentage recognition score, relative information transmitted for different features, and subject’s response times for unprocessed and processed speech were analyzed. Paired t-test was used for finding the significance between the scores of unprocessed and processed signals.

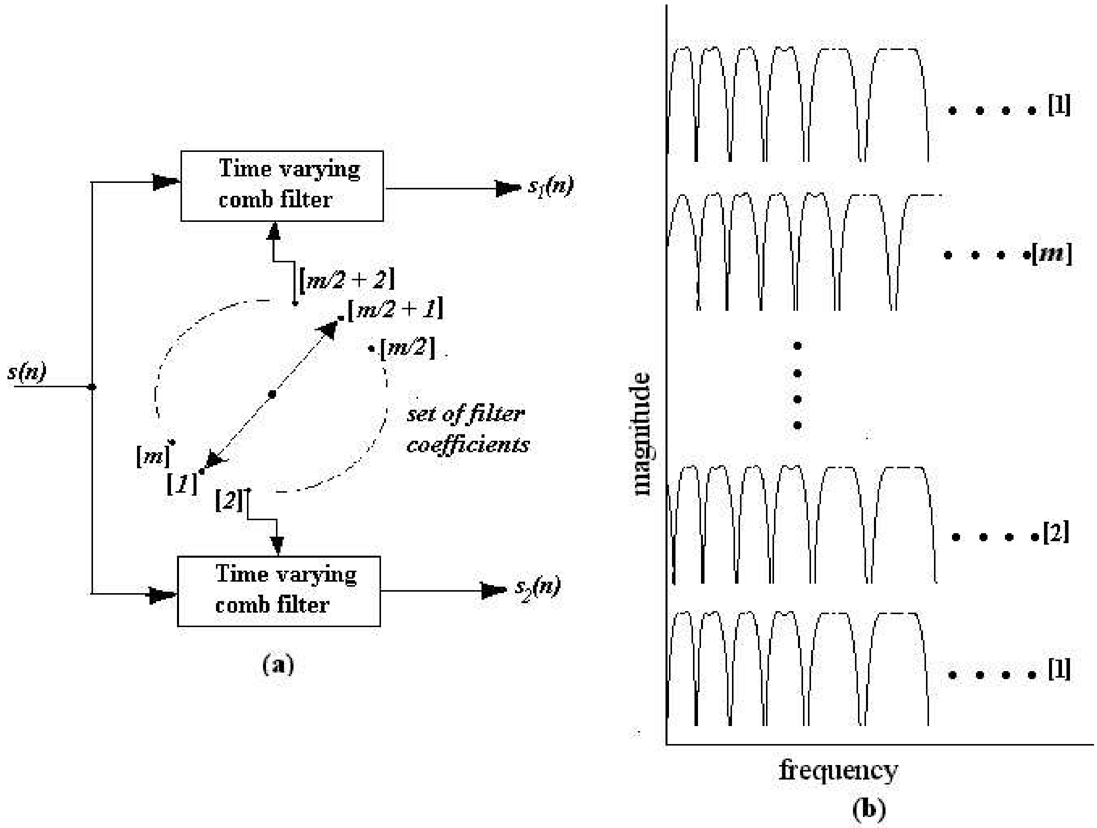

Fig. 1. (a) Schematic representation of the scheme of combined splitting using time varying comb filters and (b) Representation of magnitude response of one of the time varying comb filters which includes a set of comb filters (magnitude responses), which are swept over one after the other cyclically.

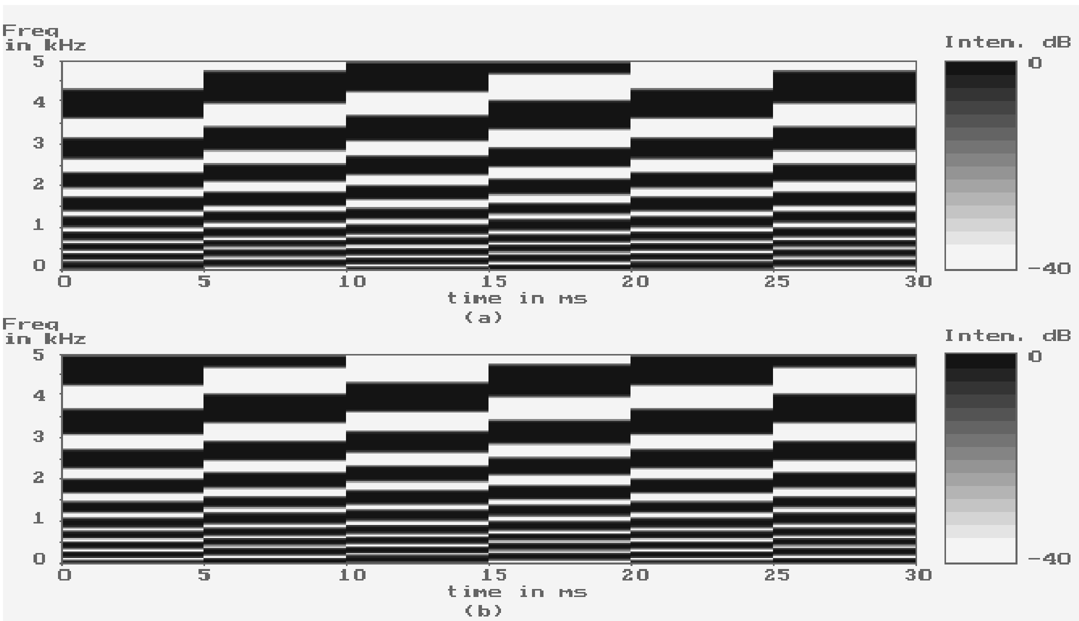

Fig. 2. An idealized representation of magnitude response of the pair of time varying comb filters using 4 shiftings (*m*). After every 5 ms (20/*m*) next pair of comb filters take over. The cycle repeats after 20 ms (a) for left ear and (b) for right ear. Effect of the previous filter coefficients have been ignored.

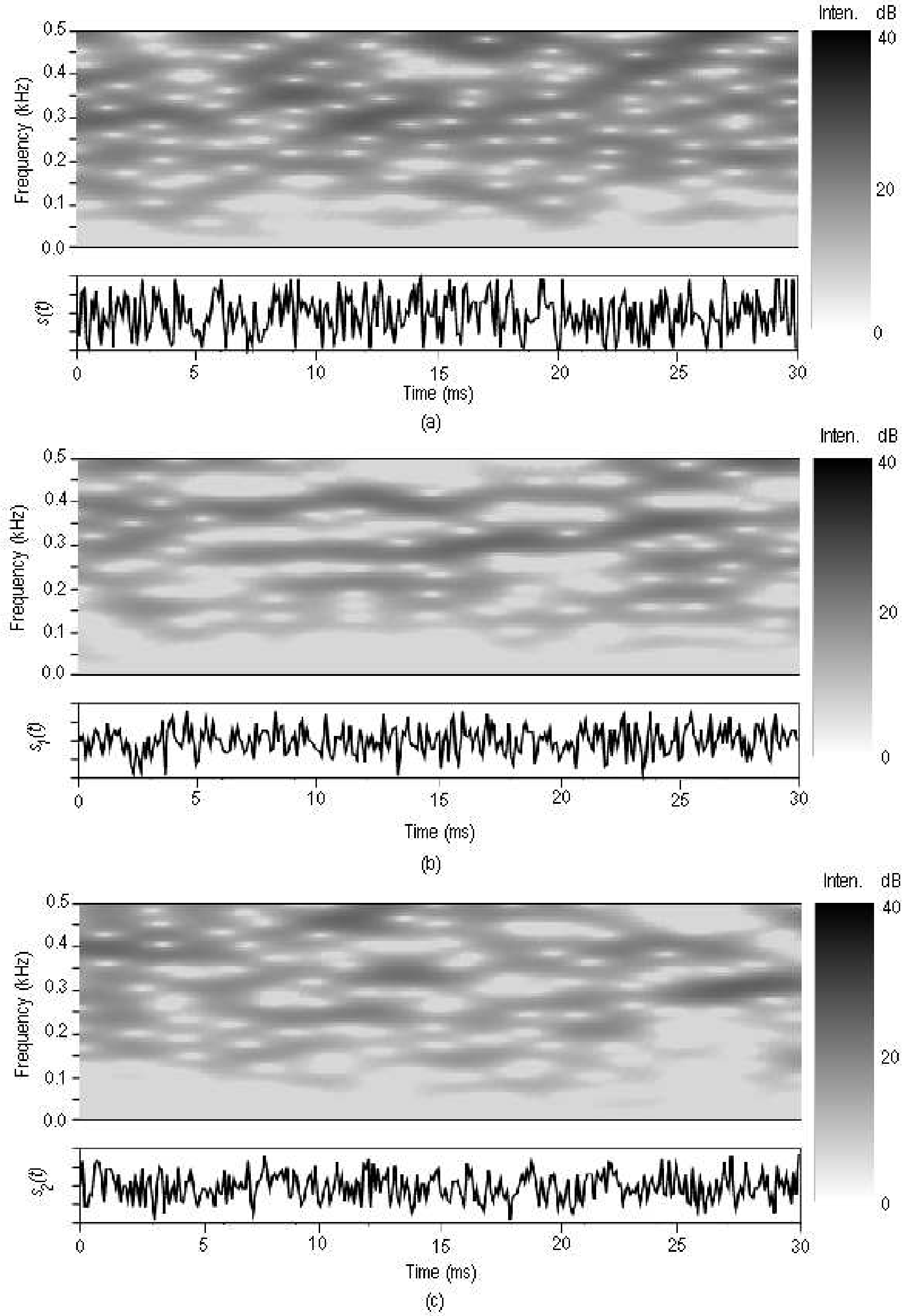

Fig. 3. Wide band spectrogram for white noise of 30 ms duration: unprocessed (a) and processed with time varying comb filters with 4 shiftings for left ear (b) and right ear (c).

Table 1. Recognition scores (%) for Unprocessed Speech, Su and relative improvement (%) for Processed Speech SpA, SpB, SpC, SpD corresponding to 4 shiftings 16, 8, 4, and 2 respectively. S: Subject, Avg.: Averaged across subjects. p: significance levels (paired t-test).

S	-3 dB SNR					-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD
S1	95	4.5	4.5	4.9	2.8	91	8.0	7.3	7.3	7.64	88.7	12.4	10.1	9.0	10.5	83.3	13.5	11.2	13.2	13.2	76.7	16.5	13.5	17.0	12.2
S2	84	19.0	19.0	18.0	18.0	75	28.0	23.6	29.4	26.7	67.0	26.8	31.8	30.3	17.9	59	33.7	33.1	39.3	10.7	59.7	26.3	23.0	38.5	14.0
S3	84	17.0	19.0	17.5	15.5	85	12.6	13.7	15.7	10.0	73.3	27.7	28.6	26.8	19.5	70	21.0	24.0	22.4	17.2	51.3	49.3	71.4	68.2	53.0
S4	90	1.84	7.0	2.6	3.3	86	7.8	7.4	11.2	5.0	86	4.3	1.5	3.5	1.5	74	13.5	18.0	15.3	16.7	63.3	12.6	16.3	13.7	12.5
S5	91	2.56	6.57	4.38	4.0	91	3.6	1.8	5.8	3.3	89	1.9	4.8	6.7	4.5	84	3.2	4.0	5.5	8.3	71.7	10.7	19.5	23.0	7.9
Avg.	88.7	8.98	11.2	9.47	8.7	85.5	12.0	10.8	13.9	10.5	81	14.6	15.4	15.3	10.8	74	17.0	18.1	19.1	13.2	64.5	23.0	28.7	32.0	20.0
p		0.032	0.009	0.021	0.025		0.015	0.015	0.009	0.023		0.021	0.026	0.017	0.014		0.007	0.005	0.006	0.0		0.006	0.013	0.006	0.02

Table 2. Information transmitted for unprocessed speech, Su and relative improvement (%) in information transmitted (%) for processed speech SpA, SpB, SpC, SpD corresponding to 4 shiftings 16, 8, 4, and 2 respectively for (a) overall, (b) place feature, and (c) duration feature. S: Subject, Avg.: Averaged across subjects.

(a) Overall

S	-3 dB SNR					-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD
S1	94.0	5.3	5.3	6.4	3.2	91.0	7.7	6.6	6.6	7.7	86.0	15.1	11.6	11.6	12.8	81.0	13.6	12.3	14.8	13.6	75.0	18.7	12.0	16.0	10.7
S2	84.0	17.9	17.9	16.7	17.9	79.0	21.5	17.7	21.5	20.3	74.0	17.6	20.3	18.9	8.1	69.0	15.9	15.9	21.7	2.9	68.0	13.2	7.4	17.6	1.5
S3	85.0	14.1	16.5	15.3	12.9	85.0	10.6	12.9	14.1	8.2	79.0	19.0	20.3	16.5	8.9	70.0	17.1	24.3	18.6	18.6	62.0	24.2	40.3	35.5	24.2
S4	92.0	0.0	4.3	0.0	0.0	88.0	3.4	3.4	6.8	1.1	86.0	3.5	3.5	1.2	-1.2	80.0	3.8	7.5	6.3	5.0	70.0	4.3	8.6	5.7	0.0
S5	91.0	2.2	6.6	4.4	4.4	92.0	3.3	1.1	5.4	1.1	90.0	2.2	4.4	4.4	3.3	84.0	3.6	4.8	4.8	8.3	74.0	10.8	14.9	17.6	1.4
Avg.	89.2	7.9	10.1	8.5	7.7	87.0	9.3	8.4	10.9	7.7	83.0	11.5	12.0	10.5	6.4	76.8	10.8	13.0	13.2	9.7	69.8	14.2	16.6	18.5	7.5

(b) Feature: place

S	-3 dB SNR					-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD
S1	80.0	22.5	22.5	25.0	20.0	71.0	32.4	29.6	26.8	33.8	61.0	60.7	45.9	41.0	47.5	51.0	64.7	47.1	56.9	68.6	47.0	38.3	40.4	46.8	34.0
S2	63.0	55.6	55.6	54.0	50.8	34.0	150	120	165	141	24.0	146	158	146	58.3	11.0	346	273	364	146	14.0	150	179	293	100
S3	71.0	28.2	36.6	32.4	29.6	68.0	25.0	29.4	35.3	19.1	41.0	100	107	92.7	58.5	35.0	68.6	71.4	62.9	62.9	12.0	225	417	425	283
S4	71.0	4.2	21.1	5.6	9.9	63.0	17.5	15.9	27.0	4.8	55.0	27.3	18.2	21.8	14.5	34.0	58.8	88.2	67.6	70.6	27.0	55.6	55.6	40.7	37.0
S5	79.0	6.3	12.7	6.3	7.6	74.0	13.5	2.7	17.6	9.5	68.0	8.8	14.7	20.6	11.8	53.0	13.2	17.0	24.5	32.1	39.0	28.2	51.3	79.5	10.3
Avg.	72.8	23.4	29.7	24.7	23.6	62.0	47.7	39.6	54.3	41.7	49.8	68.5	68.9	64.4	38.1	36.8	110	99.3	115	75.9	27.8	99.4	149	177	92.9

(c) Feature: duration

S	-3 dB SNR					-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD	Su	SpA	SpB	SpC	SpD
S1	87.0	9.2	14.9	14.9	9.2	65.0	40.0	24.6	33.8	40.0	59.0	69.5	47.5	44.1	59.3	50.0	90.0	68.0	76.0	68.0	49.0	73.5	28.6	55.1	65.3
S2	83.0	15.7	20.5	20.5	9.6	47.0	70.2	63.8	100	55.3	16.0	306	288	350	213	8.0	713	425	750	288	7.0	671	486	814	314
S3	60.0	51.7	66.7	66.7	60.0	58.0	62.1	62.1	72.4	50.0	36.0	178	178	131	94.4	42.0	35.7	88.1	52.4	47.6	7.0	586	900	757	671
S4	86.0	-2.3	11.6	-11.6	3.5	61.0	34.4	42.6	39.3	36.1	58.0	36.2	36.2	20.7	15.5	26.0	169	212	212	169	23.0	174	100	117	126
S5	68.0	11.8	11.8	16.2	41.2	88.0	1.1	-20.5	8.0	-3.4	81.0	-16.0	-3.7	4.9	12.3	48.0	8.3	41.7	54.2	95.8	35.0	54.3	74.3	82.9	80.0
Avg.	76.8	17.2	25.1	21.3	24.7	63.8	41.6	34.5	50.7	35.6	50.0	115	109	110	78.8	34.8	203	167	229	134	24.2	312	318	365	251

IV. RESULTS

Improvement was found in recognition score, response time and transmission of features with an increasing trend from high SNR to low SNRs for all shiftings. Relative improvement in recognition score was higher in 4 and 8 shiftings. Table 1 gives the recognition scores for unprocessed speech and relative improvement for processed speech, for SNR conditions of -3, -6, -9, -12, -15 dB. The relative improvement in recognition score was maximum at -15 dB SNR and was 32% and 29% respectively for 4 and 8 shiftings. Improvements are statistically significant for higher levels of noise.

Table 2(a) gives overall information transmitted for unprocessed speech and relative improvements for processed speech. With unprocessed speech, average of overall information transmitted varies from 96% under no-noise condition to 70% at -15 dB SNR condition. At this SNR condition, with processing, average of relative improvements are 19% and 17% for 4 and 8 shiftings. It is seen that, subject S3 has low recognition score with low SNR. However, relative information transmitted for this subject is not lower than for other subjects. This indicates that, errors in reception by this subject are not randomly distributed. With dichotic presentation, relative information transmitted is better for this subject than for the other subjects.

From the data on information transmitted for different features, it was observed that more improvement was obtained for duration and place features. Tables 2(b) and 2(c) give relative information transmitted for unprocessed speech and relative improvement for processed speech, for place and duration features respectively. For place feature, for unprocessed speech, average of relative information transmitted varies from 100% under no-noise condition to 29% at -15 dB SNR condition. For duration feature, for unprocessed speech, average of relative information transmitted varies from 100% under no-noise condition to 24% at -15 dB SNR condition. Subjects S2 and S3 had great difficulty in perception of these two features under poor listening conditions. With processing, the reception of place and duration features improved for all the subjects, and particularly for subjects S2 and S3. For place and duration features, relative improvements averaged across the subjects are 177%, 149% and 365%, 318% respectively for 4 and 8 shiftings at -15 dB SNR condition.

Perception for frication and manner features is also improved. With processing response times decrease, maximum decrease is for 4 and 8 shiftings. Averaged across the subjects, relative decrease in response times are 17% and 14% for 4 and 8 shiftings respectively at SNR of -15 dB.

V. CONCLUSION

The scheme for splitting speech signal using time varying comb filters provides better speech intelligibility for normal persons with simulated hearing loss and the improvements increase under adverse listening condition. The improvement in the perception of place and duration features show that the scheme has helped in reducing the effect of spectral and temporal masking. Reduction in response time indicates a reduction in the load on the perception process for the subjects. Thus the investigation has shown that the devised scheme has the potential of improving speech perception for persons using binaural hearing aids. Further investigations will help in establishing optimal values of the parameters for combined splitting. The scheme needs to be tested on patients with bilateral hearing impairment.

REFERENCES

[1] M. ter Keurs, J.M. Festen, and R. Plomp, "Effect of spectral envelope smearing on speech reception. II," *J. Acoust. Soc. Am.*, vol.93, pp. 1547-1552, 1993.

[2] 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.

[3] D.S. Chaudhari and P.C. Pandey, "Dichotic presentation of speech signal using critical filter band for bilateral sensorineural hearing impairment," *Proc. of 16th Int. Congress on Acoust. (ICA).*, Seattle, Washington, U.S.A., pp. 213-214, 1998.

[4] D.S. Chaudhari and P.C. Pandey, "Dichotic presentation of speech signal with critical band filtering for improving speech perception," *Proc. ICASSP '98*, Seattle, Washington, U.S.A., paper AE 3.1, 1998.

[5] E. Zwicker, "Subdivision of the audible range into critical bands (Frequenzgruppen)," *J. Acoust. Soc. Am.* vol. 33, pp. 248, 1961.

[6] D.S. Chaudhari and P.C. Pandey, "Real time speech processing for dichotic presentation for binaural hearing aids," *Proc. National Conf. Biomedical Engg.*, Manipal, Karnataka, India, paper II-45, 1998.

[7] D.S. Jangamashetti and P.C. Pandey, "Dichotic presentation with inter-aural switching for reducing the effect of temporal masking due to sensorineural hearing loss," *Proc. National Conf. on Biomedical Engg.*, Roorkee, India, pp. 346-353, 2000.

[8] D.S. Jangamashetti and P.C. Pandey, "Inter-aural switching with different fading functions for binaural dichotic presentation to reduce the effect of temporal masking in sensorineural hearing loss," *Proc. 4th World Multi Conf. on Systemics, Cybernetics, and Informatics (SCI'2000)*, Orlando, Florida, U.S.A, pp. 434-439, 2000.

[9] L.R Rabiner, B. Gold, and C.A McGonegal, "An approach to the approximation problem for nonrecursive digital filters," *IEEE Trans. Audio and Electroacoustics* vol. AU-16, pp.83-105, 1970.