Evaluation of speech processing schemes to improve perception of sensorineural hearing impaired

Sensorineural hearing impairment occurs due to damage to the transduction mechanism of the inner ear and due to abnormality in the auditory nerve which are responsible for sensing sounds of different pitches. Sensorineural hearing impairment is distinguished by loudness recruitment, reduced frequency selectivity, and increased spectral and temporal masking. Spectral masking results in suppression of two adjacent frequency components by each other. Spectral masking is a source of perceptual confusion in consonantal segments[1].

Most hearing aids offer facility to vary gain and frequency compensation. These aids cannot resolve the problem of speech perception due to broadening of critical bands. The sensorineural impaired person comes across difficulties while resolving neighbouring frequency components of speech signal[2]. Splitting speech into two complementary parts on the basis of critical bands and presenting it dichotically helps in lightening these difficulties. The two neighbouring bands, which may mask each other are presented to different ears in dichotic presentation. Receiving speech information binaurally and its perceptual combination have been recognized[3]. Thus, the scheme should improve the reception of consonantal segment without adversely affecting the reception of other features. Auditory frequency selectivity is illustrated in terms of an equivalent rectangular bandwidth as a function of centre frequency[4,5].

This study is based on improvement in speech intelligibility for the sensorineural hearing impaired. Speech intelligibility can be increased by using frequency bands for dichotic presentation. Binaural dichotic presentation can be a possible solution to the problem of spectral masking by splitting speech into two complementary spectra. The speech signal presentation by using two different processing schemes based on this principle can be helpful to bilateral sensorineural hearing impaired people.

A personal computer/laptop based experimental set-up (MATLAB-GUI based) was used for binaural presentation of the test stimuli. In a presentation, when a stimulus item was presented over the headphones, all stimuli were displayed on the subject terminal screen. Each stimulus corresponded to a push button on the subject terminal's screen. The subject responded by pressing appropriate push button on the terminal's screen. Presentations were completed at the comfortable listening level for the subject. These subjects were tested without adding any noise to the speech stimuli. The procedure of the listening test was explained to the subject. Subjects could listen to the test material number of times as he/she desires at the beginning of each test, to become familiar with the stimuli. In a test, 15 stimulus items were presented for six times, in a random order, leading to a total number of 90 presentations for every subject.

The time taken by the subject to respond was also recorded for each presentation. A stimulus–response confusion matrix was formed in which stimuli represented along rows and responses were represented along columns at the end of each test. The occurrence of a stimulus–response pair is represented by each entry in the cell. The diagonal elements provide the correct responses whereas off-diagonal elements represent errors. Sum of the diagonal elements gives total number of correct responses. Percentage correct recognition scores and response time statistics were also presented along with the confusion matrix.

Fifteen English consonants (p, b, t, d, k, g, m, n, s, z, f, v, r, l, y) were used in the vowel-consonant-vowel (VCV) context with vowel (a) as in farmer. Nonsense syllables were used to minimize the contribution of linguistic factors. Seven subjects with hearing impairment (BMA: M 41, DAA: M 25, FSM: M 56, KST: F 61, NB: M 58, VR: F 51, PHS: M 24) participated in the listening test. Five subjects had bilateral mild to severe sensorineural hearing impairment and two subjects had mild to moderate sensorineural hearing impairment. Subjects DAA and FSM had symmetrically sloping high frequency impairment and subject PHS had symmetrical low frequency hearing impairment. Subjects NB and KST had more loss at high frequency. Subject VR had asymmetrical high frequency impairment (less loss in one ear and more loss in the other ear). Subjects BMA and PHS had mild to moderate sensorineural impairment.

In listening test experiments, binaural dichotic presentation was used. The objective was to study possible solutions to the problem of spectral masking in case of sensorineural hearing impaired persons. In view of this, acquired speech was split based on critical bands for comb filters and on 10 frequency bands for wavelet packets. Critical bandwidths were chosen as per the auditory filter bandwidths[5]. The performance by hearing-impaired subjects saturated around eight channels, while performance by normal-hearing subjects continued to increase up to 12–16 channels[6], hence for wavelet packets, 10 frequency bands were chosen. The frequency bands were spread over a frequency up to 5 kHz. These two banks were named as left-ear and right-ear bank. The left-ear bank filter output was fed to one ear and right-ear bank filter output was fed to the other. The speech was filtered and divided into two parts in such a way that frequency components lying within one band are in one part, components lying in the next non-overlapping band are in the second part, components of third non-overlapping band are in the first part.

The processing scheme was spectral splitting with comb filters based on auditory critical bandwidths and responses optimized for low perceived spectral distortion. Filters were designed using frequency sampling technique of linear phase finite impulse response (FIR) filter[7]. Comb filters were designed with 512 coefficients. The magnitude response of the pair of comb filters is shown in Figure 1. Each comb filter has nine pass bands corresponding to auditory critical bandwidths.

The discrete wavelet transform divides the signal spectrum into frequency bands that are narrow in the lower frequencies and wide in the higher frequencies. This limits how wavelet coefficients in the upper half of the signal spectrum are classified. Wavelet packets (WP) divide the signal spectrum into frequency bands that are evenly spaced and have equal bandwidth and will be explored for use in identifying transient and quasi steady

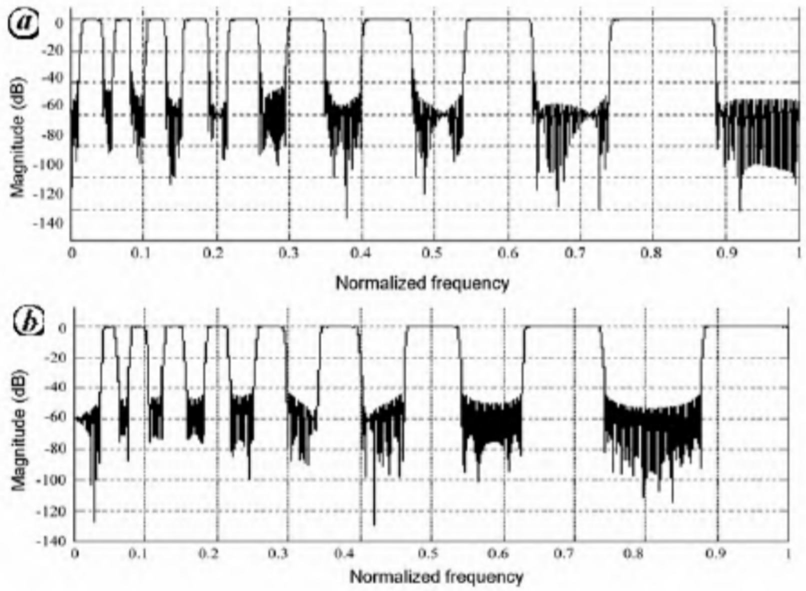

Figure 1. Filter responses for processing scheme of comb filters. ***a***, Right-ear bank filter; ***b***, Left-ear bank filter.

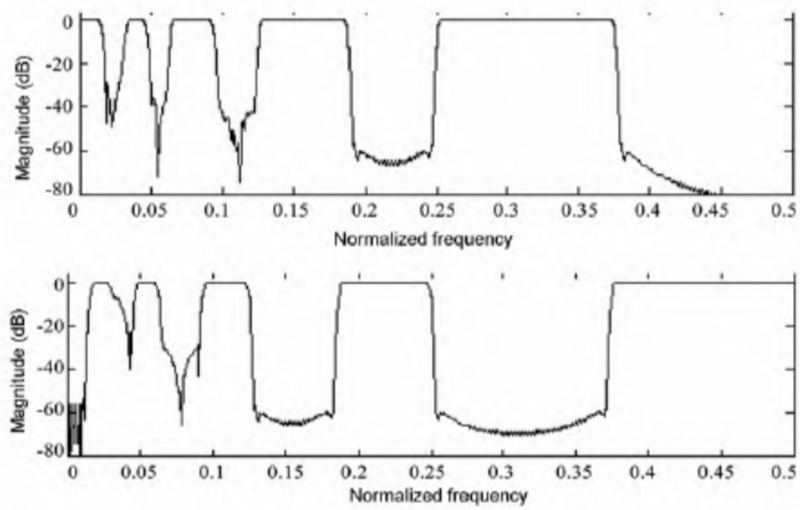

Figure 2. Filter responses for processing scheme of wavelet packet as right- and left-ear bank filters.

state speech. WP are efficient tools for speech analysis, involve using two bands splitting of the input signal by means of filtering and downsampling at each decomposition level[8–10]. Designing the WP filterbank involves choosing the decomposition tree and then selecting the filters for each decomposition level of the tree. For each decomposition level, there is a different time frequency resolution. Once the decomposition tree has been selected, the next step involves selecting an appropriate wavelet filter for each decomposition level of the tree. Here discrete wavelet transforms for one level of decomposition and WP for the second level of decomposition referred as modified wavelet packets are used. The magnitude response of the modified wavelet packets is shown in Figure 2.

Comb filters were implemented by writing a program in MATLAB[11]. The program was written for comb filters based on auditory critical bandwidths whereas modified wavelet packet was developed for symmlets wavelets using the MATLAB simulink software. Information transmission analysis can be carried out for different feature grouping of speech signals and can be used for evaluation in many studies[3]. Confusion matrix was used for calculating percentage recognition score and response time statistics. The MATLAB programming was done to calculate the confusion matrix and percentage recognition scores with respect to test data. Also, the recognition time was calculated for unprocessed and processed signals. The test was performed on seven different persons as off-line processing. Recognition scores were obtained from the confusion matrix and are as shown in Figure 3. It indicates the improvement in recognition scores for processed signals as compared to unprocessed. Recognition scores of the seven subjects were low for unprocessed signals and varied from 57.77% to 92.22%. Processing scheme of comb filters improved the scores for four subjects, while scores reduced for three subjects. The relative improvements in recognition scores varied from 1.78% to 4.44%. The relative improvements in recognition scores for processed scheme of wavelet packets were 3.33% to 22.23%. The response time was also recorded during testing and is shown in Figure 4. The response time was decreased for processed signals for almost all the persons in the processed scheme of wavelet packets.

The unprocessed and processed syllables were presented binaurally to sensorineural hearing impaired persons. It was observed that there was improvement in recognition scores using the processing scheme indicating effectiveness of the scheme. From the analysis of recognition scores and information transmission, it is observed that the scheme that gives benefit by reducing the effects of increased masking depends on the individual hearing loss configuration. Persons with low frequency hearing impairment and gradual sloping asymmetrical impairment benefit with the processing scheme of comb filters. Persons with low and high frequency hearing impairment benefit from the processing scheme of wavelet packets. Reduction in load on perception process is evident as observed in decreased

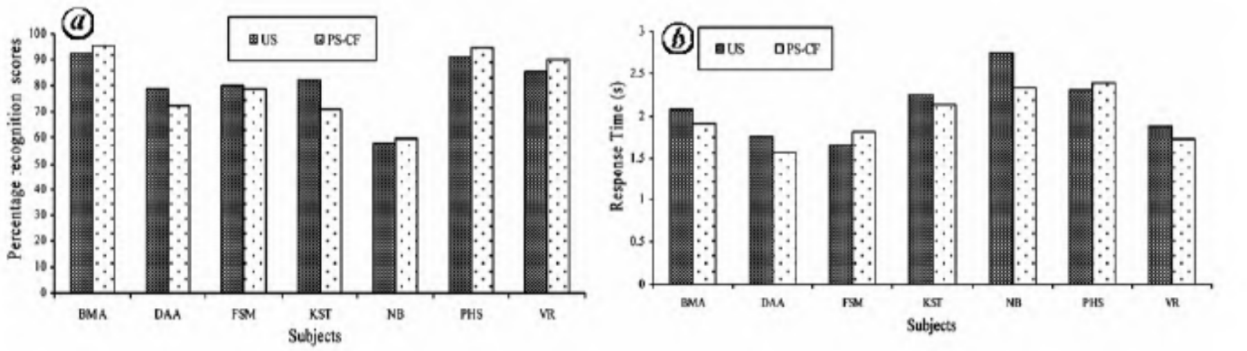

Figure 3. Processing scheme of comb filters. ***a***, Recognition score; ***b***, Response time. US: Unprocessed signal, PS-CF: Processed signal with comb filters.

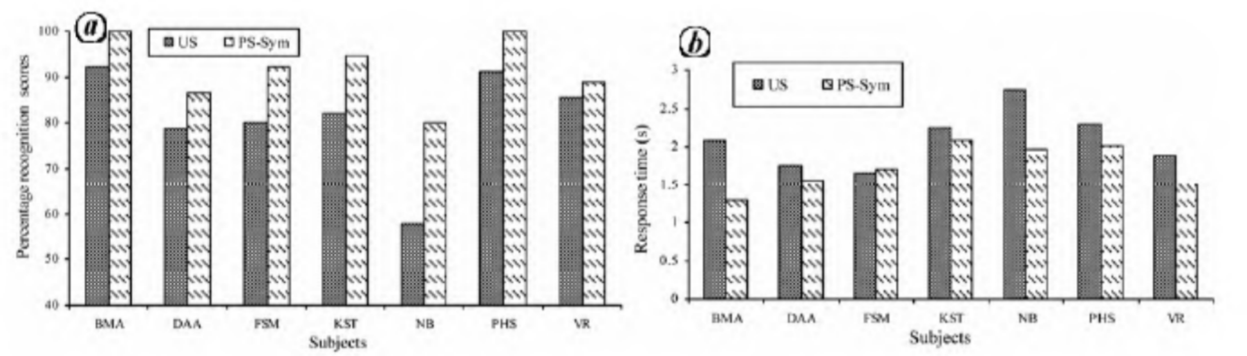

Figure 4. Processing scheme of wavelet packets. ***a***, Recognition score; ***b***, Response time. US: Unprocessed signal, PS-Sym: Processed signal with Symlet wavelet packet.

response time. Thus, filtering scheme is a remedy for reducing the effect of spectral masking and improving speech perception. The system is useful to the hearing impaired persons and in order to estimate the detailed advantages of processing schemes, extended tests with hearing impaired subjects are needed.

1. Moore, B. C. J., *An Introduction to the Psychology of Hearing*, Academic Press, California, USA, 1997.
2. Pickett, J. M., *The Acoustics of Speech Communication*, Allyn and Bacon, MA, USA, 1999.
3. Chaudhari, D. S., Dichotic presentation for improving speech perception by persons with bilateral sensorineural hearing impairment, Ph D thesis, School of Biomedical Engineering, IIT, Mumbai, 2000.
4. Moore, B. C. J., and Glasberg, B. R., *J. Acoust. Soc. Am.*, 1983, **74**, 750–753.
5. Zwicker, E., Flottorp, G. and Stevens, S. S., *J. Acoust. Soc. Am.*, 1957, **29**, 548–557.
6. Baskent, D., *J. Acoust. Soc. Am.*, 2006, **120**, 2908–2925.
7. Proakis, G. and Monoakis, D. G., *Digital Signal Processing Principles, Algorithms and Applications*, Prentice Hall, NJ, USA, 2004.
8. Daubechies, I., *Ten Lectures on Wavelets*, SIAM, CBMS-NSF Regional Conference in Applied Mathematics, Philadelphia, 1992, vol. 61.
9. Burrus, C. S., Gopinath, R. A. and Guo, H., *Introduction to Wavelets and the Wavelet Transforms: A Primer*, Prentice Hall, Upper Saddle River, NJ, 1998.
10. Nogueira, W., Giese, A., Edler, B., and Biiuchner, A., Proceedings of the IEEE International Conference on Acoustic Speech, Signal Processing (ICASSP '06), 2006, vol. 5, pp. 14–19.
11. MATLAB is a trademark of The Mathworks, Inc., Natick, USA; http://www.mathworks.com/

ACKNOWLEDGEMENT. We thank DST, New Delhi for financial support.

Received 13 January 2008; revised accepted 27 January 2010

MAHESH T. KOLTE[1,*]
D. S. CHAUDHARI[2]

[1]*Shri Chhatrapati Shivajiraje College of Engineering,*
Pune 412 206, India
[2]*Government College of Engineering,*
Jalgaon 425 001, India
**For correspondence.*
e-mail: mtkolte@yahoo.com