Optimization of Wavelet Packets to Minimize the Effect of Spectral Masking for Improving Speech Perception

Pravin A. Dhulekar
Department of E & TC
Dr. B. A. T. University
Lonere, Maharashtra, INDIA

Sanjay L. Nalbalwar
Department of E & TC
Dr. B. A. T. University
Lonere, Maharashtra, INDIA

ABSTRACT
Spectral masking occurs when the perception of one sound is affected by the presence of another sound like noise or unwanted sound of the same duration as the original sound. Earlier studies have shown that binaural dichotic presentation, using a pair of linear phase FIR comb filters with complementary magnitude responses, helps in reducing the effect of spectral masking. In the present study a spectral splitting of speech signals is done by using improved wavelet packets that provides a pair of bands of pre-recorded speech signals, which are presented dichotically that is odd bands of frequencies are given to the right ear and even bands to the left ear simultaneously. The performance of the proposed method is experimentally evaluated with speech signals of vowel-consonant-vowel syllables for twenty one English consonants. The result of Qualitative assessment, Response time and Recognition scores obtained after conducting listening tests on the five subjects, shows the effectiveness of method of improved wavelet packets over the comb filter.

General Terms
Signal Processing, Wavelet Representations et. al.

Keywords
Spectral Masking, Dichotic Presentation, Comb filter and Improved Wavelet Packets

1. INTRODUCTION
Spectral masking or simultaneous masking is a frequency domain version of temporal masking, and tends to occur in sounds with similar frequencies: a powerful spike at 1 kHz will tend to mask out a lower-level tone at 1.1 kHz. It is masking between two concurrent sounds. Sometimes called frequency masking since it is often observed when the sounds share a frequency band e.g. two sine tones at 440 and 450Hz can be perceived clearly when separated. They cannot be perceived clearly when simultaneous. In masking a sound is made inaudible by a "masker", a noise or unwanted sound of the same duration as the original sound [1].

The greatest masking is when the masker and the signal are the same frequency and these decreases as the signal frequency moves further away from the masker frequency. This phenomenon is called on-frequency masking and occurs because the masker and signal are within the same auditory filter. The simultaneous masking reduces the frequency resolution significantly, due to which it is more severe compared to the non-simultaneous masking. The auditory masking occurs because the original neural activity caused by the first signal is reduced by the neural activity of the other sound [2].

The splitting of speech signal into the two channels or bands can be carried out in a number of ways. Lunner et a1 [3] have used 8-channel constant bandwidth filtering, while Kulkarni P.N. et al [4] have used three pair of comb filter 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 bands with complementary spectra on the basis of critical band filtering for binaural dichotic presentation by using improved wavelet packets as an effective solution to problem of spectral masking. The study was carried out by processing speech signals of vowel-consonant-vowel syllables for twenty one English consonants and listening tests were conducted on five normal hearing subjects with simulated sensorineural hearing loss.

2. MATERIALS AND METHODS

2.1 The Speech Material and Subjects
Earlier studies have used CV, VC, CVC, and VCV syllables. It has been reported earlier that greater masking takes place in intervocalic consonants due to the presence of vowels on both sides [5]. Since our primary objective is to study improvement in consonantal identification due to reduction in the effect of masking, so VCV syllables are used. For the evaluation of the speech processing strategies, a set of twenty one nonsense syllables in VCV context with consonants / p, q, b, c, t, d, h, j, k, g, m, n, s, z, f, v, w, x, r, l, y / and vowel /*a*/ as in farmer were used. The features selected for study were voicing (voiced: / b d g m n z v r l y / and unvoiced: / p t k s f /), place (front: / p b m f v /, middle: / t d n s z r l /, and back: / k g y /), manner (oral stop: / p b t d k g l y /, fricative: / s z f v r /, and nasals: / m n /), nasality (oral: / p b t d k g s z f v r l y /, nasal: /m n /), frication (stop: / p b t d k g m n l y /, fricative: / s z f v r /), and duration (short: / p b t d k g m n f v l / and long: /s z r y /). Five subjects (VBA: M 23, MKA: M 25, AKS: M 26, ABC: M 27, CBB: M 31) participated in the listening test.

2.2 The Speech Processing Strategies
The discrete wavelets transform results in a logarithmic frequency resolution. High frequencies have wide bandwidth whereas low frequencies have narrow bandwidth. Wavelet Packets allow for the segmentation of the higher frequencies into narrower bands. Wavelet packets are efficient tools for speech analysis, involve using two-band splitting of the input signal by means of filtering and downsampling at each decomposition level. Designing the wavelet packets 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. Discrete wavelet transform for one level of decomposition and wavelet packets for the second level of decomposition referred as improved wavelet packets is used. Down-sampling and up-sampling are used with the wavelet based filter banks to exploit the spectral properties such as energy levels and perceptual importance. The signal is transform so that the power spectrum tends to concentrated into a few bands. MatLab software used to implement the wavelet packet based algorithm uses a natural order index to label nodes [6], [7].

Simulink models were developed based on improved wavelet packet with Daubechies wavelet function. The inverse wavelet packets transform was used to synthesize speech components from the wavelet packet representation [8]. To synthesize the speech component, wavelet coefficients were used. Table 1 and Table 2 shows frequency bands based on quasi-octave.

Table 1. Ten frequency bands for spectral splitting with compression (For left ear)

Filter for left ear		
Band	**Centre frequency kHz**	**Pass band frequency kHz**
1	0.078125	0-0.15625
3	0.390625	0.3125-0.46875
5	0.78125	0.625-0.9375
7	1.5625	1.250-1.875
9	3.125	2.500-3.75

Table 2. Ten frequency bands for spectral splitting with compression (For right ear)

Filter for right ear		
Band	**Centre frequency kHz**	**Pass band frequency kHz**
2	0.234375	0.15625-0.3125
4	0.546875	0.46875-0.625
6	1.09375	0.9375-1.25
8	2.1875	1.875-2.5
10	4.375	3.75-5

Daubechies wavelet is orthogonal wavelets that have the highest number of vanishing moments for a givens support width. The wavelet Filter Bank is a filter bank that offers a great deal of flexibility in terms of the choice of the basis filter and the decomposition tree structure. The standard DWT involves a dyadic tree structure in which the low-channel side is successively split down to a certain depth. We obtain the detail coefficients from the right-leaf node of each level and the approximation coefficients from the left-leaf node at the lowest level. WP filter bank 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. During the process of transformation, compression is achieved. A wavelet packet tree for a decomposition depth of 2 generated using the natural order index labeling of MatLab was presented in Figure 1, where the nodes represent the wavelet coefficients (at various decomposition stages) and the left and right branches represent the low- and high-pass filtering operations, respectively[9],[10].

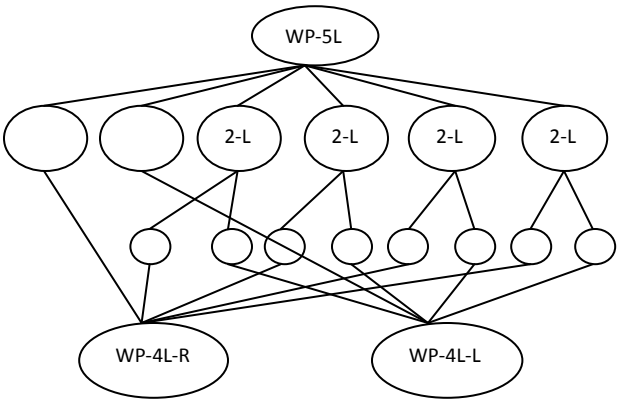

Fig 1: Decomposition tree for improved wavelet packets.

2.3 Experimental Procedure

Presentations were completed at the comfortable listening level for the subject. These subjects were tested without adding any noise to the speech stimuli. Procedure of 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, twenty one stimulus items were presented for six times, in a random order, leading to a total number of ninety 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 were 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 score and response time statistics were also presented.

A compilation of subject qualitative assessment of the test stimuli procedure under various listening conditions was carried out for learning the speech qualitative analysis. Listening tests were carried out for binaural diotic presentation of unprocessed

speech and binaural dichotic presentation of processed speech. An experimental setup using personal computer/laptop was used for binaural presentation of the test stimuli and subject's responses noted. The subject's task was to grade the processed signal with the unprocessed signal. Response times were used to assess the effectiveness of the processing schemes in reducing load on perception. Response times, Percentage correct recognition scores, Relative Decrease in response times and Relative improvement in recognition scores for unprocessed and processed signals for each subjects are discussed in the following subsections.

3. TEST RESULTS

A Comparative Results of listening tests for Qualitative assessment, Response time, Relative decrease in response times, Percentage correct recognition scores and Relative improvement in recognition scores obtained with comb filter and wavelet packets are presented in the following subsections.

3.1 Qualitative assessment

The qualitative assessment was carried out with grading below average, average, good, very good and excellent, ranking in ascending order. The subjects asked to pay attention the word three times and results were recorded on mean basis. Results of qualitative assessment of five subjects are given in Table 3 and shown in Figure 2. From the table it can be seen that, three subjects VBA, MKA, ABC and CBB ranked the quality of all schemes as the higher than the unprocessed signal. Subject AKS with ps-CF has ranked the lowest quality than others. All Subjects with processed scheme ps-DB ranked highest quality than unprocessed as well as processed scheme with ps-CF.

Table 3.Qualitative assessment

Subjects	Unprocessed Signal(US)	Processed Signal(ps-CF)	Processed Signal(ps-DB)
VBA	3.33	3.67	4.67
MKA	3.33	3.67	3.67
AKS	3.00	2.33	2.67
ABC	2.00	3.00	3.00
CBB	3.00	4.00	4.00

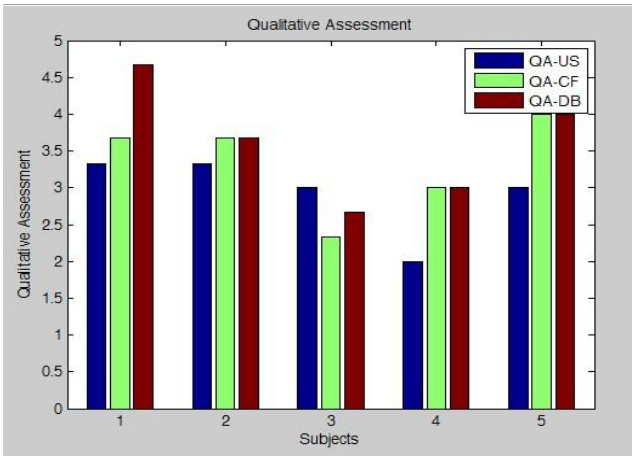

Fig 2.Comparative Result of Qualitative Assessment

3.2 Response Time

Table 4 gives the response times for unprocessed, processed signal with comb filter, processed signal with db for the five subjects VBA, MKA, AKS, ABC and CBB. The Table 5 provides relative decrease in response time (processed vs. unprocessed).

Table 4.Response Time

Subjects	Unprocessed Signal(US)	Processed Signal(ps-CF)	Processed Signal(ps-DB)
VBA	1.86	1.69	1.58
MKA	2.25	2.32	1.91
AKS	2.03	1.78	1.62
ABC	1.64	1.76	1.65
CBB	1.72	1.60	1.68

Table 5.Relative Decrease in Percentage (%)

Subjects	Processed Signal(ps-CF)	Processed Signal(ps-DB)
VBA	9.1398	15.0538
MKA	-3.1111	15.1111
AKS	12.3153	20.1970
ABC	-7.3171	-0.6098
CBB	6.9767	2.3256

Figure 3 shows the response times of the five subjects for unprocessed and processed signal. Figure 4 shows the percentage relative decrease in response times for processed signal for the five subjects. For unprocessed signal, response times varied from 1.64 to 2.25 seconds. With processing schemes ps-CF and ps-DB response times were decreased. Relative decrease in response times for processed schemes ps-CF, ps-DB were ranged from -7.3171 to 12.3153%, -0.6098 to 15.1111% respectively. Relative decreases in response times were statistically significant for the subject VBA and AKS. There is relative increase in response times for the subjects MKA and ABC.

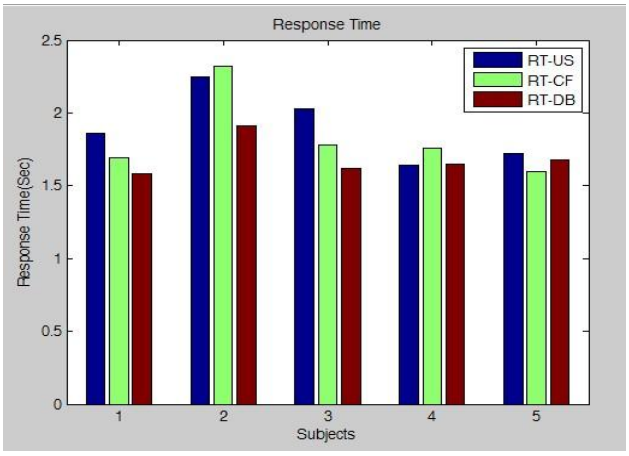

Fig 3.Response Time

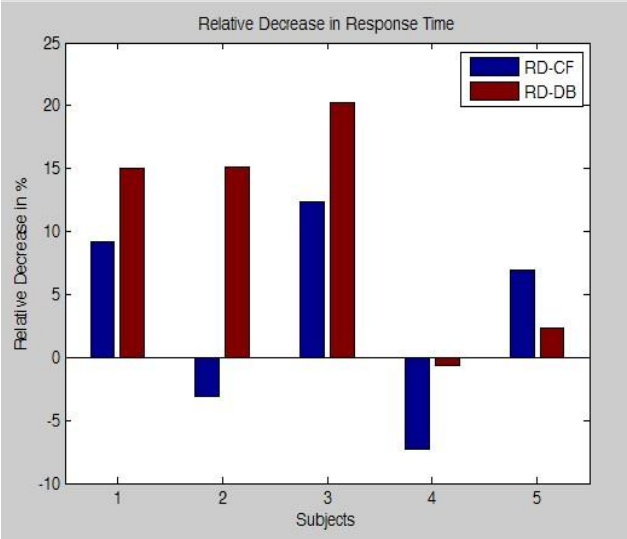

Fig 4.Relative decrease in Response Time

3.3 Recognition Scores

The recognition scores for unprocessed and processed signal and percentage relative improvement for processed signal for the five subjects are given in Table 6 and 7. Figure 5 shows percentage recognition scores for unprocessed and processed signal. Recognition scores of the five subjects were low for unprocessed signal and varied from 73.34 to 93.34%. Processing with all the schemes improved the scores for subjects, except scores reduces for subject MBT for ps-CF scheme. The relative improvements in recognition scores for processed schemes ps-CF, ps-DB were ranged from -13.33 to 6.67%, 0 to 6.67%.

Table 6.Percentage Recognition Scores

Subjects	Unprocessed Signal(US)	Processed Signal(ps-CF)	Processed Signal(ps-DB)
VBA	93.34	100	100
MKA	73.34	73.34	86.67
AKS	80	80	86.67
ABC	86.67	93.34	93.34
CBB	80	86.67	80

Table 7.Relative Improvement in % with respect to Unprocessed

Subjects	Processed Signal(ps-CF)	Processed Signal(ps-DB)
VBA	6.66	6.66
MKA	00	13.33
AKS	00	6.67
ABC	6.67	6.67
CBB	6.67	00

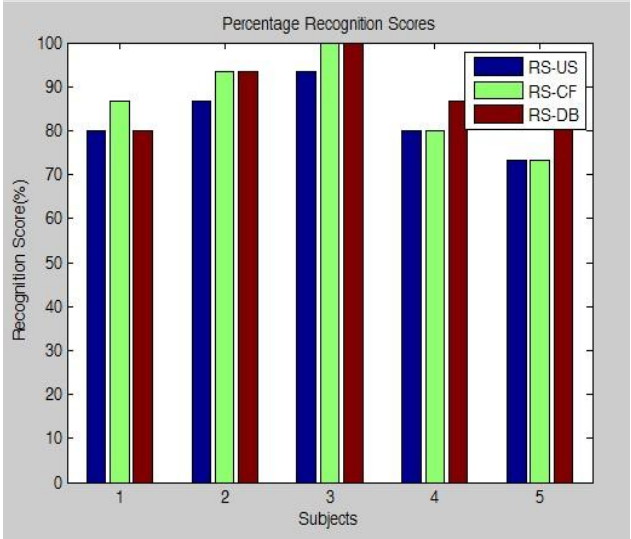

Fig 5.Percentage Recognition Scores

4. DISSCUSION AND CONCLUSIONS

An overall evaluation of the processing schemes ps-comb and ps-DB was done by conducting listening tests on five subjects. Listening tests were conducted using twenty one English consonants in VCV context. Subject's Qualitative assessment, Response times and Recognition scores were analyzed. Qualitative assessment shows that speech quality of unprocessed signal gets degraded most. Processing improved the speech quality significantly particularly the voicing, duration, and frication features.

There was a decreased response time for all the processing schemes compared with unprocessed signal, signifying reduction in burden on perception process. Relative decrease in response time was statistically significant for the processing schemes ps- DB compared to ps-CF. The extent of improvement was highest for ps-DB for all the subjects. This indicates that ps-DB is more effective in reducing perceptual load. Recognition scores indicate that binaural dichotic presentation improved consonantal identification.

All the results of qualitative assessment, response time and recognition scores of processed signal obtained by improved wavelet packets with daubechies family is much better than that of obtained with comb filter.

From the analysis, it is observed that these schemes gives maximum benefit by reducing the effects of increased masking depends on the individual hearing impairment configuration. Persons with low frequency hearing impairment and high frequency hearing impairment benefit from the processing scheme. Reception of the relatively robust consonantal features (voicing, manner, and nasality) also improves because of dichotic presentation. Hence the processing schemes for dichotic presentation have the potential of improving speech perception for persons using binaural hearing aids[11].

In listening tests involving the hearing impaired subjects, the subjects are able to perceptually integrate the dichotically presented speech signal, and the presentation results in improved speech reception. As indicated by the improvements in response time, dichotic presentation also decreases the load on the

perception process. For hearing impaired subjects, the improvement in consonantal reception and reduction in response time do not follow the same trend [12]. Therefore, in order to estimate the detailed advantages of processing schemes, extended tests with hearing impaired subjects are needed.

5. REFERENCES

[1] B. C. J. Moore, An Introduction to Psychology of Hearing. 4th ed. London: Academic, 1997.

[2] B. C. J. Moore, "Speech processing for hearing impaired: successes, failures, and implications for speech mechanisms," Speech Communication., vol. 41, pp. 81-91, 2003

[3] Lunner, T., and Hellgren, J. (**1991**). "A digital filterbank hearing aid- design, implementation, and evaluation," *Proc. in IEEE ASSP*, vol. **5**, 3661–3664.

[4] P.N.Kulkarni and P.C.Pandey,"Optimizing the comb filters for spectral splitting of speech to reduce the effect of spectral masking," in Proc. International conference on signal processing, Communication and Networking (MIT,Chennai,India), pp.69-73,Jan 4-6, 2008.

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

[6] T. Arai, K. Yasu, and N. Hodoshima, "Effective speech processing for various impaired listeners," *Proc. 18th International Congress Acoustics (ICA)*, 2004, pp. 1389 - 1392.

[7] Chaudhari D. S. and Pandey P. C. "Dichotic presentation of speech signa1 using critical filter bank for bilateral sensorineural hearing impaired". *Proceedings* of *I6'* International Congress on Acoustics*, Seattle, Washington, 1998, vol. 1, pages 213-214

[8] I. Cheikhrouhou, R.B. Atitallah, K. Ouni, A.B. Hamida, N. Mamoudi,and N. Ellouze, "Speech analysis using wavelet transforms dedicated to cochlear prosthesis stimulation strategy," 1st Intern. Symp. On Control, Communications and Signal Processing, 2004, pp. 639–642.

[9] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. on Patt. Anal. Machine Intell*, vol. 11(2), pp. 674-694, 1989.

[10] G. Tognola, F. Grandori, P. Ravazzani, "Wavelet analysis of clickevoked otoacoustic emissions," *IEEE Trans. Biomed. Eng.*, vol. 45, pp. 686-697, 1998.

[11] A. N. Cheeran, and P. C. Pandey, "Evaluation of speech processing schemes using binaural dichotic presentation to reduce the effect of masking in hearing-impaired listeners," in Proc. 18th International Congress on Acoustics (ICA 2004, Kyoto, Japan), pp. 1523 - 1526, Apr. 4–9, 2004.

[12] D. S. Chaudhari, and P. C. Pandey, "Dichotic presentation of speech signal with critical band filtering for improving speech perception," in Proc. ICASSP '98, Seattle, Washington, USA.