Wavelet Filter Banks Modeling of Human Auditory System for Robust Speech Enhancement

Ranganadh Narayanam,Hilmi Dajani and Sos Agaian

Abstract—In this research we have developed a novel perceptual wavelet filter bank architecture & wavelet filter banking coefficients for a versatile speech enhancement method based on the human auditory model. In this paper the implementation of these wavelet filer banking coefficients for a speech enhancement scheme are being described which meets the demand for quality noise reduction algorithms which are capable of operating at a very low signal to noise ratio. This is a generalized time frequency subtraction algorithm which advantageously exploits the wavelet multi-rate signal representation to preserve the critical transient information. This wavelet filer banking may be able in reducing noise in applications with little speech degradation in diverse noise environments by reducing the residual noise and improve the intelligibility of speech. MATLAB routines are developed for performing this research.

Index termas- Time-Frequency analysis, Filter banking, Robust detection, Wavelet banking, perceptual wavelet packet transform

- - - - - - - - - - - - - - - - - - - ♦ -

1 Introduction

The performance of the automatic speech processing systems degrade drastically when confronted with a great adverse noise conditions such as background noise and micro phone distortions. For this reason there is a strong demand for quality reduction algorithms capable of operating at very low signal to noise ratio in order to combat various forms of noise distortion. The solutions can be classified into two[1] main areas a) nonparametric; usually remove an estimate of the distortion from the noisy features, and b) statistical model based speech enhancements, statistical model based speech enhancement utilizes a parametric model of the signal generation process [2]. This paper research is based on a speech enhancement system which can be based on subtractive type algorithms. By subtracting the noise estimation from the noisy speech this system estimates the short time spectral magnitude of speech. One of the important parts of our main research (including our underway research) consists a generalized perceptual time-frequency subtraction method based on the masking properties of the human auditory system [4], this works in conjunction with a perceptual wavelet packet transform (PWPT) to reduce the effect of noise contamination. We describe in this paper is the proposed MATLAB implemented perceptual wavelet filer bank architecture. The main theme in this proposed method is the use of PWPT to approximate 24 critical bands of the human auditory system up to 16 kHz. It enables the components of complex sound to be appropriately segregated in frequency and time in order to mimic the frequency selectivity and temporal masking of the human auditory system. This proposed MATLAB implemented method uses PWPT to analyze for the application of improving the perceptual quality of the final processed speech. "Brain stem speech evoked potentials" research data [8] was partly required for this implementation of this paper research.

2 Proposed Perceptual Wavelet Filter Bank Architecture

Architecture for the perceptual wavelet filter bank: To design this algorithm for enhancing speech a well built psychoacoustic model of the ear which has an unsurpassed capability to adapt to noise. In this a new human auditory model that adapts to the basic structure of traditional auditory model but replace the time invariant band pass filters with WPT in order to mimic the time-frequency analysis of the critical bands according to the hearing characteristics of human cochlea [10]. A PWPT is used to decompose the speech signal from 20 Hz to 16 KHz Into 24 frequency sub -bands that approximate the critical bands, efficient seven level tree structures is implemented. This is given in the Fig 1. Two channel wavelet filter banks are used to split the low pass and high pass bands as opposed to only the low pass and high pass bands in the usual wavelet decomposition. Advantages: first, Smoothness property of wavelet is determined by the number of vanishing moments: more the vanishing moments the stringent bandwidth and stop band attenuation of each sub-band and can be more close approximation by using the wavelet decomposition. Second, according to the psychoacoustic study of human ears a frequency to bark transformation [5] needs to be performed which can be accomplished in audio processing systems by dividing the frequency range into critical bands. Using the perfect reconstruction filter bank with finite length filters using different wavelets for the analysis and synthesis scaling functions [5, 6, 7]. Let H(z) and G(z) be the low pass (LP) and high pass (HP) transfer functions, before the decimation by two operation in each stage of the analysis filter bank. F (z) and J(z) be the LP and HP transfer functions, after the up sampling by two operation in each stage of the synthesis filter bank. Then the analysis and synthesis filter banks are related by

$$g(n) = (-1)^{n} f(n) \leftrightarrow G(z) = F(-z)$$
$$j(n) = - (-1)^{n} h(n) \leftrightarrow j(z) = -H(-z) \quad (1)$$

The relationship between the LP and HP filters reduces the number of filters to be implemented for each stage of the two-channel filter bank by half. Once the LP filters, H (z) and F (z) are designed the HP filters G(z) and J(z) can be derived from the equation (1). According frequency selectivity related to critical band, temporal resolution of the human ear, and regularity property of wavelets debauchies wavelet basis is chosen as prototype filter and a seven stage WPT is adopted to build perceptual wavelet filter bank. Table I shows the mapping of the PWPT coefficients in each stage. Table II shows the comparison of lower (f_l) and upper (f_u) frequencies, center frequency (f_c) and bandwidth (Δf) in hertz between the critical band rate and the proposed perceptual wavelet packet tree scale. The critical band rate Z_B in bark is approximated by where the frequency, f , is measured in Hz.

$$Z_B = 13 \tan^{-1}(7.6 \times 10^{-4} f) + 3.5 \tan^{-1}(1.33 \times 10^{-4} f)^{2} \quad (2)$$

In case of bandwidths of the critical bands and the perceptual wavelet packet tree, critical bands have constant width at approximately 100 hz for centre frequencies upto 500 hz, and the bandwidths increase as the centre frequency increases. The critical bandwidth (CBW) is hertz is calculated by

$$CBW(f) = 25 + 75 (1 + 1.4 \times 10^{-6} f^{2})^{0.69} \quad (3)$$

Figure 2 compares the absolute threshold of hearing (ATH) [8,9] in hertz, critical band scale[6] , and perceptual wavelet packet scale. The ATH characterizes the amount of energy needed in a pure tone such that it can be detected by a listener in a pure tone such that it can be detected by a listener in a noiseless environment. The table II and figures given it makes clear regarding the proposed perceptual wavelet packet tree can closely mimic the experimental critical bands. The parameters of the discrete WPT filter used to derive the plots of figures are determined based on the auditory masking properties.

3 Conclusion

The system can consist of functional stages working cooperately to perform perceptual time-frequency subtraction[7,9] by adapting the weights of the perceptual wavelet coefficientceints. The noisy speech is first decomposed into critical bands by perceptual wavelet transform. We have developed MATLAB implementation. With this proposed wavelet filter banking architechure we proved that we can mimic the experimental critical bands. So this may be able in reducing noise in applications with little speech degradation in diverse noise environments by reducing the residual noise and improve the intelligibility of speech. We are under research for developing a speech enhancement system for the application of this wavlet filer bank architecture applicable to GPTFS method [8,9] for noise reduction & unvoiced speech [8] will also be enhanced by a soft thresholding scheme.

Author's biography: Mr. Ranganadh Narayanam, is currently an assistant professor in Dept of ECE in Marri Laxman Reddy Institute of Technology,Hyderabad. Mr. Narayanam had done research as research assistant in University of Texas at San Antonio, University of California at Berkeley and University of Ottawa under Dr. CJ Qian,Dr. Parimal Patel,Dr. Artyom Grigoryan, Dr. Sos Agaian, Dr. Ben Inglis and Dr. Hilmi Dajani in the areas of contorl systems,signal processing,digital systems and neurological signal processing. This research was partially guided by Dr. Hilmi Dajani of university of ottawa and Dr. Sos Again of University of Texas at San Antonio. Ranganadh Nnarayanam can be contacted at Marri Laxman Reddy Institute of Technology,Department of Electronics and Communications Engineering,Hyderabad,AP,India. rnara100@uottawa.ca , rnara100@gmail.com

4 References

[1] In-Chul Yoo and Dongsuk Yook, "Robust voice activity detection using the spectral spectral peaks of vowel sounds". ETRI Journal,Volume 31,Number 4,August 2009.

[2] R. Nicole, J. Sohn, N.S. Kim, and W. Sung, "A statistical Model Based Voice Activity Detection," *IEEE Signal Process. Lett.*, vol. 6, 1999, pp. 1-3.

[3] Javier Ramirez,Jos C segura, Carmen Benitez,Angel de la torre, Antonio Rubio, "Efficient voice activity detection algorithms using long-term speech information", J. Ram_ırez et al. / Speech Communication 42 (2004) 271–287.

[4] I. Krekule, "zero crosssing detection of the presence of evoked responses", Electroencephalography and clinical neurophysiology, Elsevier publishing company, Amsterdam – Printed in the netherlands.

[5] GBrson Eduardo Mog, Eduardo Parente kbeiro, "Zero Crossing determination by linear interpolation of sampled sinusoidal signal.

[6] Dajani, R.H., Purcell, D., Wong, W., Kunov, H., Picton, T.W. 2005. Recording Human Evoked Potentials That Follow the Pitch Contour of a Natural Vowel. IEEE Transactions on Biomedical Engineering 52, 1614-1618.

[7] Johnson, K.L., Nicol, G.T., Kraus, N. 2005. Brain Stem Response to Speech: A Biological Marker of Auditory Processing. Ear & Hearing 26, 424-434.

[8] Russo, N., Nicol, T., Musacchia, G., Kraus, N. 2004. Brainstem responses to speech syllables. Clinical Neurophysiology 115, 2021-2030.

[9] Galbraith GC, Arbagey PW, Branski R. Intelligible speech encoded in the human brain stem frequency-following response. *NeuroReport* 1995; 6: 2363-2367.

[10] M.S. John, T.W. Picton, MASTER: a Windows program for recording multipleauditory steady-state response. Computer Methods and Programs in Biomedicine 61 (2000) 125–150, Elsevier.

Table 1 perceptual wavelet filter banks coefficients.

Sub band Z_w	l	Coefficients k_a-k_b	Transform stage j
1	1	0-0	7
2	1	1-1	7
3	1	2-2	7
4	1	3-3	7
5	1	4-4	7
6	1	5-5	7
7	1	6-6	7
8	1	7-7	7
9	2	8-9	6
10	2	10-11	6
11	2	12-13	6
12	2	14-15	6
13	2	16-17	6
14	2	18-19	6
15	4	20-23	5
16	4	24-27	5
17	4	28-31	5
18	8	32-39	4
19	8	40-47	4
20	8	48-55	4
21	8	56-63	4
22	16	64-79	3
23	16	80-95	3
24	32	96-127	2

Table 2 Critical band rate Z and perceptual wavelet filter banks W

Z	Bark Scale	Bark Scale	Bark Scale	Wavelet scale	Wavelet scale	Wavelet scale
	[f_l f_u]	f_c	Δ_f	[f_l f_u]	f_c	Δ_f
1	[0 100]	50	100	[0 125]	62.5	125
2	[100 200]	150	100	[125 250]	187.5	125
3	[200 300]	250	100	[250 375]	312.5	125
4	[300 400]	350	100	[375 500]	437.5	125
5	[400 510]	450	110	[500 625]	562.5	125

6	[510 630]	570	120	[625 750]	687.5	125
7	[630 770]	700	140	[750 875]	812.5	125
8	[770 920]	840	150	[875 1000]	937.5	125
9	[920 1080]	1000	160	[1000 1250]	1125	125
10	[1080 1270]	1170	190	[1250 1500]	1375	250
11	[1270 1480]	1370	210	[1500 1750]	1625	250
12	[1480 1720]	1600	240	[1750 2000]	1875	250
13	[1720 2000]	1850	280	[2000 2250]	2125	250
14	[2000 2320]	2150	320	[2250 2500]	2375	250
15	[2320 2700]	2500	380	[2500 3000]	2750	500
16	[2700 3150]	2900	450	[3000 3500]	3250	500
17	[3150 3700]	3400	550	[3500 4000]	3750	500
18	[3700 4400]	4000	700	[4000 5000]	4500	1000
19	[4400 5300]	4800	900	[5000 6000]	5500	1000
20	[5300 6400]	5800	1100	[6000 7000]	6500	1000
21	[6400 7700]	7000	1300	[7000 8000]	7500	1000
22	[7700 9500]	8500	1800	[8000 10000]	9000	2000
23	[9500 12000]	10500	2500	[10000 12000]	11000	2000
24	[12000 15500]	13500	3500	[12000 16000]	14000	4000

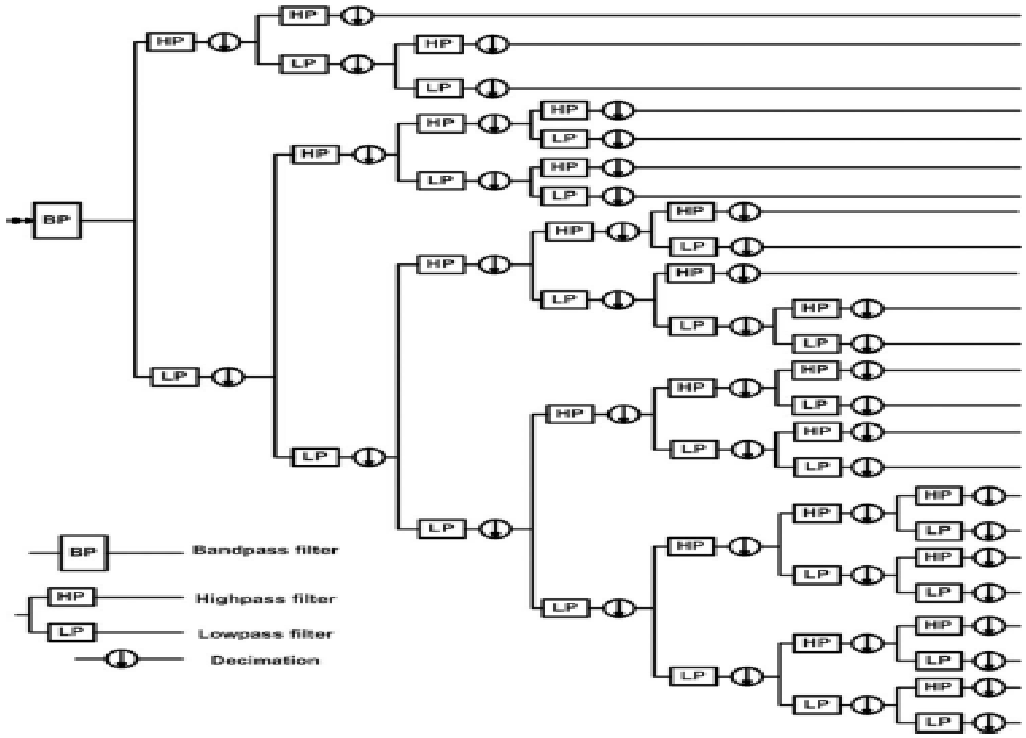

Figure 1. Perceptual wavelet packet decomposition tree (PWPT)

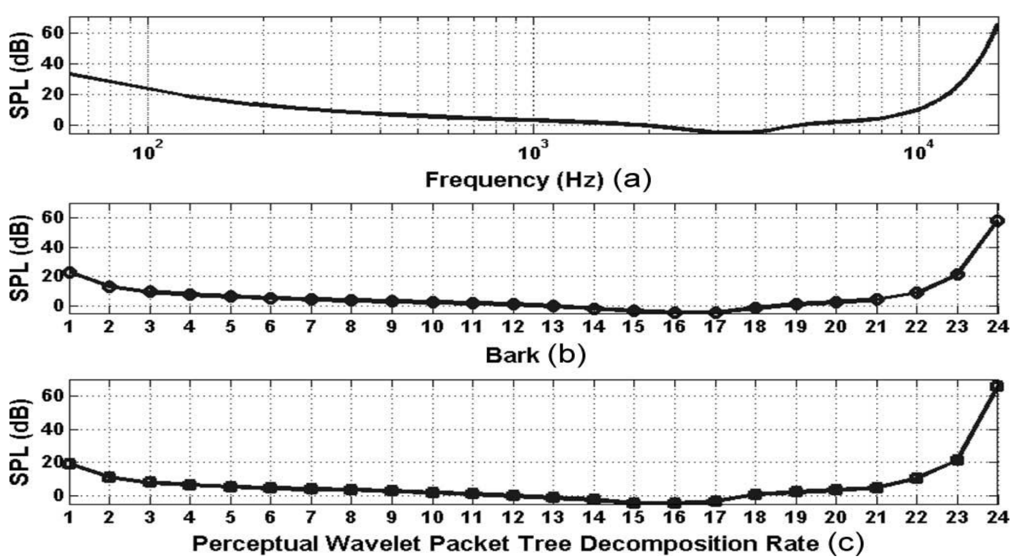

Figure 2. ATH in (a) frequency, (b) bark, and (c) perceptual wavelet packet tree scales.