International Journal of Technology (2015) 2: 190-197
ISSN 2086-9614

TRADITIONAL PSYCHOACOUSTIC MODEL AND DAUBECHIES WAVELETS FOR ENHANCED SPEECH CODER PERFORMANCE

Sheetal D. Gunjal[1*], Rajeshree D. Raut[2]

[1]*Department of Electronics Engineering, Amrutvahini College of Engineering, Pune Road, Near Pune Nashik Highway, Sangamner, Maharashtra 422608, India*
[2]*Department of Electronics Engineering, Ramdeobaba College of Engineering, Katol Rd, Nagpur, Maharashtra 440013, India*

(Received: February 2015 / Revised: April 2015 / Accepted: April 2015)

ABSTRACT

Speech compression techniques based on the traditional psychoacoustic model have been proposed by many researchers. We propose the Discrete Wavelet Transform (DWT) supported by the same psychoacoustic model for speech compression. This paper presents a traditional psychoacoustic model for processing equal partitions of the total bandwidth spectrum of audio signal frequencies in order to reduce redundancy by filtering out the tones and noise maskers in the speech signal. Here, uniform filter banks are used for efficient computation, for selection of appropriate threshold levels, and for better compression of Discrete Wavelet Transform coefficients. A Daubechies wavelet filter bank is nonlinear and asymmetric. It is equivalent to a cochlear filter in the human hearing system. The similarity between the Daubechies filter bank and our hearing system was the basis for developing a novel speech coder. Better performance in terms of the compression factor (CF) and the signal-to-noise ratio (SNR) resulted, as compared to the earlier methods.

Keywords: Daubechies wavelet; Discrete Wavelet Transform; Psychoacoustic model; Thresholding

1. INTRODUCTION

To meet the heavy demands of mobile and internet users with high fidelity and accuracy, speech compression has become a prime concern for present and future communication (Baumgarte, 2002). In this regard, speech coding with good quality reproducibility contributes much (Frank et al., 2002; Painter & Spanias, 1997). Speech compression is one of the most important signal processing steps that can reduce the requirements for transmission bandwidth, storage, and processing overhead time. Many algorithms, such as lossy or lossless algorithms, were proposed earlier for speech signal compression. Ours is a lossy algorithm in which part of the signal information is lost; nonetheless, the loss doesn't impinge on the reproducibility of the signal. The algorithm offers good compression as well as a good quality speech signal (Krimi et al., 2007; China et al., 2013).

In lossy compression, information redundancy is reduced by means of coding. The only information that is necessary to reproduce the signal is kept to ensure good playback quality (Jagdeesh et al., 2014). Parametric coders and waveform coders have also been proposed for compression of the speech signal. A parametric coder, such as the Code Excited Linear Predictive (CELP) coder, deals with the parameters that characterize filter behavior to encode

[*] Corresponding author's email: gunjalsheetal@yahoo.com, Tel. +91-02425 259016, Fax. +91-02425 259017

Permalink/DOI: http://dx.doi.org/10.14716/ijtech.v6i2.761

data and then the data are used by the decoder for speech synthesis. In contrast, waveform coders attempt to replicate most accurately the waveform of the original signal by exploiting the correlation between the selected filter bank and the signal components in the transform domain (Sheetal et al., 2012).

A transformation is applied to create a spectral representation of the input signal from the corresponding bank of sampled band pass filters. It analyzes the information from the input signal in terms of fewer coefficients to which processing steps such as thresholding and quantization can be applied for further performance improvement. In the proposed encoder (see Figure 1), we are suggesting the Discrete Wavelet Transform as the main coder, supported by the traditional psychoacoustic model to enhance the compression factor (CF), signal-to-noise ratio (SNR) and energy content of the speech signal. Discrete Wavelet Transform is an effective mathematical tool for compression of non-stationary signals as it is a time-frequency localization-based transform. The time-frequency characteristics of a wavelet filter bank match speech signal characteristics very well (Abid et al., 2010; Sheetal et al., 2012). This paper presents a speech coder based on the Discrete Wavelet Transform (DWT) using the Daubechies wavelet family. The Daubechies wavelet family was selected because of its smooth and compactly supported orthogonal low pass and high pass FIR filter banks.

2. PROPOSED SPEECH CODER

A block diagram of proposed speech coder is given below:

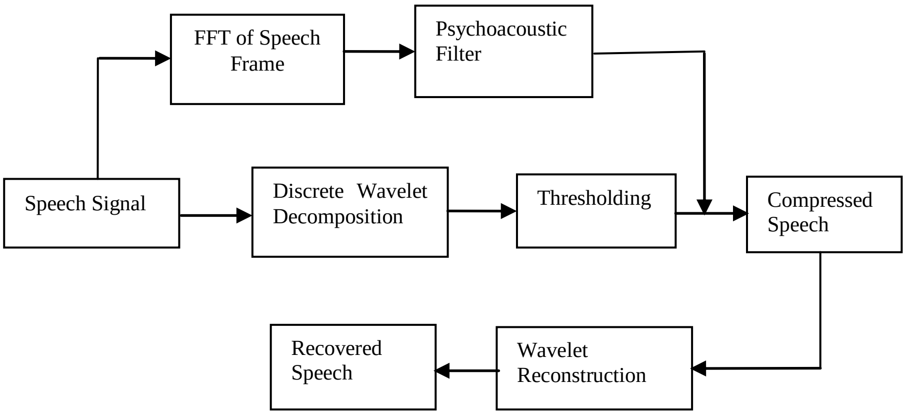


Figure 1 Proposed speech coder

The proposed speech coder consists of two important blocks: the psychoacoustic model and the Discrete Wavelet Transform in the form of Discrete Wavelet Decomposition and thresholding, with details given below.

2.1. The Psychoacoustic Model

The psychoacoustic model is an important component of the proposed encoder. It is based on research on human perception properties. The two main properties which characterize the psychoacoustic model are the absolute hearing threshold and auditory masking (Mourad et al., 2013; Abdul et al., 2003). Even though the human hearing frequency range is 20 Hz to 20 KHz, all frequencies are not heard in the same manner. A person can hear lower frequencies more accurately than higher frequencies. It indicates that our hearing system has a better ability to detect differences in pitch at lower frequencies. In addition, a signal whose frequency component lacks the power specified as the absolute threshold of hearing (ATH) can be removed (Sheetal et al., 2012). A very small frequency difference makes a low power signal inaudible to the listener; hence the signal needs a power boost greater than the frequency of the masker tone. If the signal is constant except for a sharp peak, then it is considered as a tone; otherwise it is a noise signal (Khalifa et al., 2005). Noise components that are detected in the critical band are added together to get the resulting noise component. The result of the noise components is used for passing/testing the threshold of the the transformed signal. The individual noise masker threshold can be represented by the following equation:

For tones:

$$P_{nm\,(j)_{-}} 0.27Z_{(j)\,+}\ SF_{(i,j)\,-}\ 6.025 \tag{1}$$

For noise:

$$T_{nm}\,(i,j) = P_{nm}\,(j) - 0.175Z_{(j)} + SF(i,j) - 2.025 \tag{2}$$

where $SF_{(i,j)}$ is a low level masking noise occurring in the tails of the basilar excitation pattern, and $P_{nm\,(j)}$ denotes the SPL of the noise masker in frequency bin j. $Z_{(j)}$ represents the bark frequency of bin j. For multiple tones and noise, the overall effect is additive. The psychoacoustic model steps are:

Perform a FFT analysis.

1. Calculate the energy in each frame.
2. Convolve the energy with the spreading function.
3. Calculate the tonality index (0 to 1) to separate tones.
4. Calculate the energy threshold for every frame.
5. Compare the value with the absolute threshold of hearing and consider the highest value as the energy threshold of the speech signal.

2.2. Wavelet Transform

To overcome the mismatch between uniform filter banks and spectral decomposition of the cochlea, a non-uniform cochlear filter bank was proposed for the psychoacoustic model (Shao et al., 2011; Trina et al., 2000). We use a traditional psychoacoustic model, and to overcome the above-mentioned problem, the Daubechies wavelet family was used. This is the only non-uniform FIR filter bank that combines perfect signal reconstruction with energy conservation and removes the aliasing effect. In DWT analysis, filters are used with different cutoff frequencies at different scales (Srinivasan & Jamieson, 1999; Davis et al., 1995). DWT offers a compact representation of the signal in the time and frequency domains along with efficient computation in the form of Equations 3 and 4.

$$d_{\mathrm{jk}} = \int x(t)\,\mathrm{dt} = 2^{\frac{j}{2}} \int x(t)\emptyset_{jk}\,(2^{\mathrm{j}}\ t - k)dt \tag{3}$$

$$\emptyset_{jk}\,(\mathrm{t}) = 2^{-\frac{j}{2}}\,\emptyset_{\mathrm{jk}}(2^{-j}t - k),\quad \mathrm{j,k}\ \in Z \tag{4}$$

where, d_{jk} is the wavelet coefficient and $x(t)$ is the time signal. The number of vanishing moments is related to the smoothness and flat frequency response of wavelet filters. A large number of vanishing moments produces a more compact signal. A vanishing moment must satisfy the condition given in Equation 5.

$$\int x^k\,\psi(x)\,dx = \ 0\,for\ 0 \le k \le K \tag{5}$$

where K is the degree of the polynomial (0 to K). However, the length of filter increases with the number of vanishing moments at the cost of complexity and time for computation (Davis et al., 1995). To process a speech signal, a decomposition level up to five is sufficient. Then the decomposed speech signal is ready for thresholding. Level-dependent thresholding and soft thresholding are applied to the decomposed signal (Agbinya et al., 2011). These types of thresholding help to modify the compression factor according to each application's needs. The threshold value is determined using a Birge-Massart strategy, which is well suited for speech signals. It processes the detail coefficients without disturbing the approximation coefficients. The number of detail coefficients is given by:

$$n = \frac{M}{(j+2-i)^{\propto}} \tag{6}$$

where level i starts from 1 to j, α is compression parameter, and M depends upon the number of approximation coefficients. This thresholding approach provides maximum absolute coefficients at each level. After hard thresholding, the transform vector needs to be compressed further. Here, we have used soft thresholding to achieve a good value for the compression factor. This method has produced improvement in the compression factor as compared to other methods used before.

3. RESULTS AND DISCUSSION

The software code for the proposed system was developed using MATLAB 2013a without any consideration of rate control for encoding purposes.

The performance parameters, that is, the compression factor (CF) and signal to noise ratio (SNR) of the coder for all five levels, are given in Table 1.

Table 1 Performance parameters in the proposed coder

Level	Parameter	DB 02	DB 04	DB 06	DB 08	DB 10
1	CR	1.97	1.98	1.99	1.99	1.99
	SNR	20.38	20.99	21.48	21.50	21.51
2	CR	3.77	3.80	3.83	3.85	3.85
	SNR	17.22	18.20	18.65	18.65	18.70
3	CR	6.68	6.90	6.98	6.99	7.02
	SNR	13.56	14.36	14.87	14.98	15.60
4	CR	7.58	7.71	7.64	7.72	7.73
	SNR	12.07	12.63	13.07	13.07	13.45
5	CR	7.60	8.07	8.53	8.82	9.05
	SNR	11.63	12.07	12.44	12.52	12.70

The dependencies of CR, SNR and the decomposition level is shown in Figures 2 and 3. As per Equation 5, the number of vanishing moments is determined by the length of the filter. An increase in vanishing moments naturally supports the increase in CR. However, instead of increasing the complexity of filter design, we can achieve the same end by increasing the number of decomposition levels. From Table 1, we can deduce that the compression ratio increases as the level number increases and increase in number of Daubechies filter bank i.e. DB02, DB04 and so on. Nonetheless, further increase in the CR value is limited by the SNR value.

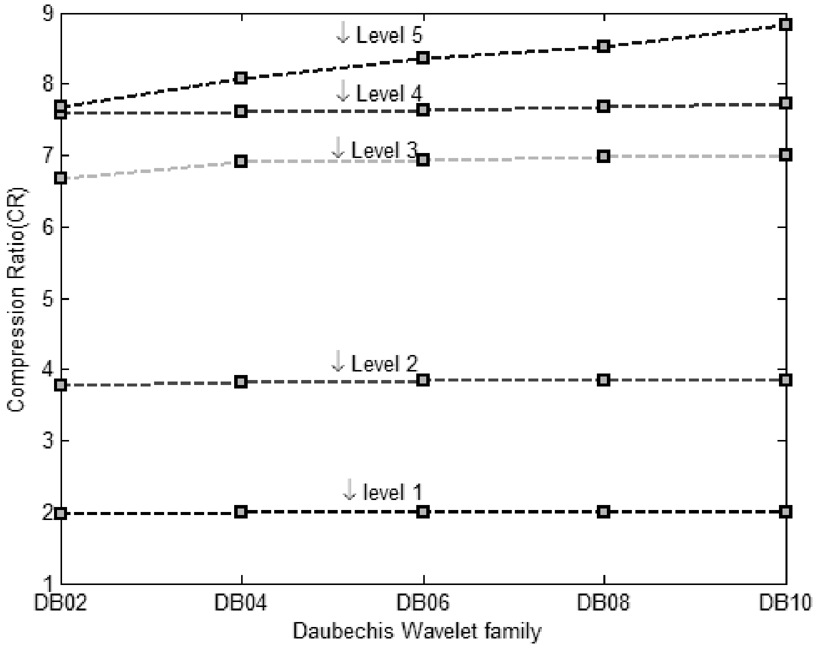


Figure 2 Daubechies wavelet family vs CR

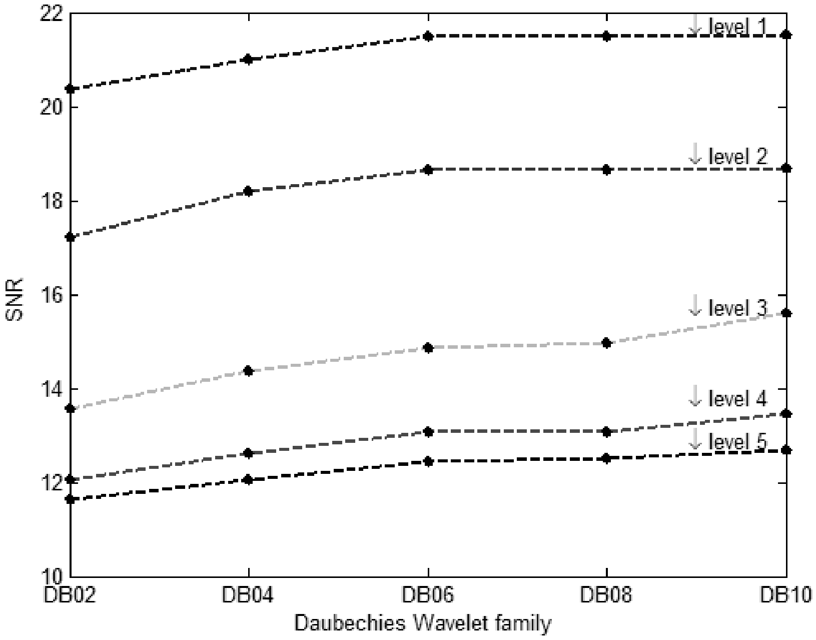
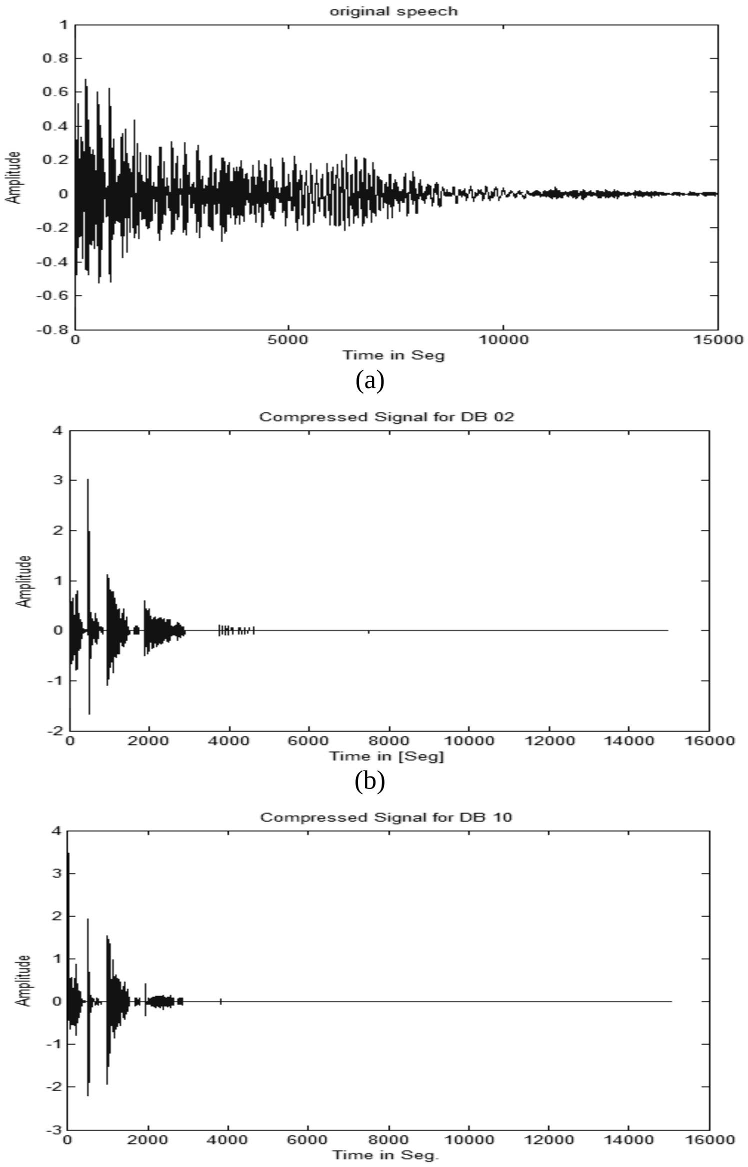


Figure 3 Daubechies wavelet family vs SNR

The number of vanishing moments is increased by means of an increase in the decomposition level and in the filter banks. Hence, at a higher power level, the signal part which is essential for accurate reproduction of the speech signal may get added into the vanishing moment. This causes deterioration of the signal quality in the compressed signal against the noise part. Therefore, the SNR value is higher at the minimum level. There is a slight increase in SNR with the Daubechies filter bank, as shown in Figure 3. The waveforms for the original speech signal and for the corresponding compressed signal at the fifth level are shown in Figures 4. These figures show that increase in the number of levels does not contribute further to listening clarity in the recovered signal.

Figure 4 illustrates an example of the compressed speech signal obtained by applying the proposed speech compression technique.



(a)

(b)

(c)

Figure 4 (a) Original speech signal; (b) compressed speech for DB 02; (c) compressed speech for DB 10

The second part of the evaluation is to compare the compression ratios obtained from the proposed coder with a classical MPEG1 coder, a coder based on the DWT transform, and the dynamic gammachirp psychoacoustic model. Table 1 presents the data measured at a constant bit rate of 372 Kbps and the result obtained in (Samar et al., 2007) is from different bit rates.

Table 2 Comparison of CR values in the classical MPEG1 coder, the DWT Coder and the proposed coder

Classical MPEG1 Coder 160 Kbps	DWT Coder 160 Kbps	Proposed Coder 372 Kbps
7.393	8.239	9.05

The bit rate of our proposed coder is 372 Kbps and that of the classical MPEG1 coder and DWT coder is 160 Kbps; the CR value is even higher. However, the slower the bit rate, the higher the compression ratio (Samar et al., 2007). Table 2 reveals that speech compression using the classical psychoacoustic model and the Daubechies Wavelet family is the best.

The performance of a coder was evaluated with a subjective listening test of recovered speech (Trina et al., 2000). The test was conducted for monophonic samples of five seconds' duration. The quality test was conducted on the basis of selected options by the listeners for the Daubechies wavelet filter bank (DB2, DB4,... DB10) and decomposition levels from one to five. The options given were: 'not comparable', 'comparable' and 'good'. The signal reconstructed with the higher degree filter bank was quite close to the original signal, but the same was not reflected in the quality of the output signal. For the DB10 family at levels four and five, the majority of the listener's remarks were 'comparable'.

4. CONCLUSION

Even though the traditional psychoacoustic model was used in our proposed coder, the selection of the Daubechies wavelet family with DWT yielded comparable improvement in the performance parameters with a good quality reconstruction of the speech signal. The compression factor improves at the cost of the SNR with progressive levels. At levels 3, 4 and 5, the variation in CF and SNR is much more consistent. One can select level 3 for good performance, with a moderate number of filter banks.

Adaptive filter banks can be used in combination with a psychoacoustic model for more effective coding. The proposed method can be modified for variable/smaller bit rates for monophonic CD quality audio signals and for mobile communication, such as cognitive radio, as the energy content of the recovered speech signal is maintained at 98% with slight variations at all levels using the Daubechies filter bank.

5. REFERENCES

Abdul Mawla M.A. Najih, Abdul Rahman Bin Ramli, V. Prakash, Syed A.R., 2003. Speech Compression using Discrete Wavelet Transform, *4th IEEE International Conference on Telecommunication Technology Proceedings,*

Abid, K., Ouni, K., Ellouze, N., 2010. Audio Compression Codec using a Dynamic Gammachirp Psychoacoustic Model and a DWT Multiresolution Analysis, *IJCSE,* Volume 2(4), pp. 1340–1354

Agbinya, J.I., 2012. Discrete Wavelet Transform Techniques in Speech Processing. *IEEE Tencon,*

Baumgarte, F., 2002. Improved Audio Coding using Psychoacoustic Model Based on a Cochlear Filter Bank, *IEEE Transactions on Speech and Audio Processing,* Volume 10(7),

China, S., Venkateswaralu, Sridhar, V., Subba R.R.A., Satya P.K., 2013. Audio Compression using Munich and Cambridge Filters for Audio Coding with Morlet Wavelet, *Global Journal of Computer Science and Technology Software and Data Engg.,* Volume 13(5), Davis Pan Motorola, 1995. A Tutorial on MPEG/Audio Compression, *IEEE Transaction on Signal Processing,*

Khalifa, O.O., Harding, S.H., Aisha-Hashin, H.A., 2005. Compression using Wavelet Transform, *Int. Journal: Signal Processing,* Volume 2(5),

Krimi, S., Auni, K., Ellouze, N., 2007. An Improved Psychoacoustic Model for Audio Coding Based on Wavelet Packet, *IEEE Int. Conf. SETIT 2007* Tumisia

Mourad Talbi, Chafik Barbarnoussi, Cherif Adnane, (2013), Speech Compression Based on Psychoacoustic Model and a General Approach for Filter Bank Design Using Optimization, *International Arab Conference on Information Technology (ACIT 2013).*

Painter, T., Spanias, A., 1997. A Review of Algorithms for Perceptual Coding of Digital Audio Signal, *IEEE Proceedings on Digital Signal Processing,*

Shao, Y., Chang, C.H., 2011. Bayesian Separation with Scarcity Promotion in Perceptual Wavelet Domain for Speech Enhancement and Hybrid Speech Recognition, *IEEE Transaction on System Man and Cybernetics,* Part A: System and Humans, Volume 41(2), pp. 284–293

Sheetal D. Gunjal, Rajeshee D. Raut, 2012. Advance Source Coding Techniques for Audio/Speech Signal: A Survey, *International Journal for Computer Technology and Applications,* Volume 3(4), pp. 1335–1342

Srinivasan, P., Jamieson, L.H., 1999. High Quality Audio Compression using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modeling, *IEEE Transaction On Signal Processing,*

Trina Adrian de Perez, Mini li, Hector McAllister, Norman D. Black, (2000), Noise Reduction and Loudness Compression in a Wavelet Modelling of The Auditory System, *IEEE Transaction on Signal Processing,*