IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 6, Dec-Jan, 2014
ISSN: 2320 - 8791
www.ijreat.org

Psychoacoustic Model-1 using Wavelet Tree Decomposition

Jagadeesh B Kanade[1], Dr.B.Sivakumar[2]

[1]**Department of ECE, VTU/HKBKCE**
Bangalore, Karnataka-560045 , India

[2]**Department of TCE, VTU/ DR.AIT**
Bangalore, Karnataka- ZIP/Zone, India

Abstract
Audio compression is the lossy compression technique of renovating sound indicates directly into a good successfully encoded stream that can afterwards be decoded to produce a close approximation on the original data.
In this work we are looking into implementation regarding psychoacoustics model intended for MPEG-1 making use of wavelet decomposition. For implementing this algorithm a design of psycho-acoustic model was developed. In this work psychoacoustic model implementation is based on Wavelet tree decomposition instead of conventional Fast Fourier transform (FFT).
Keywords: *Psycho-Acoustic, ATH, DWT, FFT.SMR, CB, MMT*

1. Introduction
Data compression describes strategy for decreasing data size without affecting the quality of the data, Audio data compression can be form of facts data compression, To get compressed Audio data compression procedures are developed along with carried out. This specific varies by easy way to many improve along with difficult in which normally takes awareness in the individual ear canal. In the act associated with sound recording data compression perceptual restriction associated with individual ear canal can be used. Limitation inside individual reading let to lose a few sound facts is just not identified within the unique sign. Hearing is usually a body organ associated with huge feeling, along with able to present a higher solution plus a excellent energetic choice of the sign. Should the blocking operations is just not appropriate subsequently it is going to bring about decrease of high quality in the aural sound. MPEG cross facts data compression (involves equally lossy along with lossless compression) process protocol achieves data compression by exploiting the perceptual restriction in the individual ear canal. Through the use of sound recording data compression algorithms you'll be able to receive compact digital camera representations associated with sound recording indicators for the purpose of successful indication without impairing the product quality in the receiving conclude. The principle reason for the sound recording coding should be to stand for the sound recording sign with a minimal amount of parts though obtaining see-through sign replica.
The absolute threshold of hearing (ATH) characterizes how much energy needed in a pure tone such that it might be detected by a listener in a noiseless environment. The absolute threshold is typically expressed regarding dB SPL. The frequency dependence of this threshold was quantified as soon as 1940, when Fletcher reported test results for a selection of listeners that were generated in a National Institutes of Health examine of typical American hearing acuity. The quiet (absolute) threshold is well approximated from the nonlinear function [1].

Absolute threshold of hearing is used to shape the coding distortion spectrum is the initial step toward perceptual coding. Absolute threshold is of limited value in the coding context. Finding threshold for spectrally complex quantization noise can be a modified version of the actual absolute threshold, with its shape determined by the stimuli present at any moment. Since stimuli are in general time-varying, the detection threshold is additionally a time-varying function in the input signal. Auditory masking is some sort of psycho acoustical phenomenon when a weak signal is masked in the presence of a tougher signal, the stronger signal is named masker and the signal that's masked by stronger signal is named maskee. Exploiting this phenomenon within perceptual audio compression is achieved in order that the original audio signal is treated being a masker for distortions launched by lossy data.

The psychoacoustic model utilized in the perceptual audio coder is founded on the Psychoacoustic Model-1 through the MPEG -1 audio. The MPEG-1 Audio Normal describes two different psychoacoustic designs, i. e. psychoacoustic models-1 and psychoacoustic models-2 the very first being computationally simpler and suited to coding at higher bit rates along with the second being more complex but also more reliable at lower bit rates As a result of complexity associated with the construction of

your psychoacoustic model-2. In this work we've got used psychoacoustic models-1 for our implementation.
Many MPEG coders applies a psycho-acoustic design for coding and uses the particular filter bank to approximate the frequency selectivity on the human auditory system. Figure1 (a) and Figure1 (b) shows a diagram on the structure of a generic perceptual sound coder. Figure 1(a) shows the structure on the encoder, which has three main stages and also a fourth is bit stream formatting stage and Figure 1(b) shows the decoder, which has three stages. Decoder operates around the encoded input audio signal and outputs the encoded bitstream plus the bitstream and reconstructs the unique signal. The three stages inside decoder, as a result, usually are reverse operations of encoder. Three stages inside encoder. Namely, the signal examination, Quantization and encoding, and bitstream formatting stages on the encoder correspond to the transmission synthesis, de-quantization and decoding, and bitstream extraction stages of the particular decoder, respectively. The extra stage in the encoder is

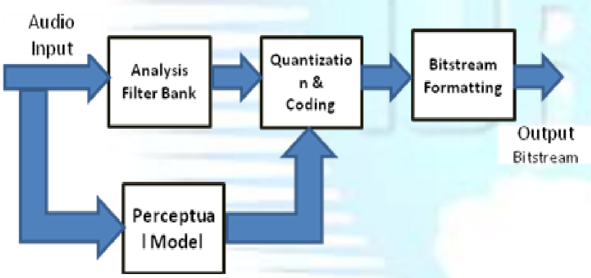

Fig1 (a) Encoder

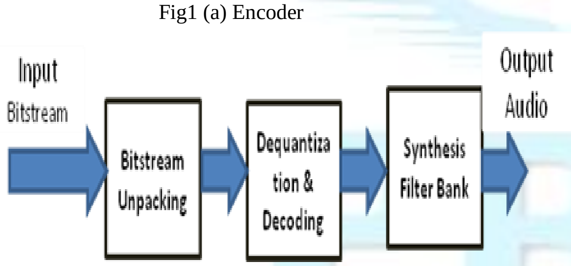

Fig1 (b) Decoder

the psychoacoustic model, which is not required in the decoder since the information is implicitly encoded as side-information. This means that perceptual coders are asymmetrical in that the encoder has a greater computational requirement than the decoder, which actually can be desirable in certain applications where one \server" encodes the signal for many clients".
The discrete wavelet transform can conveniently decompose the signal into an auditory critical band-like partition [2][11].Signal decomposition into critical bands resulting from Wavelet analysis needs to satisfy the spectral resolution requirements of the human auditory system.

2. Wavelet Tree Decomposition

Wavelet tree decomposition [4] provides a solution that makes possible for a finer an adjustable resolution of frequencies at high frequencies. This makes adaptation to particular signals [5]. Psychoacoustic model achieves an improved decomposition of the signal into critical bands(CB) using the discrete wavelet decomposition transform (DWT).This results in a spectral partition which approximates the critical band distribution much closer than before. Furthermore, the

Masking thresholds are computed entirely in the Wavelet domain To get approximation of critical bands using Wavelets analysis should meet the spectral resolution requirements. The wavelet basis also plays important role to satisfy temporal resolution of the signal. The typical range is less than 10ms at high frequencies to 100ms [6] at low frequincies.Individual masking threshold will be calculated for each component then deduction of tonal and nontonal components from the spectrum that has resulted from wavelet decomposition will give the global masking threshold[3]. The continuous wavelet transform (CWT)of signal x relative to the basic wavelet is given by:

$$W_{\psi}x(a,b) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{+\infty} x(t)\psi^{*}\left(\frac{t-b}{a}\right) \qquad (1)$$

Where a,b (a,b∈ $\mathbb{R}$, a≠0) are respectively the translation and scale parameters. Furthermore, ψ(t-b/a) represents the wavelet basis functions that are derived from a single mother wavelet function, ψ(t), through dilations a and translations b. The wavelet basis functions represent an Orthonormal basis to the space of $L^2(R)$ such that,

$L^2(R)$ = span { ϕ ab (t); a ∈ R+, b ∈ R}

If the basic wavelet satisfies the admissibility condition, then the wavelet reconstruction formula is:

$$x(t) = \iint_{R} W_{\psi}x(a,b)\psi_{a,b}(t)\frac{dadb}{a^2} \qquad (2)$$

The standard DWT involves a dyadic tree structure in which the low-channel side is successively split down to a certain depth. Wavelet decomposition is a wavelet transform in which the signal is passed through more number of filters .The detail coefficients will be obtained from the right-leaf node of each level and the approximation coefficients will be obtained from the left-leaf node at the lowest level . Fig.2 illustrates DWT where the nodes represent the wavelet coefficients (at various decomposition levels).

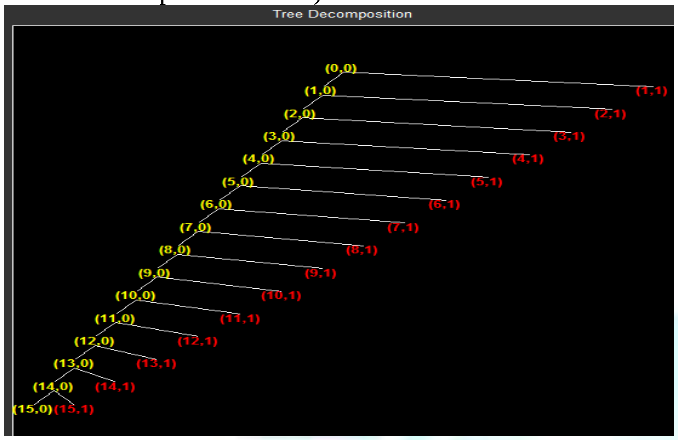

Fig2 .Decomposition tree structure for DWT

Wavelet decomposition with depth one splits the signal into high pass and low pass bands. With depth two will splits the low pass spectrum from depth one. Each stage wavelet decomposition splits low pass spectrum from previous stage, this yields an octave band pass filter bank wherein sampling rate of each subband is proportional to its bandwidth.Wavet analysis is efficient because of portions of the frequency towards the low frequency the psychoacoustic model is based on many studies of human perception. Studies have proven that the average human does not able to hear all frequencies as same.

While choosing specific wavelet decomposition the author have considered some restrictions to create orthogonal translates and dilates of the wavelet (the same number of coefficients than the scaling functions), and to ensure regularity (fast decay of coefficients controlled by choosing wavelets with large number of vanishing moments).

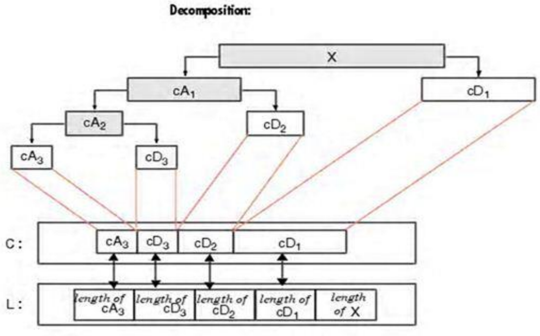

Figure 3: Wavelet decomposition

In this work author used orthogonal family of wavelets with name “daubechies”. There are so many types of “daubechies” wavelet. This work used “db10” figure3 shows the wavelet decomposition. Here “C” denotes the coefficients in the various decomposed branches of the tree and “L” denotes the number of the coefficients in the corresponding nodes of tree

3. Proposed design of psychoacoustic model-1 with Wavelet Decomposition

Fig 4 shows the implementation of psychoacoustic model with the wavelet decomposition.

Figure 4: Design of psychoacoustic model with wavelet Decomposition

The signal spectrum will be obtained by applying wavelet decomposition and whose connections are selected in such a way that sub bands correspond to the best possible one to the critical bands.

The implementation steps as follows,

Step 1: Framing

The “.wav” file is actually an uncompressed audio signal, divide the audio signal into different frames each frame of size 2048, Sampling frequency is 44100

Step 2: Wavelet Decomposition

Apply wavelet decomposition to each frame the wavelet used here is daubechies family wavelet specifically db10 with depth level 15 such that it will result in 30 subands that will replicate the critical bands in human auditory system.

Step 3: Local maxima

After wavelet decomposition, we are concerned with finding the local maxima among the various coefficients in

$$X(i) > X(i \pm 1)$$

a particular frame. The formula for finding the local maxima is [1].

Step 4: Elimination of frequency components masked by Absolute Threshold of Hearing (ATH)

The absolute threshold of hearing gives the information about the amount of energy needed in a pure tone such that it can be detected by a healthy listener in noiseless environment [1] .The local maxima calculated in sstep3 helps in determining those coefficients of an audio signal which are irrelevant for the healthy listener. Using ATH the aim is to retain only those coefficients which lies above the curve as in the figure4 .The quiet (absolute) threshold is well approximated [17] by the nonlinear function.

$$T_q(f)=3.64(f/1000)^{-0.8}-6.5e^{-0.6(f/1000-3.3)^2}+10^{-3}(f/1000)^4 \text{ (dB SPL)}$$

Which is representative of a young listener with acute hearing [1] as shown in the fig 5

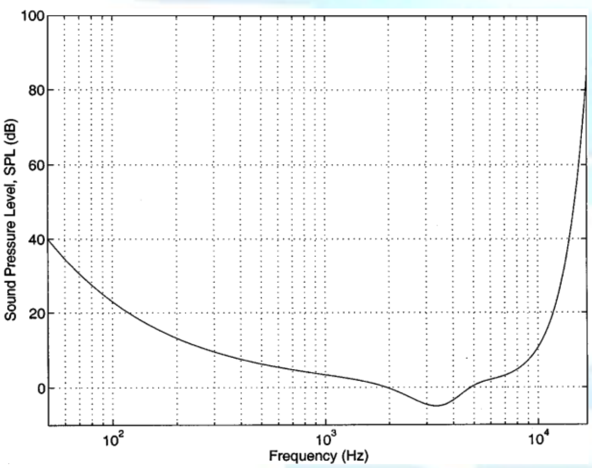

Figure5.Absolute threshold of hearing in quiet[1]

Step 5: Localization of tonal and non tonal components

After all the estimations and normalizations done next step is to identifying the tonal and non tonal masking components from local maxima that has obtained. If frequency bin (component) exceeds neighboring components within a bark distance by at least 7 db then it will treated as "tonal" otherwise it will be considered as "non tonal". Notions of critical bandwidth and simultaneous masking in the audio coding context give rise to some convenient terminology illustrated in Fig. 6. More detail is provided in [1].

Step 6: Individual masking threshold

Every tonal and non tonal component has its effect on the neighboring coefficients. The effect can be analyzed as its individual masking threshold. This can be analyzed by the spreading function, The spreading function[1] is given by

$$SF_{db}(x)=15.81+7.5(x+0.474)-17.5\sqrt{1+(x+0.474)^2} \text{ db}$$

Where unit of x is Bark and $SF_{db}(x)$ is expressed in dB.

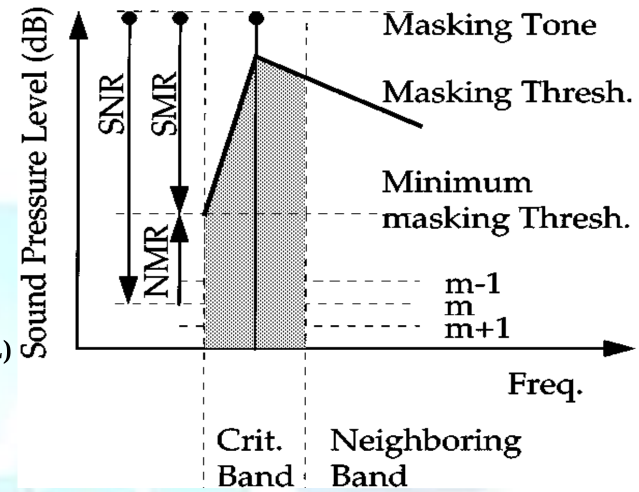

Figure 6. Schematic representation of simultaneous Masking [1].

Critical band analysis is done and the spread of masking has been accounted for, masking thresholds in perceptual coders are often established by the decibel relations[1].

$$TH_N=E_T-14.5-B \quad \text{and} \quad TH_T=E_N-K$$

Where TH_N and TH_T noise and tone masking thresholds, respectively figure 6 shows the schematic representation of simultaneous masking.

Step 7: Global masking threshold

Global masking threshold is additive in nature. It is calculated by adding the effect of all the thresholds calculated above. An adjustable resolution of frequencies at high frequencies can be achieved through wavelet decomposition. It gives a structure that allows adaptation to particular signal

Step 8: Minimum masking threshold (MMT)

The minimum of the global masking threshold for each sub band is calculated and is referred as MMT.

4. Experimental Results

In this work we have implemented Psychoacoustic model-1 based on the standard MPEG-1 audio with wavelet decomposition instead of FFT. Figures7 to figure10 captures the results of various stages of the algorithm. Figure 11 shows the plot of minimum masking threshold for the various subands this plot is the result of Psychoacoustic implementation using wavelet decomposition. Figure12 shows the implementation using FFT.From the figure11 and Figure12 it is evident that minimum masking threshold that is resulted from wavelet decompostion is better estimate for subbands compare to FFT method.

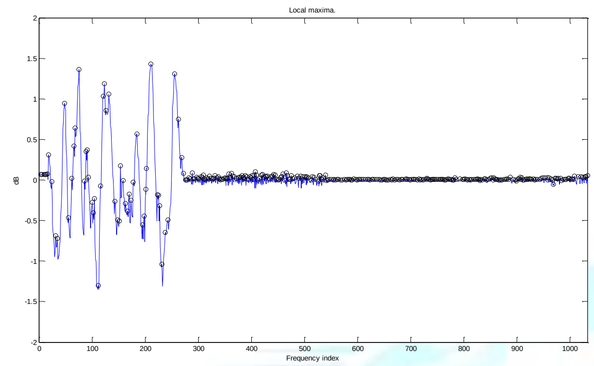

Figure7. Local Maxima

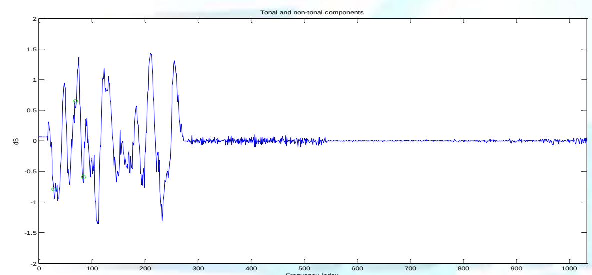

Figure8. Tonal and Non Tonal Components

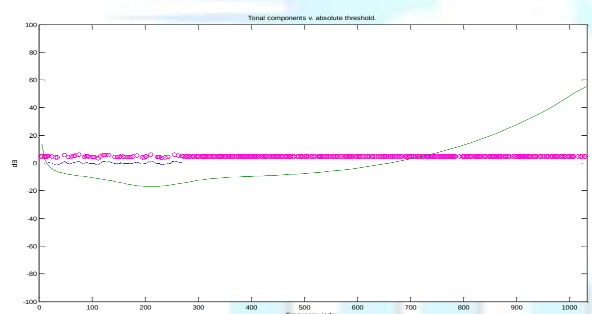

Figure9. Tonal Components with ATH

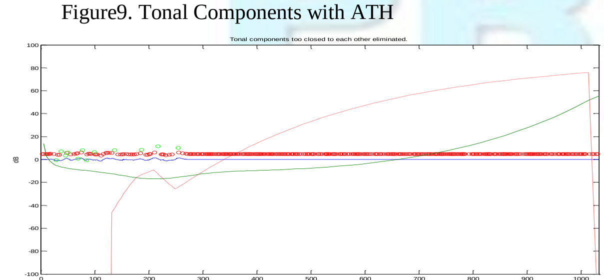

Figure10. Tonal Components closed to each other is eliminated

5. Conclusions and Future work

The psychoacoustic model-1 based on wavelet decomposition takes account of the critical bands and takes an account of the masking phenomenon. The specialty of the proposed model is that it gives an analysis by wavelet decomposition on the frequency bands that gives the closer approximation of the critical bands of the ear. The result of implementation in this paper can be evaluated using signal to mask ratio, later this can be integrated with the other blocks of the MPEG audio codecs to get overall compression ratio (CR).

References

[1] T. Painter and A. Spanias,"Perceptual coding of digital audio," Proc. IEEE Trans.Sig, vol. 88, Apr. 2000, pp. 51-513

[2] D.Sinha and A.Tewfik, Low Bit Rate Transparent Audio Compression using adapted wavelets, IEEE Trans.Sig.Proc, pp.3463-3479, Dec 1993.

[3] Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1, 5 Mbit/s-Part3: Audio, 1999.

[4] M. V. Wickerhauser, "Adapted wavelet analysis from theory to software," Wellesley, assachusetts, 1994.

[5] C.S.Burrus, R.A. Gopinath and H.Guo, "Introduction to wavelets and wavelets transforms: A Primer, "Prentice Hall, 1998.

[6] M. Bosi and R. E. Goldberg, Introduction to Digital Audio Coding and Standards, Kluwer Academic Publishers, New York, NY, USA, 2003.

[7] M. R. Zurera, F. L. Ferreras, M. P. J. Amores, S.M. Basc´ on, and N. R. Reyes, "A new algorithm for translating psycho-acoustic information to the wavelet domain," Signal Processing, vol. 81, no. 3, pp. 519–531, 2001.

[8] B. Lincoln, "An experimental high fidelity perceptual audio coder," Project in MUS420 Win97, March 1998.

[9] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," IEEE Journal on Selected Areas in Communications,vol. 6, no. 2, pp. 314–323, 1988.

[10] Irino, T., Patterson, R. D., A time-domain, level-dependent auditory filter: The gamma chirp, JASA, Vol. 101, No. 1,pp. 412-419, January 1997.

[11] B. Carnero and A. Drygajlo, "Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms," IEEE Transactions on Signal Processing,vol. 47, no. 6, pp. 1622–1635, 1999.

[12] M. R. Zurera, F. L. Ferreras, M. P. J. Amores, S.M. Basc´ on, and N. R. Reyes, "A new algorithm for translating psycho-acoustic information to the wavelet domain," Signal Processing, vol. 81,no. 3, pp. 519–531, 2001

[13] Strang, G., Nguyen, T., 1996. "Wavelets and Filter Banks" Wellesley-Cambridge Press, Wellesley, Massachusetts.

[14]Chui,C.K. an Wang,J.Z.(1992), "On compactly supported spline wavelets and a duality principle", rans. Amer. Mathematic Soc. 330 903-915.

[15]Meyer, Y. (1989), "Orthonormal Wavelets ", in wavelets, Proc. Int. Conf. France, Springer, Berlin.

[16] Ogden RT. (1997), "Essential Wavelets for Statistical Applications and Data Analysis".Birkhauser, Boston.

[17] E. Terhardt, "Calculating virtual pitch," Hearing Res., vol. 1, pp.155–182, 1979.