Compression of Audio Signal Using Gammachirp Wavelet

B. Sree Deepthi
M.Tech Student, Department of ECE,
K.L.University, Vaddeswaram, Andhra Pradesh, India
deepthiece1@gmail.com

Abstract

The main goal of the proposed algorithm in this project is to compress high quality audio signal and maintaining transparent quality at low bit rates. Most psychoacoustic models for coding applications use a uniform spectral decomposition to approximate the frequency selectivity of the human auditory system. However the equal filter properties of the uniform subbands do not match the non uniform characteristics of the cochlear filters. For implementing this algorithm a design of psycho-acoustic model was developed following the model used in the standard MPEG-1 audio layer 3.Wavelet transform can be efficiently used to represent human speech with less number of bits that can be decoded to produce a close approximation of the original speech signal. This architecture is based on appropriate wavelet packet decomposition instead of short time Fourier transformation. This study evaluates the performance of Morlet Munich coder and Gamma chirp coder.

Keywords- *Psychoacoustic model, FFT, Wavelet packet decomposition, Munich and Gamma chirp models*

INTRODUCTION:

The techniques of coding make possible to reach high compression ratios while preserving a good subjective quality. Many methods of compression were developed; from the simplest one that consists to make a banal under sampling, to the most advanced that takes account of the sensitivity of the human ear. For these last methods, the ear presents in effect, some limit hearing that let to eliminate some sound information not perceived in the original signal.

However, the ear is an organ of large sensibility, presenting a high resolution and a great dynamic of the signal: a bad filtering can lead to a loss of an aural quality

The MPEG/Audio is a standard for both transmitting and recording compressed ratio. The MPEG algorithm achieves compression by exploiting the perceptual limitation of the human ear. Audio compression algorithms are used to obtain compact digital representations of high-fidelity audio signals for the purpose of efficient transmission. The main objective in audio coding is to represent the signal with a minimum number of bits while achieving transparent signal reproduction. The majority of MPEG1coders apply a psycho-acoustic model for coding applications using filter bank to approximate the frequency selectivity of the human auditory system.

In this paper, a new design for modeling auditory masking is based on wavelet packet decomposition. Wavelet unlike FFT, whose basic functions are sinusoids, is based on small waves, called wavelets, of varying frequency and limited duration. It's most important characteristics are conceived to analyze temporal and spectral properties of non-stationary signals such as audio. A new architecture of a psycho-acoustic model based on wavelets decomposition can be applied to the audio coders in sub-bands approximating the critical bands. Our study is made up of four parts. The first part will focus on the proposed psychoacoustic model. The second part aims at presenting the two models (Munich and Gammachirp).

The third part deals with a comparison between the compressions ratios of the Proposed coders (Munich and Gammachirp).

THE PROPOSED PSYCHO-ACOUSTIC MODEL:

Psycho-acoustic studies show that each frequency takes place in the cochlea, along the basilar membrane. Therefore, the cochlea can be viewed from a signal-processing perspective as a bank of highly overlapping band-pass filters. The efficient algorithm of the new design is based on some consecutive steps as shown in figure1.

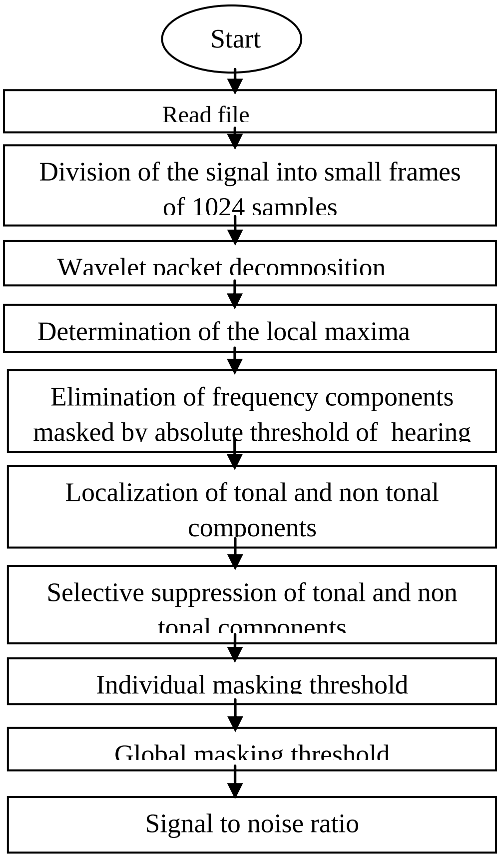

Fig.1. The new psycho-acoustic model using wavelet packet decomposition

Wavelet packet decomposition on 1024 points of the wave signal is applied. An approximation of the critical bands is adopted using 26 sub-bands. It has approximately the same architecture used by Tewfik and Sinha .

Then, we determine tonal and non tonal components. This step begins with the determination of the local maxima, followed by extracting the tonal components (sinusoidal) and non tonal components (noise) in every bandwidth of critical bands. The selective suppression of tonal and non tonal components of masking is a procedure used to reduce the number of maskers taken into account for the calculation of the global masking threshold. The remaining tonal and non tonal components are those which are above the hearing absolute threshold. Individual masking threshold takes into account the masking threshold for each remaining component.

Lastly, global masking threshold is calculated by the whole of tonal and non tonal components which are deduced from the spectrum of the wavelet packet decomposition. The wavelet packet transform constitutes a solution that permits a finer an adjustable resolution of frequencies at high frequencies and gives a rich structure that allows adaptation to particular signals or signals classes

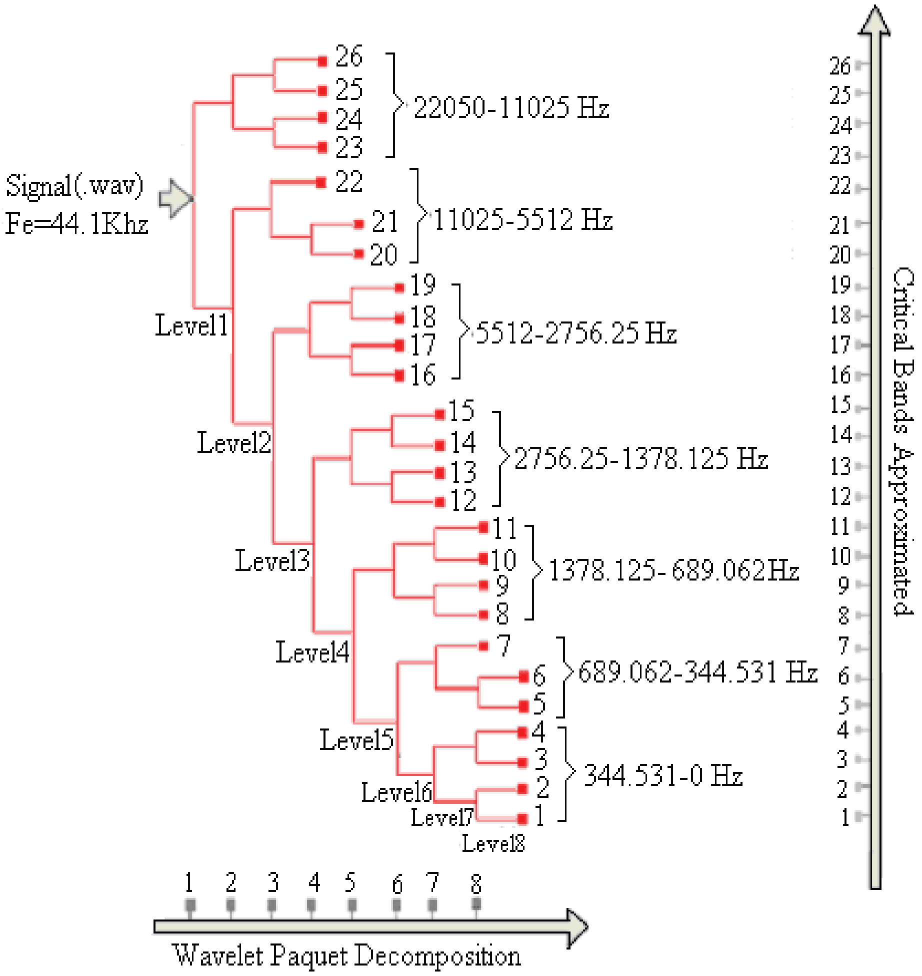

Fig. 2. The proposed wavelet packet decomposition

Auditory masking is a well-known psycho acoustical phenomenon in which a weak signal (maskee) becomes inaudible (masked) in the presence of a stronger masker signal. Exploiting this phenomenon in perceptual audio coding is achieved so that the original audio signal is treated as a masker for distortions introduced by lossy data compression. The gammachirp filter underwent a good success in psychoacoustic research. Indeed, it fulfils some important requirements and complexities of the cochlear filter

THE MUNICH FILTER BANK:

Morlet wavelet is chosen as the mother wavelet to decompose the audible spectrum for the cochlear filters bank. We use the Morlet wavelet based on the fact that it has good properties in joint time frequency localization and has a well defined impulse response. The Morlet wavelet transform has proved to be beneficial for many types of wave signal processing and has shown a good performance in tasks like audio coding. The Morlet wavelet shape is not simply analytical, but also very resolute in time and frequency, regular, locally periodic and with non compact support. It is formed by complex values of the shape of a complex sinus modulated by a Gaussian envelope. It is characterized by narrow frequency response, which offers a higher spectral resolution.

The Morlet wavelet is defined as:

$$\psi(t) = \frac{1}{\sqrt{2\pi\lambda}} e^{\frac{-t^2}{2\lambda^2}} e^{-i2\pi f_o t}$$

Considering a constant surface of elementary paving in time and frequency domain, we can prove that

$$\lambda_t * \lambda_f = \frac{1}{2\pi}$$

The result of a mathematical demonstration can be
summarized as

$$\lambda = \lambda_t = \frac{\sqrt{0.7}}{B\pi}$$

If $\gamma(f)$ is a Fourier transform of $\psi(t)$:

$$\gamma(f) = \sqrt{\lambda} e^{\frac{-1}{2}(2\pi f - 2\pi\lambda f_o)^2}$$

Then the following condition must be satisfied

$$c_g = \int_{\infty}^{\infty} \frac{|\psi(\omega)|^2}{\omega} d\omega < \infty$$

Finally this wavelet is defined as

$$\psi(t) = \sqrt{\frac{\pi}{1.4}}(B).e^{\frac{-(B_M)^2 t^2 \pi^2}{1.4}} e^{-i2\pi f_o t}$$

Where $e^{\frac{-B^2 t^2 \pi^2}{1.4}}$ is the Gaussian envelope, f_o is the center frequency of the mother wavelet and "B" is the equivalent rectangular bandwidth of the cochlea. The peripheral auditory system is equivalent to a band-pass filter bank whose respective bands would have the critical bandwidth. To approximate these critical bands, we adopted in our work two models called Munich and Cambridge. In fact, there is a similarity between the widths of bands of these models and those of the ear.

This model is defined by the following function

$$B_M = 25 + 75(1.4F^2 + 1)^{0.69}$$

Which F is the central frequency of the band (in KHz).

Consequently, the Morlet Munich wavelet is given by:

$$\psi_M(t) = \sqrt{\frac{\pi}{1.4}}(B_M).e^{\frac{-(B_M)^2 t^2 \pi^2}{1.4}} e^{-i2\pi f_0 t}$$

The spectral analysis of Munich models give the following results:

Fig3. Morlet Munich filter bank

THE GAMMACHIRP FILTER BANK:

Gammachirp filter is popular for auditory speech analysis. This function was introduced by Irino and Patterson.

It has the following classical form

$$\psi(t) = A.t^{n-1}.e^{-2\pi.b.ERB(f_o).t}.e^{-i.(2\pi.f_o.t + c.\ln(t) + \varphi)}$$

Where time t > 0, "A" is the amplitude, "n" and "b" are
parameters defining the envelope of the gamma distribution, and f_o is the asymptotic frequency. "c" is a parameter for the frequency modulation or the chirp rate, φ is the initial phase, ln(t) is a natural logarithm of time, and $ERB(f_o)$ is the equivalent rectangular bandwidth of an auditory filter at f_o and is defined by the following expression

$$ERB(f_o) = 0.108.f_o + 24.7$$

THE GAMMACHIRP FILTER AS A WAVELET

This wavelet has the following properties: It is with no compact support, non symmetric, non orthogonal and not even presenting a scale function. For this family of wavelets.

The modulation frequencies are $f_m = f_o.s_o^{-m}$ and the bandwidths are $\beta_m = \beta_0.s_0^{-m}$

Where $\beta == b.ERB(f_o)$ and s_o is the dilation parameter

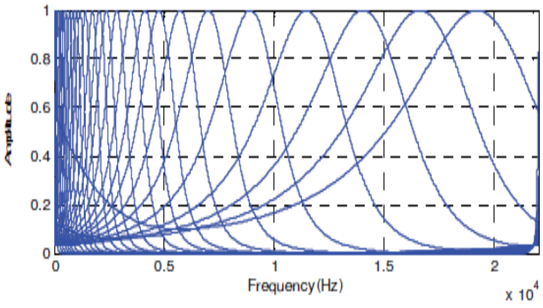

Fig4. Gammachirp filter bank

The choice of the gammachirp filter is based on two reasons:

1. First reason is that the gammachirp filter has a well defined impulse response, unlike the conventional roex auditory filter, and so it is an excellent candidate for an asymmetric, level-dependent auditory filter bank in time domain models of auditory processing.
2. Second reason is that this filter was derived by Irino as a theoretically optimal auditory filter that can achieve minimum uncertainty in a joint time-scale representation.

Table 1: sound compression ratio obtained by gamma chirp Wavelet packet

Sounds wav	Duration (s)	Capacity (ko)	96	112	128	160
Rock	10	948	8.323	7.091	6.167	4.873
Classic	12	1003	8.316	7.084	6.160	4.866
Jazz	9	851	8.310	7.078	6.154	4.860
Voice	9	801	8.305	7.073	6.149	4.855
Opera	11	521	8.301	7.069	6.145	4.851

Evaluation of sound compression ratio:

In order to evaluate Munich and Gamma chirp coders, we used for various bitrates between 64 Kbits/s and 160 Kbits/s some types of sound such as Soul, Slow, Rock, Arabic music and voice

Table 1 indicates the type of the sound files used for the test, their duration, their capacity as well as the compression ratio,The evaluation is based on the compression ratio defined as

$CR = \frac{length(wav\ file)}{length(mp3\ file)}$ for various flows (Kbit/s).

The types of the sounds chosen for the tests try to cover some difficult aspects to code such as percussions and the pure sounds.

• Rock music: this type of sound contains the electric guitar, it is not dense.

• Classical music: this type of sound contains violin like some percussions.

• Jazz music: this type of sound contains piano.

• Voice: a recorded sentence made by the first author.

The recording was made in a calm medium.

• Opera: this type of sound contains the voice of Maria Callas in addition to the violin and the piano.

The highest compression ratio is given for the flow of 96Kbit/s and the weakest one is given for the flow of 160Kbit/s. however, higher the compression ratio is, less its quality become intelligible. In the continuation, we will clarify the compromise done between
compression ratio and sound quality. Then, we have proceeded to a comparison of the sound compression ratio of the coder based on FFT analysis and the coder based on wavelet packet analysis with gamma chirp function.

The gammachirp wavelet packet takes account of the critical bands and takes account of the masking phenomenon, on the other hand the FFT does not point on the critical bands.

This new psycho-acoustic model has been implemented in reference coder based on the standard MPEG-1 audio layer 3. The spectrum of the original and compressed signal using the two models (Munich and Gammachirp) are shown in figure 5.

Those figures show that high frequencies are canceled for compression since their amplitudes are low and converge to 0 dB. Those amplitudes will be not listen since their values are below to the absolute threshold of hearing.We remark that the Gammachirp model is quiet better than the Munich one. In fact, for Munich model, high frequencies are canceled from 15000 Hz, as for the Gammachirp one they are canceled from 12500 Hz.

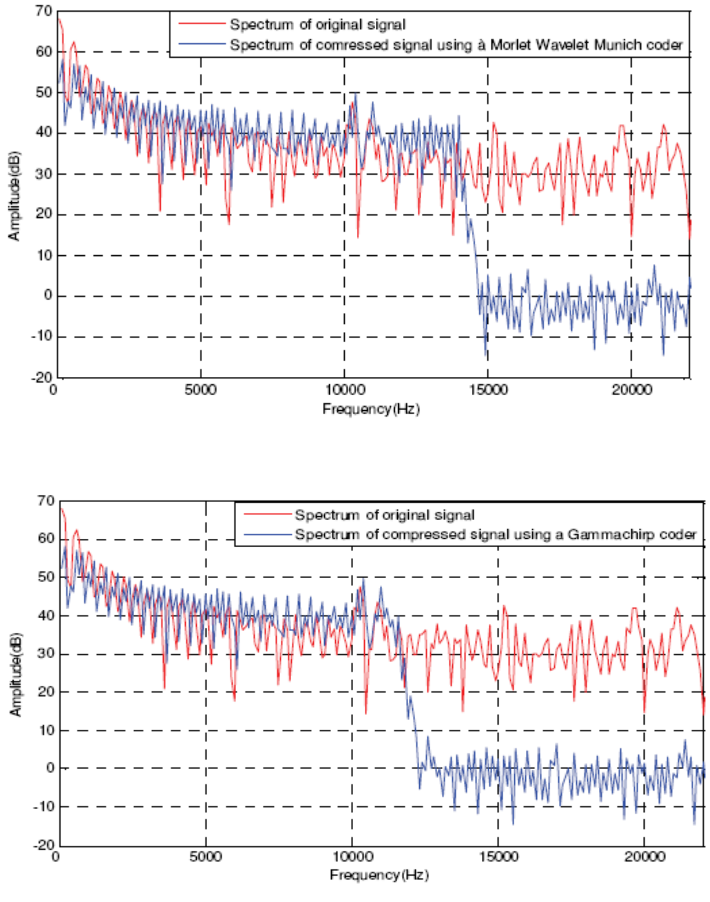

Fig.5. Spectrum of original signal and compressed signal

Using the two models: the Munich model (the first) and the Gammachirp model (the second)

CONCLUSION

This paper studies audio compression using a Munich and Gammachirp transforms. Our goal is to identify the best model that can be applied to the audio coders in sub-bands approximating the critical bands and finally to improve the psycho-acoustic model. This study shows that the Gammachirp offers a good performance. In fact, it gives the best compression ratio and sound quality in comparison to the FFT and Munich coders. A high selectivity was noticed and can lead to some interesting perspectives on audio coding using this type of psycho-acoustic model.

REFERENCES

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

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

Alex Park, Using the gamma chirp filter for auditory analysis of speech May 14, 2003. 18.327: Wavelets and Filter banks

G. Hernandez, M. Mendoza, B. Reusch, and L. Salinas, "*Shiftability and filter bank design using Morlet wavelet*," Computer Science Society, pp.141-148, Nov.2004

I. Daubechies, "*Ten lectures on wavelet*," SIAM, Philadelphia, PA, 1992

J. O Smith III and J. S Abel, "*Bark and ERB Bilinear transforms* ," IEEE Transactions on speech and Audio Processing, 1999.

T. Painter and A. Spanias, "Perceptual coding of digital audio," Proc. IEEE Trans. Sig , vol. 88, Apr. 2000, pp. 451-513

C. K. Chui, "An Introduction to Wavelets," Academic Press, 1992