Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 13 Issue 5 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Audio Compression using Munich and Cambridge Filters for Audio Coding with Morlet Wavelet

By S. China Venkateswarlu, V. Sridhar, A. Subba Rami Reddy & K. Satya Prasad

Holy Mary Institute of Technology & Science

Abstract - The main aim of work is to develop morlet wavelet using Munich and Cambridge filters, for audio compression and most psycho-acoustic models for coding applications use a uniform -equal bandwidth, spectral decomposition for compression. In this paper we present a new design of a psycho-acoustic model for audio coding following the model used in the standard MPEG-1 audio layer 3. This architecture is based on appropriate wavelet packet decomposition instead of a short term Fourier transformation. To fulfill this aim, the following objectives are carried out: Approximate the frequency selectivity of the human auditory system. However, the equal filter properties of the uniform sub- bands do not match the non uniform characteristics of cochlear filters and reduce the precision of psycho-acoustic modeling. This architecture is based on appropriate wavelet packet decomposition instead of a short term Fourier transformation. In this paper Morlet Munich coder shows best performance. 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.

Keywords : audio compression, acoustic, human voice, morlet wavelet, munich, psycho-acoustic, music, wavelet transform.

GJCST-C Classification : H.5.5

AUDIO COMPRESSION USING MUNICH AND CAMBRIDGE FILTERS FOR AUDIO CODING WITH MORLET WAVELET

Strictly as per the compliance and regulations of:

Audio Compression using Munich and Cambridge Filters for Audio Coding with Morlet Wavelet

S. China Venkateswarlu [α], V. Sridhar [σ], A. Subba Rami Reddy [ρ] & K. Satya Prasad [ω]

Abstract - The main aim of work is to develop morlet wavelet using Munich and Cambridge filters, for audio compression and most psycho-acoustic models for coding applications use a uniform -equal bandwidth, spectral decomposition for compression. In this paper we present a new design of a psycho-acoustic model for audio coding following the model used in the standard MPEG-1 audio layer 3. This architecture is based on appropriate wavelet packet decomposition instead of a short term Fourier transformation. To fulfill this aim, the following objectives are carried out: Approximate the frequency selectivity of the human auditory system. However, the equal filter properties of the uniform sub- bands do not match the non uniform characteristics of cochlear filters and reduce the precision of psycho-acoustic modeling. This architecture is based on appropriate wavelet packet decomposition instead of a short term Fourier transformation. In this paper Morlet Munich coder shows best performance. 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. The ear is an organ of large sensibility, presenting a high resolution and a great dynamic of the signal. The ear is more sensitive to low frequencies than to high ones and our hearing threshold is very intense in the high frequency regions. The study of its characteristics makes it possible to exploit the intrinsic properties of the ear in order to carry out systems of compression with loss of information. Most MP3 coders use the Fourier transform whose fundamental roles are limited only to the spectral properties of the signal. In this paper, a new design for modeling auditory masking is based on wavelet packet decomposition.

Keywords : audio compression, acoustic, human voice, morlet wavelet, munich, psycho-acoustic, music, wavelet transform.

Author α : Research Scholar JNTUK-Kakinada.
E-mail : cvenkateswarlus@gmail.com
Author σ : Asst. Professor ECE, VJIT, Principal SKIT, Tirupati.
Author ρ : Rector JNTUK-Kakinada.

I. Introduction

Audio compression is a form of data compression designed to reduce the transmission bandwidth requirement of digital audio_streams and the storage size of audio files. Audio compression algorithms are implemented in computer software as audio codecs. Generic data compression algorithms perform poorly with audio data, seldom reducing data size much below 87% from the original and are not designed for use in real time applications. Consequently, specifically optimized audio lossless and lossy algorithms have been created. Lossy algorithms provide greater compression rates and are used in mainstream consumer audio devices.In both lossy and lossless compression, information__redundancy is reduced, using methods such as coding, pattern recognition and linear prediction to reduce the amount of information used to represent the uncompressed data. The trade-off between slightly reduced audio quality and transmission or storage size is outweighed by the latter for most practical audio applications in which users may not perceive the loss in playback rendition quality. For example, one Compact Disc holds approximately one hour of uncompressed high fidelity music, less than 2 hours of music compressed losslessly, or 7 hours of music compressed in the MP3 format at medium bit rates.

a) Lossless Audio Compression

Lossless audio compression produces a representation of digital data that can be expanded to an exact digital duplicate of the original audio stream. This is in contrast to the irreversible changes upon playback from lossy compression techniques such as Vorbis and MP3. Compression ratios are similar to those for generic lossless data compression (around 50–60% of original size, and substantially less than for lossy compression, which typically yield 5–20% of original size.

b) Lossless Compressed Audio Formats

A lossless compressed format requires much more processing time than an uncompressed format but is more efficient in space usage. Uncompressed audio formats encode both sound and silence with the same number of bits per unit of time. Encoding an uncompressed minute of absolute silence produces a file of the same size as encoding an uncompressed minute of symphonic orchestra music. In a lossless compressed format, however, the music would occupy a marginally smaller file and the silence take up almost no space at all. Lossless compression formats (such as the most widespread FLAC, WavPack, Monkey's Audio, ALAC/Apple Lossless) provide a compression ratio of about 2:1. Development in lossless compression formats aims to reduce processing time while maintaining a good compression ratio.

II. Audio Signal Processing

Audio signal processing, sometimes referred to as audio processing, is the intentional alteration of auditory signals, or sound. As audio signals may be electronically represented in either digital or analog format, signal processing may occur in either domain. Analog processors operate directly on the electrical signal, while digital processors operate mathematically on the digital representation of that signal. Audio processing was necessary for early radio broadcasting-as there were many problems with studio to transmitter links. *Analog signals*-An analog representation is usually a continuous, non-discrete, electrical; a voltage level represents the air pressure waveform of the sound. *Digital signals*- A digital representation expresses the pressure wave-form as a sequence of symbols, usually binary numbers. This permits signal processing using digital circuits such as microprocessors and computers. Although such a conversion can be prone to loss, most modern audio systems use this approach as the techniques of digital signal processing are much more powerful and efficient than analog domain signal processing.

a) Sound Recording and Reproduction

Sound recording and reproduction is an electrical or mechanical inscription and re-creation of sound waves, such as spoken voice, singing, instrumental music, or sound effects. The two main classes of sound recording technology are analog recording and digital recording. Acoustic analog recording is achieved by a small microphone diaphragm that can detect changes in atmospheric pressure (acoustic sound waves) and record them as a graphic representation of the sound waves on a medium such as a phonograph (in which a stylus senses grooves on a record). In magnetic tape recording, the sound waves vibrate the microphone diaphragm and are converted into a varying electric current, which is then converted to a varying magnetic field by an electromagnet, which makes a representation of the sound as magnetized areas on a plastic tape with a magnetic coating on it. Analog sound reproduction is the reverse process, with a bigger loudspeaker diaphragm causing changes to atmospheric pressure to form acoustic sound waves. Electronically generated sound waves may also be recorded directly from devices such as an electric guitar pickup or a synthesizer, without the use of acoustics in the recording process other than the need for musicians to hear how well they are playing during recording sessions.

Digital recording and reproduction converts the analog sound signal picked up by the microphone to a digital form by a process of digitization, allowing it to be stored and transmitted by a wider variety of media. Digital recording stores audio as a series of binary numbers representing samples of the amplitude of the audio signal at equal time intervals, at a sample rate so fast that the human ear perceives the result as continuous sound. Digital recordings are considered higher quality than analog recordings not necessarily because they have higher fidelity (wider frequency response or dynamic range), but because the digital format can prevent much loss of quality found in analog recording due to noise and electromagnetic interference in playback, and mechanical deterioration or damage to the storage medium. A digital audio signal must be reconverted to analog form during playback before it is applied to a loudspeaker or earphones.

b) Speech Compression

Speech compression may mean different things: Speech encoding refers to compression for transmission or storage, possibly to an unintelligible state, with decompression used prior to playback. Time-compressed speech refers to voice compression for immediate playback, without any decompression (so that the final speech sounds faster to the listener). Wavelet unlike FFT, whose basic functions are sinusoids, is based on small waves, called wavelets, of varying frequency and limited duration. Its 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. Models of human auditory pathway are commonly used as a part of audio coding systems or algorithms for objective Evaluation of sound quality. In order to properly simulate the nonlinear, signal-dependent behavior of the cochlea, physiologically accurate models were developed, such as the Munich model which was concepted by Zwicker and as for the Cambridge one by Moore. Our study is made up of fourth parts. The first part will focus on the proposed psycho-acoustic model. The second part aims at presenting the two models (Munich and Cambridge). The third part deals with a comparison between the compressions ratios of the proposed coders (FFT and wavelet). Finally, a sound quality evaluation will be carried out to conclude this work. 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. The ear is an organ of large sensibility, presenting a high resolution and a great dynamic of the signal. The ear is more sensitive to low frequencies than to high ones and our hearing threshold is very intense in the high frequency regions. The study of its characteristics makes it possible to exploit the intrinsic properties of the ear in order to carry out systems of compression with loss of information. Most MP3 coders use the Fourier transform whose fundamental roles are limited only to the spectral properties of the signal. In this paper, a new design for modelling 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. Its 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. Models of human auditory pathway are commonly used as a part of audio coding systems or algorithms for objective Evaluation of sound quality. In order to properly simulate the nonlinear, signal-dependent behaviour of the cochlea, physiologically accurate models were developed, such as the Munich model which was concepted by Zwicker and as for the Cambridge one by Moore. Our study is made up of fourth parts. The first part will focus on the proposed psycho-acoustic model. The second part aims at presenting the two models (Munich and Cambridge). The third part deals with a comparison between the compression ratios of the proposed coders (FFT and wavelet). Finally, a sound quality evaluation will be carried out to conclude this work.

III. Wavelet Transform

Given a real, mother wavelet function $\psi(t) \in L^2(R)$, which satisfies the admissibility condition,

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad (3.1)$$

Where $\Psi(\omega)$ is the Fourier transform of $\psi(t)$, the scaling parameter p and the shifting parameter q are introduced to construct a family of basis functions or wavelets $\{\psi_{p,q}(t)\},\ p,q \in R,\ p \neq 0$, where

$$\psi_{p,q}(t) = \frac{1}{\sqrt{|p|}} \psi\left(\frac{t-q}{p}\right) \quad (3.2)$$

If $\Psi(\omega)$ has sufficient decay, the admissibility condition reduces to the constraint that $\Psi(0) = 0$ or

$$\int_{-\infty}^{\infty} \psi(t) dt = \Psi(0) = 0 \quad (3.3)$$

Because the Fourier transform is zero at the origin, and the spectrum decays at high frequencies, the wavelet behaves as a bandpass filter. The continuous wavelet transform (CWT) of a function or a signal $x(t) \in L^2(R)$ is then defined as

$$X_{CWT}(p,q) = \int_{-\infty}^{\infty} x(t) \psi_{p,q}(t) dt \quad (3.4)$$

and the inverse wavelet transform is given by

$$x(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X_{CWT}(p,q) \psi_{p,q}(t) \frac{dpdq}{p^2} \quad (3.5)$$

Since $\{\psi_{p,q}(t)\}$ is a redundant basis set, it implies that, by discretizing the scaling and shifting parameters as follows:

$$p = 2^k, q = 2^k l,\ k,l \in Z \quad (3.6)$$

an orthonormal basis $\{\psi_{k,l}(t)\}$ is obtained,

$$\psi_{k,l}(x) = 2^{-k/2} \psi(2^{-k} x - l),\ k,l \in Z \quad (3.7)$$

which comply with the condition,

$$\int_{-\infty}^{\infty} \psi_{k,l}(t) \psi_{k',l'}(t) dt = \delta(k-k') \delta(l-l'), \forall k,k',l,l' \in Z \quad (3.8)$$

The wavelet transform is an appropriate tool for multiresolution analysis (MRA) [47]. In MRA, a *scaling function* $\phi(x)$ and the associated mother wavelet function $\psi(x)$, which are needed in the construction of a *complete basis* (see below), must satisfy the two-scale difference equations:

$$\phi(x) = \sqrt{2}\sum_n h(n)\phi(2x-n)$$
$$\psi(x) = \sqrt{2}\sum_n g(n)\phi(2x-n) \quad (3.9)$$

Where the coefficients $h(n)$ and $g(n)$ satisfy the following:

$$g(n) = (-1)^n h(1-n) \quad (3.10)$$

Similar to the construction of $\{\psi_{k,l}(t)\}$, a family of orthonormal basis $\{\phi_{k,l}(x)\}$ can be obtained through translation and dilation of the kernel $\phi(x)$.

$$\phi_{k,l}(x) = 2^{-k/2}\phi(2^{-k}x-l), k,l \in Z \quad (3.11)$$

A series of nested subspaces V_j, which is spanned by the orthonormal basis $\{\phi_{j,k}(x)\}, k \in Z$, forms a multiresolution space. The subspace W_j, which is spanned by the orthonormal basis $\{\psi_{j,k}(x)\}, k \in Z$, is the complementary space of V_j in the subspace V_{j-1}.

$$V_{j-1} = V_j \oplus W_j \quad (3.12)$$

The subspaces V_j and W_j are called the approximate and residue spaces respectively at resolution j.

For the discrete case, the coefficients $h(n)$ and $g(n)$ also play an important role because the continuous forms of $\phi(x)$ and $\psi(x)$ can be neglected, and the coefficients can be directly applied to the discrete signal using the following iterations:

$$a_{j+1}(n) = \sum_k a_j(k)h(k-2n)$$
$$r_{j+1}(n) = \sum_k a_j(k)g(k-2n) \quad (3.13)$$

Where $a_j(n)$ and $r_j(n)$ are the coefficients at resolution j.

Thus, for a J–level discrete wavelet decomposition of the given coefficients $a_0(n)$, a series of coefficients, $\{a_J(n), r_J(n), \ldots, r_1(n)\}$, is obtained. Because of the orthonormality of the wavelet transform, the number of coefficients after the decomposition is equivalent to that before decomposition. The synthesis of the signal from the wavelet coefficients obeys the following:

$$a_j(n) = \sum_k a_{j+1}(k)h(k-2n) + \sum_k r_{j+1}(k)g(k-2n) \quad (3.14)$$

If $\tilde{h}(n)$ and $\tilde{g}(n)$ are defined as

$$\tilde{h}(n) = h(-n), \quad \tilde{g}(n) = g(-n), \quad (3.15)$$

and regarded as the respective impulse responses of the *quadrature mirror filters* (QMF), H_Q and G_Q, which correspond to the halfband lowpass and highpass filter respectively, Eq. (2.24) can be conveniently implemented using convolution and downsampling (here the downsampling rate is 2 which implies that every other sample is omitted) techniques in the signal processing literature. Refer to Fig. 2.1 below for the diagram of the implementation for the 2-D case. If the wavelet decomposition is recursively applied to the output of the lowpass filter, a pyramid-structured decomposition is obtained.

It is interesting to note that the wavelet framework also links the tree-structured wavelet transform with wavelet packets. The library of wavelet packet basis functions $\{W_n\}_{n=0}^{\infty}$ can be obtained from a given W_0 as follows:

$$W_{2n} = \sqrt{2}\sum_k h(k)W_n(2x-k)$$
$$W_{2n+1} = \sqrt{2}\sum_k g(k)W_n(2x-k) \quad (3.16)$$

Where the functions W_0 and W_1 are set to the scaling function $\phi(x)$ and the mother wavelet function $\psi(x)$, respectively. Then, the functions $W_n(2^k x - l), k,l \in Z, n \in N$ form the orthogonal wavelet packet basis. The implementation of the wavelet packets will naturally lead to a tree-structured decomposition, thereby implying that both the outputs of the lowpass and highpass filters are recursively decomposed.

a) *Munich and Cambridge morlet 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

Global Journal of Computer Science and Technology (C) Volume XIII Issue V Version I Year 2013

that it has good properties in joint time frequency localization and has a well defined impulse response. In the time domain, the mother wavelet is a high-frequency vibration whose amplitude decreases when the time varies from zero to infinity. 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. This wavelet is defined.

IV. Evaluation of Sound Quality

Approval and intelligibility are the two most significant criteria to consider subjective quality. For this purpose, we invited several subjects to hear some mp3 files resulting from the coder based on FFT analysis, then on Munich and Cambridge Morlet wavelet transforms.

a) Simulation Results

In order to evaluate Morlet Munich and Cambridge wavelet compression, 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 (this type of sound contains a dialogue between two persons). The evaluation is based on the compression ratio defined as the quotient between the size of the original file and MP3 file. Tables 2 and 3 contains the type of the sound files used for the test, the average of compression ratio (Comp), their capacity (Cp), their duration (t) and the compression ratio calculated for each bitrate. The weakest compression ratio is given for the flow of 160 Kbits/s and the highest is given for the flow of 64 Kbits/s. However, the weaker the bitrate, the higher the compression ratio is and the less intelligible its quality becomes. Based on the results obtained in tables 2 and 3, the Morlet Munich filters bank takes account of the masking phenomenon and takes account of the critical bands. The second part of our evaluation will focus on comparing the compression ratio obtained from the coder based on wavelet analysis and the coder based on an FFT analysis versus using a bitrate of 128 kbits/s which gives the best compromise between compression ratio and sound quality [10]. Table 1 contains the type of the sound files used for the test, the average compression ratio (Comp) and the compression ratio values using FFT and wavelet (Munich and Cambridge). It reveals that sound compression using Morlet Munich wavelet is the best one. In fact, its average compression ratio presents an improvement of 12.67% and 3.2% in comparison to the FFT and Cambridge coders. The protocol of evaluation consists in listening to the compressed sound file. Then, the possibility is offered to the listeners to listen to it as long as they wish. The listeners are 24: 12 men and 12 women between 15 and 35 years old. Our aim is to classify the three coders for each type of sound in a statistic card as shown in Figure 7. The listeners rate the speech they have just heard on a five-point opinion scale, ranging from "bad" to "excellent". The ratings are assigned integer scores ranging from 1 for"bad" to 5 for "excellent". This comparison, based on results obtained from the statistic card, shows that Morlet Munich coder gives satisfactory results especially for Slow, Soul and Arabic sound. We noticed that for those sound files our coder received a significant number of notes ranging between 3 and 4. Consequently; it gives the best statistics for quality in comparison to the FFT and Cambridge coders.

V. Synthesis Result

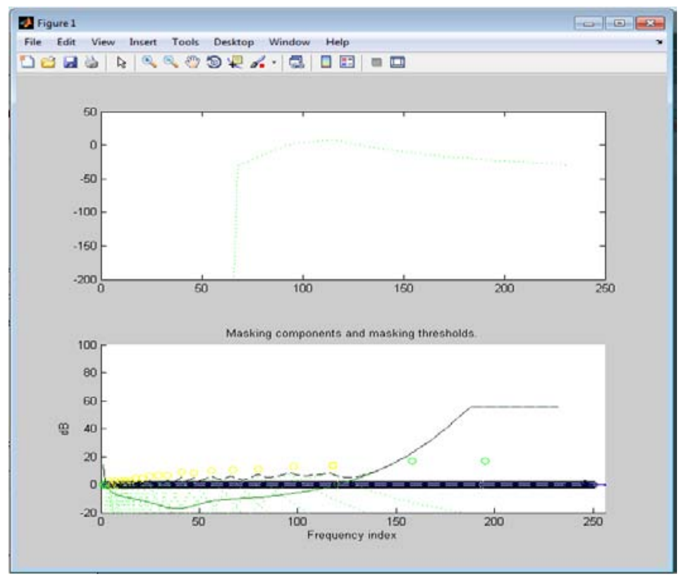

Figure 5.1 : Masking components thresholds

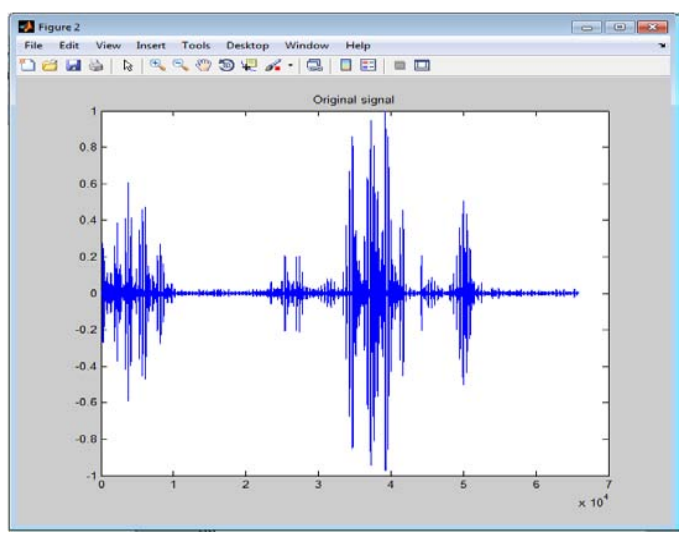

Figure 5.2 : Original Signal

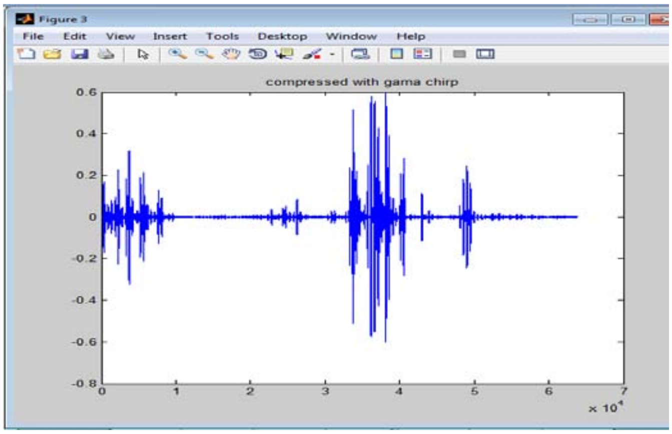

Figure 5.3 : Compressed with game chirp

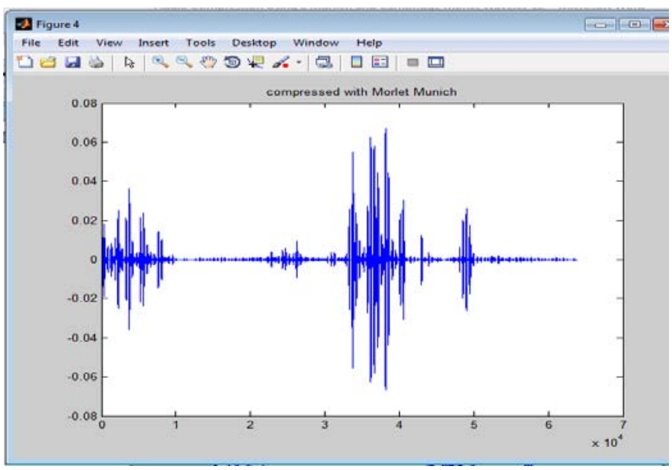

Figure 5.4 : Compressed with Morlet Munich

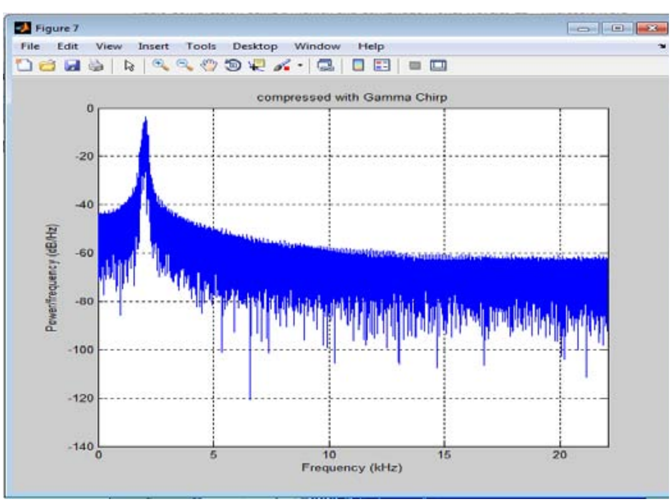

Figure 5.5 : Compressed with game chirp

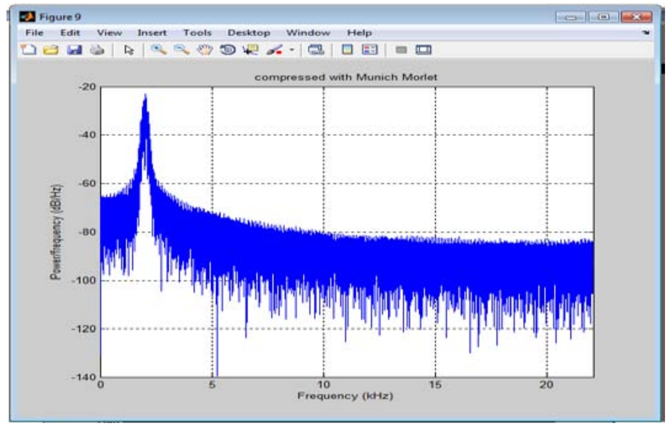

Figure 5.6 : Compressed with Munich Morlet

VI. Conclusion

This paper studies audio compression using a Munich and Cambridge Morlet wavelet 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 Munich Morlet wavelet packet decomposition offers a good performance. In fact, it gives the best compression ratio and sound quality in comparison to the FFT and Cambridge coders. A high selectivity was noticed and can lead to some interesting perspectives on audio coding using this type of psycho-acoustic model.

VII. Acknowledgement

The work is carried out through the research facility at the Department of Computer Science & Engineering, Department of Electronics & Communication Engineering and Department of Electrical & Electronics Engineering, HITS College of Engineering, Bogaram(V), Keesara(M), R.R. District, Hyderabad A.P., and Dept., Electronics & Communication Engineering from VJIT-Hyderabad. The Authors also would like to thank the authorities of JNT Kakinada, SKIT for encouraging this research work.

References Références Referencias

1. ISO /IEC 11172-3 (F), "Information technology - Coding of moving picture and associated audio for digital storage media at up to about 1.5Mbits/s Part3: Audio," 1999.
2. G. Hernandez, M. Mendoza, B. Reusch, and L. Salinas, "Shiftability and filter bank design using Morlet wavelet," Computer Science Society, Nov. 2004, pp. 141- 148.
3. D. Sinha and A. Tewfik,"Low Bitrate Transparent Audio Compression using adapted wavelets," Proc. IEEE Trans.Sig, Dec. 1993, pp. 3463-3479.
4. B. Carnero and A. Drygajlo, "Perceptual speech coding and enhancement using frame-synchronized

fast wavelet packet transform algorithm," Proc. IEEE Transaction on speech and audio processing, vol 47, NO 6, June 1999.
5. Daubechies, "Ten lectures on wavelet," SIAM, Philadelphia, PA, 1992.
6. J. O. SmithIII and J. S. Abel,"Bark and ERB Bilinear transforms," Proc. IEEE Transactions on Speech and Audio Processing, 1999.
7. W. W. Chang and C. T. Wang, "Auditory patterns," Proc. IEEE Transactions on speech and audio processing," vol. 4, NO 2, March 1996.
8. E. Zwicker and E. Terhardt, "Analytical expressions for critical band rate and critical bandwidth as a function of frequency," J. Acoust. Soc. Am, vol. 68, NO 5, Nov. 1980.
9. C. Wang and Y. C. Tong,"An improved critical-band transform processor for speech applications," Circuits and Systems, May 2004, pp. 461-464.
10. G. Davidson, L. Fielder, and M. Antill, "High-quality audio transform coding at 128 kbits/s," Proc. IEEE ICASSP, vol.2, 1990, pp. 1117-1120.
11. F. Baumgarte, "A computationally efficient cochlear filter bank for perceptual audio coding storage," Proc. IEEE ICASSP, vol.5, 2001, pp.3265-3268.
12. H. Fletcher, "Auditory patterns," Rev. Mod. Phys, vol.12, 1940, pp. 47-65.

This page is intentionally left blank