REALIZATION OF A PSYCHOACOUSTIC MODEL FOR MPEG 1 USING GAMMACHIRP WAVELET TRANSFORM

Samar Krimi, Kaïs Ouni & Noureddine Ellouze

Laboratory of Systems and Signal Processing (LSTS), National Engineering School of Tunis (ENIT)
BP 37, Le Belvédère 1002, Tunis, Tunisia
(Samar.Krimi, kais.ouni, N.ellouze) @enit.rnu.tn

ABSTRACT

In this paper, we present a new design of a psychoacoustic model following the example model used in audio standard MPEG-1 layer 3 with the gammachirp wavelet packet decomposition. The essential characteristic of this model is that it proposes an analysis by wavelet packet transformation on the frequency bands that come closer the critical bands of the ear that differs from the existing model based on an analysis by a short-term Fourier transformation.

1. INTRODUCTION

The techniques of coding make possible to reach high compression ratios while preserving a good subjective quality [1]. 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 [2][3]. 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 [4].
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. This compression is supposed to be transparent when the requirements of the model psychoacoustics are satisfied.
Nowadays, the international interest in audio coding is centred on the ISO/MPEG audio standards [1]. The designed model can be applied to the audio coders in sub bands based on a wavelets transformation, this when the decomposition in sub bands approximates the frequential decomposition of the sounds in the inner ear [10].

2. THE PROPOSED PSYCHOACOUSTIC MODEL

In this work, a new approach for modelling auditory masking based on gammachirp filters for application areas including speech /audio coding is developed, figure 1.
The perceptual encoding uses as a first psychoacoustic effect the mask one. The psychoacoustic model is based on the functioning of human ear.
This model analyzes the input signal on several consecutive stages and determines for every pad the spectrum of the signal. It models next the mask property of the auditory human system and estimates the minimal audible level.
Auditory masking is a well-known psychoacoustical phenomenon in which a weak signal (maskee) becomes inaudible (masked) in the presence of a stronger masker signal [2].
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 [4][7].
The notion of critical bands is primordial in psychoacoustic. In each of these bands whose width varies according to the precision of the ear at these frequencies.
In order to take into account the perception of the human ear and the psychoacoustic phenomena, a new frequency scale is introduced, the scale Bark in which, a growth of 1 Bark corresponds at frequency-division increase of one critical band.

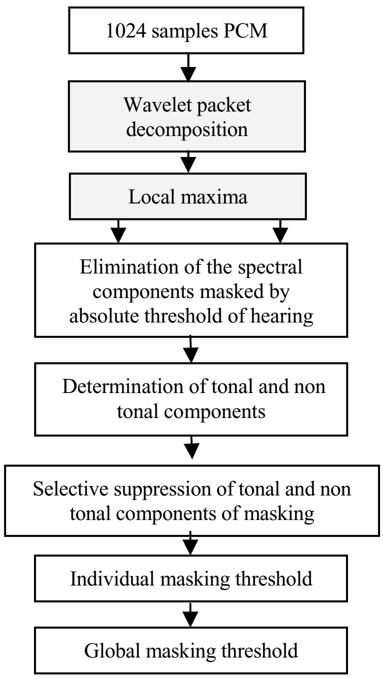

Figure 1: Design of psychoacoustic model I for Layer 3 by an analysis with gammachirp wavelets

In a first stage, we applied a decomposition of wavelets packet on 1024 points of the wav signal. We adopted the decomposition of Sinha and Tewfik (28 subbands) [9] which is a good approximation of the critical bands.
In a second stage, we calculate tonal and non tonal components. This stage begins with the determination of the local maxima, followed by the extraction of the tonal components (sinusoidal) and non tonal components (noise), in a bandwidth of a critical band.
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 tonal and non tonal components remaining 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 transform of the wavelets packet decomposition.

The calculation of the spectrum of the input signal was carried out using decomposition of wavelets packets whose connections are selected in such a way that subbands correspond to the best possible one to the critical bands.

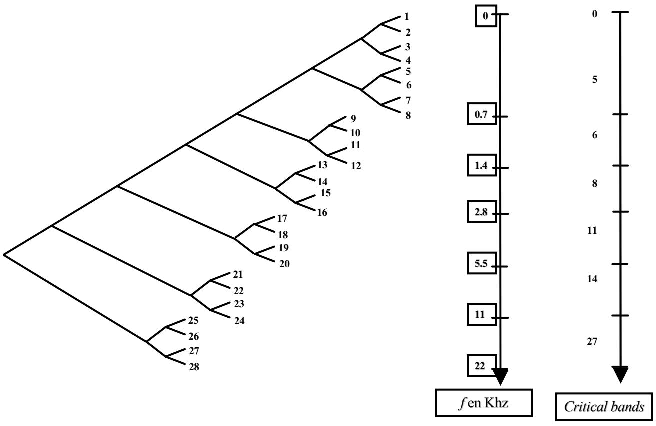

Figure 2: wavelet packet decomposition proposed by Sinha and Tewfik [9]

Psychoacoustic studies show that a frequency-to-place transformation 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. Moreover, the cochlea filter pass bands are of non uniform bandwidth. The 'critical bandwidth' is a function of frequency that quantifies the cochlear filter pass bands.
The wavelet packet decomposition defined by Sinha and Tewfik [9] can be adapted to the critical-band decomposition as shown in the figure 2.
The wavelet packet decomposition induces an organization of information according to a frequential segmentation as shown in figure 2: on each level of the tree, the information corresponding to the entire signal is divided into frequency bands of equal widths in bark scale.

2.1 The gammachirp psychoacoustical modelling for audio coding

It is known that human auditory frequency-selectivity is largely determined by signal processing in the cochlea [3]. The basilar membrane inside the cochlea is usually conceived (in psychoacoustical auditory masking models) as a bank of band-pass filters that have increasing bandwidths with increasing central frequency [4]. Many models have been proposed to simulate filtering properties in the inner ear. Among them, the gammachirp filters are a reasonably accurate description for auditory filtering at moderate intensity levels [5] [6].
The choice of the gammachirp filter is based on two reasons:

- 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 filterbank in time-domain models of auditory processing [5][7][11].
- 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 [7].

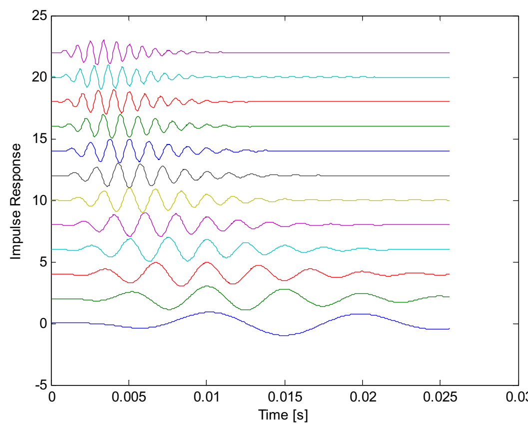

Figure 3: examples of gammachirp impulse responses centred on different frequencies

The gammachirp impulse response, shown in Figure 3, is essentially identical to that of the gammatone [6], but also includes a chirp term, c, in the carrier tone (c=0 for the gammatone filter).
The gammatone is the 1st order approximation of gammachirp, which fits better psychoacoustics and physiological data.
The gammachirp filter is a gamma distribution modulated at frequency f_0. It has as impulse response the following function [7]:

$$\psi(t) = A\, t^{n-1+ic} e^{-2\pi\beta t} e^{i2\pi f_0 t + i\phi} \qquad (1)$$

With $t > 0$
n: a parameter defining the order of the corresponding filter.
f_0: the frequency of modulation of the function gamma.
Φ: the initial phase.
A: amplitude.
The term $\beta = b\, ERB(f_0)$ characterizes the equivalent rectangular bandwidth (Equivalent Rectangular Bandwidth)

of the filter and b a parameter defining the envelope of the gammachirp filter.
The function $ERB(f_0)$ is defined by the expression [7]:

$$ERB(f_0) = 24,7 + 0,108 f_0 \quad \text{Hz} \qquad (2)$$

c: a factor introducing the asymmetry of this filter.

2.2 The gammachirp filter as a wavelet

The Gammachirp function which is a window modulated in amplitude by the frequency f_0 and modulated in phase by the parameter -c- can thus be seen as wavelet roughly analytical [11].
This wavelet has the following properties: it is with non compact support, it is not symmetric, it is non orthogonal and it does not present a scale function [11].
For this family of wavelet, the frequencies of modulation are $f_m = s_0^{-m} f_0$ and the bandwidths are $\beta_m = s_0^{-m}\beta$. s_0 is the dilation parameter and $m \in Z$.
To determine the central frequency of the gammachirp function, there is a compromise to make between three parameters which are dependent from each other. Thus, the central frequency of the mother wavelet f_0, the step fixes the selected u_0 and the number M of the daughters' wavelet generated of the wavelet mother, are all dependent.
The results showed that for a family of 512 daughters' wavelet gammachirp, value 1000 Hz is most compatible central frequency of the gammachirp function.
In the same way the average quadratic analysis error between the frequencies of optimal paving and the frequencies generated by the scale s_0 show that the scale s_0 = 1,13 is closest to the curve of optimal paving [11]. The choice of this step gives spectral widths of the daughters' wavelet deduced from the mother wavelet gammachirp close relations of those of the critical bands of the cochlear filters [11].
The frequency responses of the gammachirp filters, as seen in Figure 4, are asymmetric and exhibit a sharp drop off on high frequency side of the center frequency. This corresponds well to auditory filter shapes derived from masking data [8].

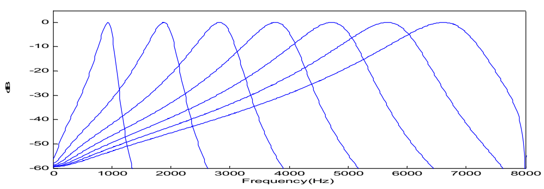

Figure 4: example of gammachirp filter bank

3. EVALUATION

It is a question of implementing the psychoacoustic model of audio standard MPEG-1 layer 3 using gammachirp wavelet packet. The goal is the optimization of the storage space of the whole file while keeping the sound quality of the same original file.
The program carried out converts a *wav* file format into an *mp3* file in agreement with the specifications of the international standard of the ISO/IEC 11172-3. The wave file used has the following characteristics:
- Mode: monophonic
- The sampling rate of 44.1 Khz and in format PCM.
- Binary rate: 96, 112, 128, 160 (Kbit/s)

3.1 Evaluation of sound compression ratio

3.1.1 Compression with Gammachirp wavelet packet

Table 1 indicates the type of the sound files used for the test, their duration, their capacity as well as the compression ratio, $CR = \frac{length(wav\ file)}{length(mp3\ file)}$, 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 96 Kbit/s and the weakest one is given for the flow of 160 Kbit/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 gammachirp function.

Table 1: sound compression ratio obtained by gammachirp 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	809	8.305	7.073	6.149	4.855
Opera	11	521	8.301	7.069	6.145	4.851

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.

3.1.2 Comparison of sound compression ratio between FFT and Gammachirp

Table 2 presents the compression ratio obtained from the coder based on an FFT analysis versus the coder based on wavelet analysis.
We notice that for a flow of 128 kbit/s which gives the best compromise between compression ratio and sound quality,

the wavelet packet gammachirp gives the best compression ratio, it is the wavelet which approaches the cochlear filters of the ear.

Table 2: Compression ratio obtained from the FFT analysis coder versus the wavelet analysis coder

Type of sound file	Gammachirp	FFT
Rock	6.167	5.210
Classic	6.160	5.203
Jazz	6.154	5.197
Voice	6.149	5.192
Opera	6.145	5.188

3.2 Evaluation of sound quality

There are several criteria to consider subjective quality: intelligibility and approval are the two most significant criteria since they measure the listening comfort or the effort to be produced and comprehensibility. We invited for this purpose several subjects to hear, through a reader mp3, the file mp3 resulting from our coder based on wavelet packet decomposition (gammachirp) then the same file resulting from the coder based on FFT analysis. The population chosen to carry out these statistics is composed of 20 peoples: 10 men and 10 women between 20 and 30 years old. These subjects are not experts of the field.
The protocol of comparison consists in listening to the compressed original sound file. Then, the possibility is offered to the tester to listen to it as much time as he wishes. The test on the following sound file is accessible when a note was validated. The order of appearance of the sound files is drawn randomly so as to limit the influence of the position of the files on the general average of the notes (factors such as training or tiredness).
Each subject classifies the two coders for each type of sound in a statistic card. The coder who gives best sound quality approaching original sound quality will have 4 points. The coder giving a sound quality less near to original quality will have 1 point.
Based on results obtained from this comparison the gammachirp wavelet packet coder gives satisfactory results especially for the voice file and the opera file as shown on figure 5. We noticed that for rock, voice and opera sound files our coder received a significant number of notes ranging between 3 and 4 consequently it is the coder who gives the best statistics to the level of quality compared to the FFT coders.

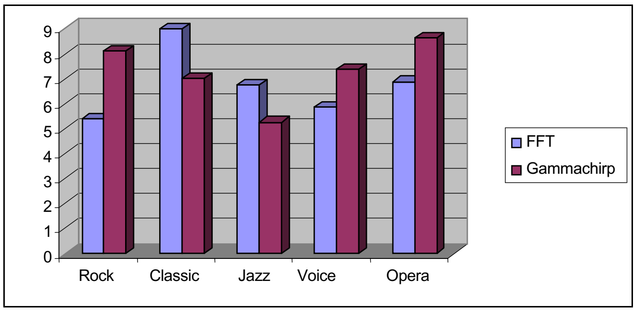

Figure 5: Sound quality histogram evaluation

4. CONCLUSION

We have presented in this paper a psychoacoustic model for audio coding based on a wavelet packet decomposition. The wavelet packet uses a gammachirp filter which is an accurate model for the cochlear spectral processing.
To evaluate this wavelet packet coder, we have achieved two types of tests, an objective test and a subjective test. The first one is based on the obtained compression ratio of sound files from various types of music, versus the same files obtained from the existing FFT MPEG coder. The second one aims to test subjective quality by some subjects chosen to carry out a measure of this quality in comparison to the FFT coder.
In both cases, the obtained results show that the gammachirp wavelet packet coder gives, in these preliminary tests, good performances and can lead to some interesting perspectives on audio coding using this type of psychoacoustic model.

REFERENCES

[1] ISO/IEC 11172-3 (F), Norme Internationale : Technologies de l'information-Codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1,5 Mbits/s – Partie 3 : Audio, 1993.
[2] B. C. J. Moore. An Introduction to the Psychology of Hearing. Academic press, San Diego, 1997, ISBN 0-12-505627-3.
[3] J. E. Hawkins Jr. and S. S. Stevens The masking of pure tones and of speech by white noise. J. Acoust. Soc. Am., 1950, vol. 22, pp. 6-13.
[4] H. Fletcher. Auditory patterns. Rev. Mod. Phys., 1940, 12, pp. 47-65.
[5] R. D. Patterson, M. Allerhand and C. Giguère. Time domain modelling of peripheral auditory processing: A modular architecture and a software platform. J. Acoust. Soc. Am., 1995, vol. 98, pp. 1890-1894.
[6] E. Ambikairajah, J. Epps and L. Lin. Wideband speech and audio coding using gammatone filter banks. Proc. ICASSP'01, Salt Lake City, USA, May 2001, vol. 2, pp. 773-776.
[7] Irino, T., Patterson, R. D., A time-domain, level-dependent auditory filter: The gammachirp, JASA, Vol. 101, No. 1, pp. 412-419, January 1997.
[8] Alex Park, Using the gammachirp filter for auditory analysis of speech May 14, 2003. 18.327: Wavelets and Filter banks
[9] D.Sinha and A.Tewfik, Low Bit Rate Transparent Audio Compression using adapted wavelets, IEEE Trans.Sig.Proc, pp.3463-3479,Déc 1993
[10] M.Rosa Zurera, F.Lopez Ferreras, M.P Jarabo Amores, S.Maldonado Bascon, N.Ruiz Reyes, A new algorithm for translating a psychoacoustic information to the wavelet domain. Signal Processing 81(2001) 519-531
[11] Ouni Kais," Contribution to the vocal signal analysis using knowledges on the auditory perception and multiresolution time frequency representation of the speech signals" (in french), PhD Thesis on Electrical Engineering. National Engineering School of Tunis. February 2003.