Asian Journal of Information Technology 6 (5): 640-646, 2007

Implementation of a Psychoacoustic Model for Audio Watermarking

Said Ghnimi and Adnen Cherif
Laboratory of Signal Processing, Faculty of Science Tunis 2092, El Manar, Tunisia

Abstract: In this study, we will present 2 psychoacoustic models used in an audio watermarking chain. The objective of this study is to present the psychoacoustics model of the human auditory mecanism and its implementation. The main application is the speech masking inherent in the human ear that will be used in the general outline of watermarking. The idea consists in modifying the power spectral concentration of the original code in order to insure a high sound audibility. It consists in generating a threshold starting from the masking phenomenon and a relative ponderation of the coefficients of MPA output signal contents.

Key words: Watermarking, audio and speech processing, MPA (psychoacoustic auditory model), masking, SNR, Matlab

INTRODUCTION

The insertion of watermarking consists in adding to an audio signal a random pseudo sequence. In order to guarantee the watermarking's speech inaudibility, the sequence must be made in spectral according to a masking threshold obtained using a psychoacoustic model (Zwicker and Zwicker, 1991).

Masking is the phenomena where a sound (masked) is made inaudible because of the presence of another sound (named masking) (Painter and Spanias, 2000) it is used in audio compression and watermarking. We can observe in particular two kinds of masking: Simultaneous masking (frequential) and temporal masking (John, 1998). In the literature, several psychoacoustic models are considered. The most important are these described by Leandro (2002) and Garcia (1999).

The purpose of this study is to set up a psychoacoustics model adapted to a system of watermarking. For that, we formalize the constraints to be respected by the psychoacoustic model (MPA) and define thereafter criteria of performance for the context of watermarking.

PRINCIPLE OF THE PSYCHOACOUSTIC MODEL

Psychoacoustic elements: Figure 1 gives synoptic operation of the human auditive system (Kim, 2000). It includes the outer ear, the middle ear and the inner ear. The outer ear collects the signal form the difference in pressure of the air.

The role of the middle ear (tympanum, hammer, clamp and anvil) is to transmit the audio signal to the cochlea. The mechanical vibration of this zone transmits the signal

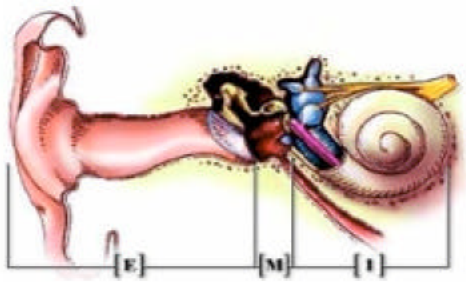

Fig. 1: Diagram of the human auditive mecanism (Pujol, 2004): Outer ear [E], Middle ear [M] and the Inner ear [I]

the inner ear. The latter contains the basilar membrane and supports the body of Corti connected to the auditive nerve and cilices cells. These cells generate electrochemical impulses associated with the vibrations of the membrane. The transmission of information is made after several transformations (mechanical and electric) from entry in the auditive system until its arrival to the brain. The principal role of frequency response system is played by the basilar membrane located in the cochlea. Its geometrical and mechanical characteristics make it a particular system of resonance.

Hearing threshold: The absolute threshold varies according to the frequency. It is given in an experimental way while making listen on a subject of the noises with narrow bands, centered at different frequencies. According to Fig. 2, any noise below this curve is unperceivable by the human ear. One can approximate the

Corresponding Author: Said Ghnimi, Laboratory of Signal Processing, Faculty of Science Tunis 2092, El Manar, Tunisia

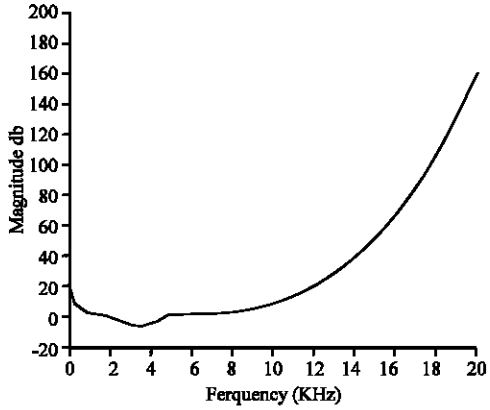

Fig. 2: Absolute threshold of hearing

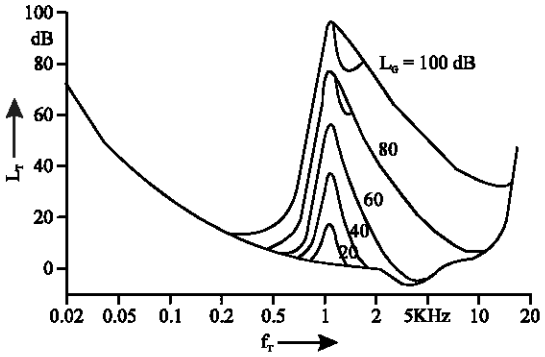

Fig. 3: Curve of masking of a pure sound (Painter and Spanias, 2000)

absolute threshold of hearing by the following function which expresses the absolute threshold of hearing in dB according to the frequency in KHz (Johnston, 1988):

$$S(f)=3.64f^{-018}+0.001f^{4}-6.5\exp(-0.6(f-3.3)^{2})-0.18\max(f-5.0)^{2} \quad (1)$$

Concept of critical band: The critical bands are defined as the audible areas of the frequency spectrum where the human auditive system is able to carry out a frequential integration, which means that on these areas, the ear carries out a sum of the contributions of the sounds being around a frequency given to generate an auditive stimulus. The human ear is more perceptible a nonlinear scale called barks given by next expression (Zwicker and Fedtkeller, 1981).

$$F_{bork}=13\arctan(0.76\frac{f_{hertz}}{1000})+3.5\arctan\left[\left(\frac{f_{hertz}}{7500}\right)^{2}\right]$$

Concept of frequency masking: Masking is a phenomenon used in audio compression and watermarking. Masking is the phenomenon where a sound (named its masked) is made inaudible because of the presence of another sound (named its masking). We distinguish in particular two types of masking: simultaneous masking (or frequential) and temporal masking.

In the curve represented in Fig. 3 for a periodic signal with f_0 = 1 kHz and for several values of $(\sigma_0)^2$, we observe that around f_0, the curve of hearing threshold has a triangular form. The index of masking $\frac{(\sigma_0)^2}{(\sigma)^2}$ depends only on more or less tonal nature on masking approximately 20 dB for a pure sound and it passes to approximately 5 dB for a noise to narrow band.

IMPLEMENTATION OF THE PSYCHOACOUSTIC MODEL

According to Fig. 4, we used the psychoacoustic model MPA to extract from the audio signal a threshold of masking translating the frequential limit of audio inaudibility depending of the speaker. This threshold is applied to the filter H(f) which formats the modulated signal so that its power spectral concentration coincides with the threshold of audio signal masking.

Two psychoacoustics models are considered in the literature on the chain of watermarking. The first one is the model described by Léandro (2000) and the second is described by Garcia (1999).

In this study, we propose to describe the characteristics of these models.

Model of Leandro: The threshold of masking by Gomés is a simplified version of the MPEG 1 model (Codage, 1993). The DSP (power spectral concentration) of the audio signal is divided into 4 bands. In each one of these sub bands, the DSP is submitted to a dynamic compression then a smoothing operation. The threshold obtained is then modified to obtain a power report/ratio between the masking threshold and the signal audio (RMS) acceptable. This threshold offers finally the advantage of a very low complexity of calculation.

Leandro MPA algorithm: The Leandro algorithm is illustrated by Fig. 5. First, we compute the speech power spectral concentration by the periodogramme method. Then, we divide the frequency axis into four intervals. In

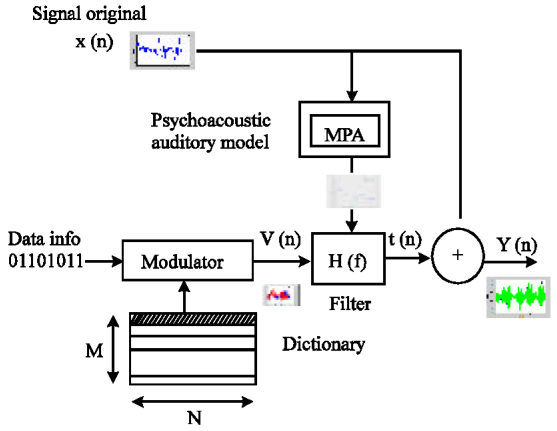

Fig. 4: Diagram of the watermarking system (transmitter)

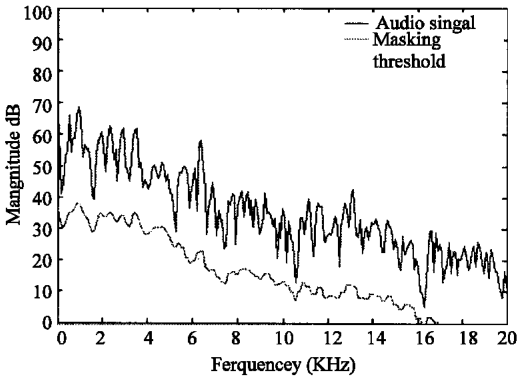

Fig. 6: The masking threshold by the Leandro model

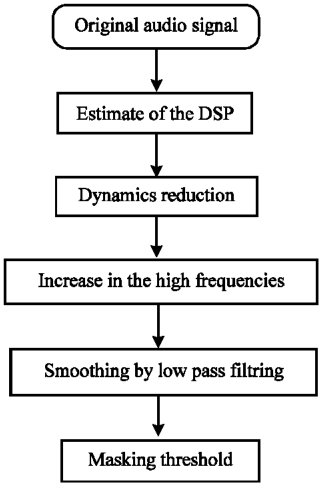

Fig. 5: Description of the Leandro algorithm

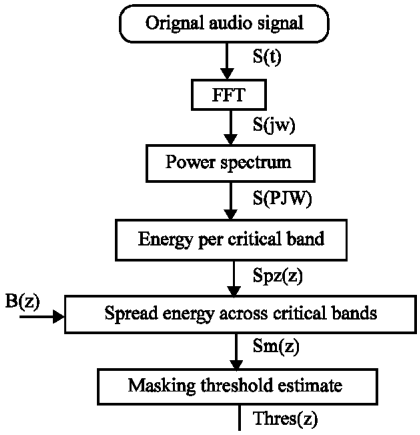

Fig. 7: Description of the Garcia algorithm

each interval, we calculate the DSP average and then we compress the curve around this average in order to reduce its dynamics to 50%.

Here, Léandro does not give any justification for the parameter of reduction. Certainly, a good alternative of this parameter will give a better estimate of the threshold (Colomes *et al*., 1995).

To rise of the masking threshold in the high frequencies, we add a few decibels to the curve on the last quarter. The added quantity varies linearly from zero (for F = 3fe/8) to 25dB (for f = fe). This is done because the ear becomes less sensitive in the high frequencies. We remove then the fast variations of the curve using a low passe filtering. The cut-off frequency of the filter is adjusted has Fc = Fe.

Figure 6 illustrates the simulation masking threshold and the original audio signal.

The curve of the masking threshold is moved to the bottom of a parameter fixes a priori has -10 dB. This parameter can be organised according to the desired power of watermarking. A good alternative of this parameter will guarantee a more powerful model psychoacoustics to us.

Garcia model: The psychoacoustic model of Garcia takes as a starting point the MPA of MPEG 2. Garcia proposes to calculate the masking threshold on the scale of the critical bands in order to respect the physiology of auditive perception as well as possible. The DSP of the audio signal is transposed in the scale of Barks by integration of the power in each under band. A single function of spreading out (without reference to tonality of the components) is applied to model the effects of masking of the basilar membrane.

Garcia MPA algorithm: The Garcia algorithm is illustrated by Fig. 7.

After acquisition, we compute the short time DSP and energy of the speech signal x(n). To calculate the

threshold of masking, one needs a scale in Barks. One notes by Kmin and Kmax the frequential indices representing the lower and higher critical band limits. In the model of Garcia, we finally, we must take into account the absolute threshold of hearing, by using the following equation:

$$Spz(z)=\sum_{k=k_{min}}^{k_{max}^{(z)}} Sp(k) \quad (3)$$

With $z=1, 2, 3...Z_T$

The third step of the Garcia algorithm is the calculation of the energy spread across critical bands. We model the curve of masking by convoluting the basilar spectrum by a function of spreading out: This we obtain the basilar excitation. In this model, we work with energy by critical band. We obtain an energy spread out by band criticizes Smz (z) which one compares to the excitation in critical band Z given by the function of spreading out (Beerends and Stemerdink, 1994):

$$Smz(z) = Spz(z) * B(z) \quad (4)$$

With :

$$B(z) = 15.81 + 7.5(z + 0.474 - 17.5\sqrt{1+(z+0.474)^2})$$

Finally, it remains to calculate the index of masking. For that, it is necessary to have an estimation of the tonality of the spectrum. Garcia calculates the Spectral Flatness Measure (SFM) as:

$$SFM_{db} = 10\log_{10}\left\{\frac{\left(\prod_{z=1}^{Z_t} Spz(z)\right)^{\frac{1}{z_t}}}{\frac{1}{Z_t}\sum_{z=1}^{z_t} Spz(z)}\right\}^{\frac{1}{z_t}} \quad (5)$$

With Z_T is the total number of critical bands in the current window. The index of tonality is thus calculated:

$$\alpha = \min\left(\frac{SFM_{db}}{SFM_{db\,max}}, 1\right) \quad (6)$$

With: $SFM_{dbmax} = -60db$.

If the analyzed screen is tone-like, á will be close to 1 and if it is noise-like then á will be close to 0. We calculate then the index of masking given by:

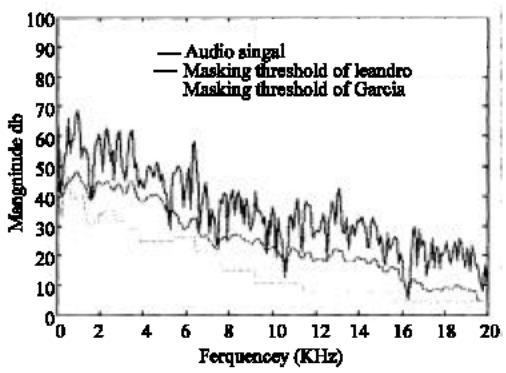

Fig. 8: Comparison of the masking thresholds for the two models: Leandro and Garcia

$$O(z) = \alpha(14.5 + z) + 5.5(1 - \alpha). \quad (7)$$

This will enable us to calculate the threshold of masking per critical band:

$$Seuil(z) = 10^{\log_{10}(sm(z)) - \frac{O(z)}{10}} \quad (8)$$

We then calculate the standardized threshold as:

$$Seuil_{norm}(z) = \frac{Seuil(z)}{k_{max}(z) - k_{min}(z) + 1} \quad (9)$$

Finally, we must take into account the absolute threshold of hearing, by using the following equation:

$$Seuil_{Garcia}(z) = \max(seuil_{norm}(z), TH) \quad (10)$$

Where TH is the absolute threshold of hearing given by:

$$TH = \max(|P_p(jw)|) \quad (11)$$

$P_p(jw)$ is the power of the signal p (t) = sin (2 p 4000t).

Comparison of the thresholds: Figure 8 show the masking thresholds simulated by the models of Garcia and Leandro.

The observation of the various thresholds shows that for the Léandro model we obtain a curve which follows the variations of the DSP of the signal, but its dynamics and its power are less significant. Hence, For the Garcia model, we obtained a continuous curve. It is the description on which the basilar membrane varies from a critical band to another.

RESULTS AND DISCUSSION

A psychoacoustic model in a context of watermarking must to respect a certain number of constraints:

- Guarantee the inaudibility of watermarking.
- The robustness with the various audio disturbances.

Experimental protocol: During this resaerch, we worked with 10 sample pieces of music and speech signal of various kinds. We watermarking different speech sequences of 12 sec and used 512 points analysis frames (20 ms/trame) to deduce the performances from each model of them.

The literature presents various criteria and protocols rof measurement of the sound quality of the watermark signal (David, 1997) such as:

- Subjective measurements based on listeners decisions
- Objective measurements based on SNR values (UIT-RBS, 1998)
- Robustness of watermarking (Mitchell *et al*., 1998).

For all the 10 tested pieces described in the experimental protocol, we obtained like average note -0.2 for the model Leandro, -0.8 for the Garcia model.

Table 1: shows the notes allotted for each model for the various pieces of speech.

We note that the Leandro model gives the best results. Indeed, all the notes given for this model are in the vicinity of 0, which implies that watermarking is almostunperceivable. For the Garcia model, the majority of the notes are acceptable.

Comparison of the masking threshold before and after watermarking: In this study, to compare the performances of the two psychoacoustic models, we will compare the masking thresholds before and after watermarking as illustrated in Fig. 9 and 10.

For the Leandro model, the difference between the thresholds of masking always presents a peak in the vicinity of 5 kHz. For the Garcia model, the difference becomes more significant in the high frequencies (Fig. 11 and 12).

Table 1: Standard reference auditory speech signal

Sistorsion	Signal reference	SDG
Imperceptible	5	0
Persceptable	4	-1
Persceptable but litte hardness	3	-2
Persceptable but hardness	2	-3
Persceptable but very hardness	1	-4

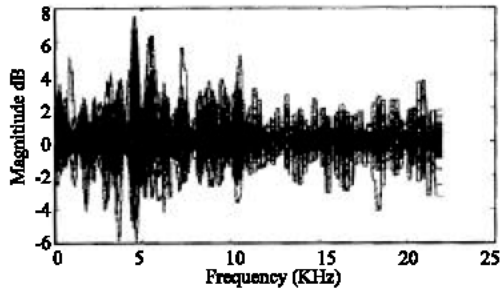

Fig. 9: Difference of the masking thresholds: Leandro Frequency (Khz)

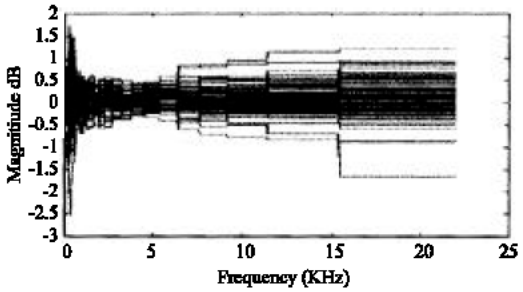

Fig. 10: Difference of the masking thresholds: Garcia

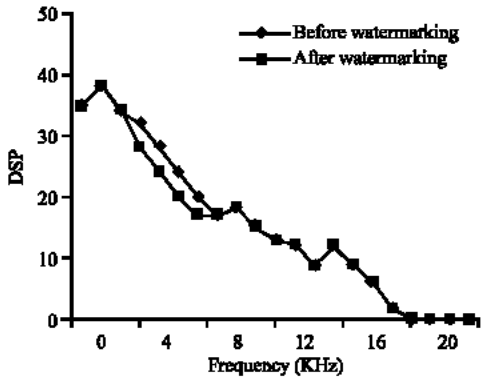

Fig. 11: Measures the masking threshold of Leandro

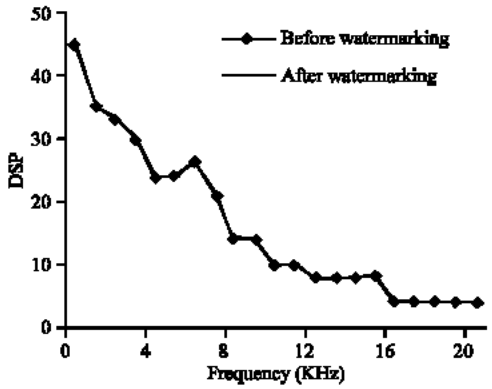

Fig. 12: Measures the masking threshold of Garcia

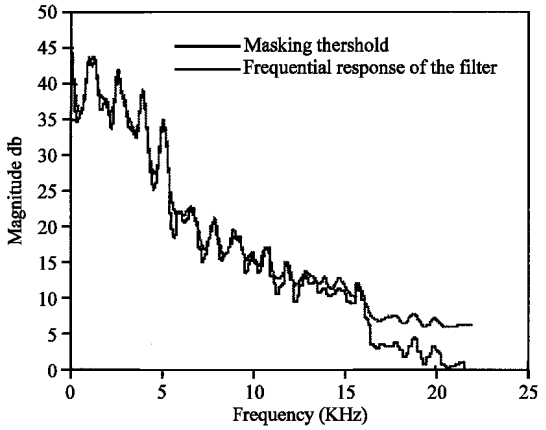

Fig. 13: Comparison of the threshold and the frequency response for the Leandro model

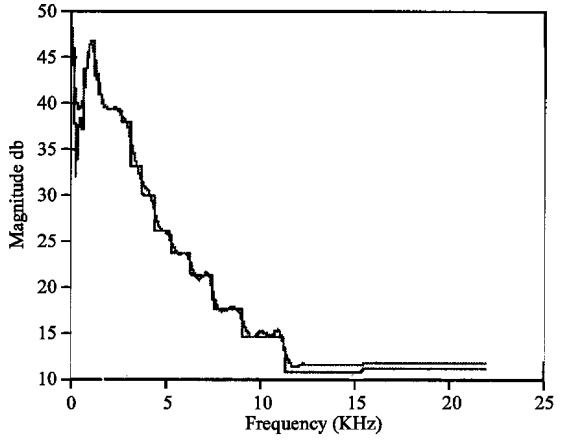

Fig. 14: Comparison of the masking threshold and the frequency response for the Garcia model

The results given by the Garcia model in terms of difference between the threshold of masking of the original signal and of the watermarking signal are better. The model of Garcia seems to be more robust and we obtain a better estimate of the threshold in the reception, which implies a better restitution of the modulated signal.

Comparison of the frequency response: In Fig. 13 and 14, we compare for each model the masking threshold and the filter frequency response to identify the model which ensures the best formatted signal.

For the various tested speech signals, the results given by Garcia model are better then those obtained from Leandro MPA model. This implies that this model presents the advantage of ensuring the best formatted watermarking audio signal. This can be explained by the shape of the curve given by Garcia which is easier to approximate by an auto regression filter AR than the curves of Leandro.

CONCLUSION

In this study, we have succeeded in implementing two psychoacoustics models in a chain of audio watermarking. In comparison of the performances of the two models of Garcia and Leandro, we find that the Leandro model ensures a very good quality of watermarking in term of inaudibility. However, the Garcia model has a better robustness. Our principal contribution in this study is the adaptation of the Matlab, our system can answer as much as possible to problems encountered in audio watermarking, such as the constraints of robustness and imperceptibility. This study is very interesting because it can be integrated and implemented in a real time speech watermarking system such as a DSP or FPGA.

REFERENCES

Beerends, J.G. and J.A. Stemerdink, 1994. A Perceptual Speech-Quality Measure Based on a Psychoacoustic Sound Representation J. Audio Eng. Soc., 42: 115-123.

Codage, 1993. de l'image animée et du son associe pour les supports de stockage numérique jusqu'a environ 1.5 mbit s 1, Tech. Rep, ISO/CEIL 11172,

Colomes, C., M. Lever, J.B. Rault, Y.F. Dehery and G. Faucon, 1995. A Perceptual Model Applied to Audio Bit-Rate Reduction, J. Audio Eng. Soc., 43: 233-239.

David, S., 1997. Méthodes d'évolution subjective des dégradations faibles dans les systèmes audio compris les systèmes sonores multivoies, Genève, BS 11116.

Garcia, R.A., 1999. Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum, Theory 107th AES convention, New York.

John Watkinston, 1998. La réduction de débit en audio et vidéo. Eyrolles.

Johnston, J.D., 1988. Transform Coding of Audio Signals Using Perceptual Noise Criteria, IEEE J. Selected Areas in Commun., 6: 314 - 323.

Kim, H., 2000. Stochastic model based audio watermark and whitening filter for improved detection. Proceedings of the international conference on Acaustic, speech signal processing.

Leandro de Campos, 2002. Tatouage de signaux audio. PHD thesis, France.

Méthodes, 1998. de mesure objective de la qualité du son perçu. Recommandation UIT-RBS.1387.

Painter, T. and A. Spanias, 2000. Perceptual coding of digital audio In Proc IEEE., 88: 451-513.

Pujol, R., 2004. Promenade autour de la cochlée, http://www.iurc.montp.inserm.fr/cric/audition/index.htm

Zwicker, E. and E. Feldtkeller, 1981. Psychoacoustique, l'oreille récepteur d'information, Masson, collection technique et scientifique de télécommunication.

Zwicker, E. and U.T. Zwicker, 1991. Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System, J. Audio Eng. Soc., 39: 115 -126.