IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834, p- ISSN: 2278-8735.Volume 4, Issue 6 (Jan. - Feb. 2013), PP 73-78
www.iosrjournals.org

Digital Audio watermarking using perceptual masking: A Review

Ms. Anupama Barai[1], Associate Prof. Rohini Deshpande[2]
[1, 2](EXTC, K J Somaiya College of Engineering/ Mumbai University, India)

Abstract : *Sharing electronic files on the internet has grown extremely fast over the last decade due to large volume of the mobile phones. These files include diverse forms of multimedia such as music, video, text and image. However digital files can be easily copied distributed and altered leading to copyright infringement. It is this ease of reproducing that causes copyright violations. Composers and distributors are more focused on implementing digital watermarking techniques to protect their material against illegal copying. Digital audio watermarking technique protects intellectual property by embedding watermark data into the audio file and recovering that information without affecting the audio quality of the original data. In this paper an overview of fundamental concepts of digital audio watermarking using perceptual masking is presented which includes a watermarking procedure to embed copyright protection into digital audio by directly modifying the audio samples. The procedure directly exploits temporal and frequency perceptual masking to guarantee that the embedded watermark is inaudible and robust. The watermark is constructed by breaking each audio clip into smaller segments and adding a perceptually shaped pseudo-random sequence. The noise-like watermark is statistically undetectable to prevent unauthorized removal.*

Keywords -*Audio watermarking, Copyright Protection, Embedding, Frequency Masking Perceptual masking, Psychoacoustic Auditory Model, Temporal Masking.*

I. INTRODUCTION

On line distribution of digital media including images, audio, video and documents has proliferated rapidly in recent years. In such environment it is convenient to get the access to various information resources. Although with the ease by which the digital formatted data can be copied and edited, copyright infringement like illegal reproduction and distribution has arisen and greatly spoilt the originator's passion for innovation. To prevent such iniquities, the enforcement of ownership management has claimed more and more attention. As a result, novel watermarking technique is introduced for copyright protection [1].

Watermarking is the process of encoding hidden copyright information in digital data by making small modifications to the data samples[3]. Unlike encryption, watermarking does not restrict access to the data. Once encrypted data is decrypted, the media is no longer protected. A watermark is designed to *permanently* reside in the host data. When the ownership of a digital work is in question, the information can be extracted to completely characterize the owner. Most schemes utilize the fact that digital media contain perceptually insignificant components which may be replaced or modified to embed copyright protection. However, the techniques do not *directly* exploit spatial/temporal and frequency masking. Thus, the watermark is not guaranteed inaudible. Furthermore, robustness is not maximized. The amount of modifications made to each coefficient to embed the watermark is estimated and not necessarily the maximum amount possible.

In this paper, we introduce a novel watermarking scheme for audio which exploits the human auditory system (HAS) to guarantee that the embedded watermark is imperceptible. As the perceptual characteristics of individual audio signals vary, the watermark adapts to and is highly dependent on the audio being watermarked[2]. The watermark is generated by filtering a pseudo-random sequence (Author id) with a filter that approximates the frequency masking characteristics of the HAS. The resulting sequence is further shaped by the temporal masking properties of the audio. Based on pseudorandom sequences, the noise-like watermark is statistically undetectable.

II. DIGITAL AUDIO WATERMARKING

Digital watermarking is the process that embeds copyright information as watermark into the multimedia object, so that the watermark can be extracted to make an assertion about the ownership. The general schematic diagram of watermarking is shown in figure 1.

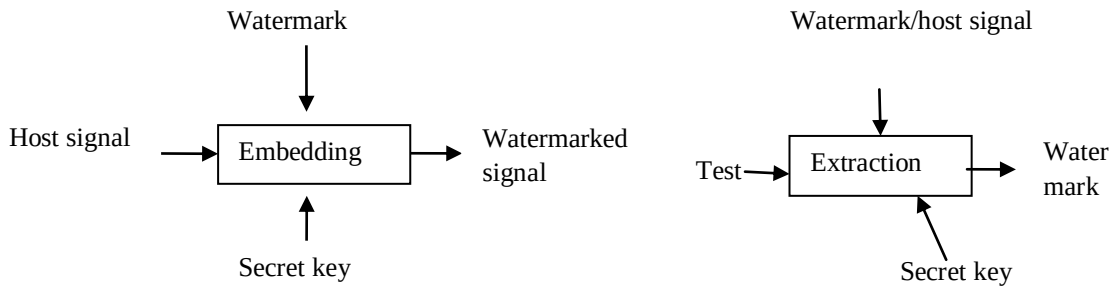

Fig.1 General Schematic diagram of watermarking

From the point of view of information hiding, watermarking is a technique of hiding messages, may be secret signatures, in the host carrier for the purpose of identification annotation and rights management. For example, a common form of hidden writing is using invisible inks. So the secret message can only be read by processed with some prescribed chemicals in a certain sort of way. Specifically, research is focused on embedding imperceptible robust and secure watermark for copyright protection. These watermarks are permanent signatures, difficult to remove without degrading the quality of the host media. When disputes happen the watermark could be extracted as reliable proofs for assuring the authentication[5]. The further subheadings cover the fundamental concepts of the basic theories Psychoacoustic masking and Spread Spectrum, watermarking and extraction.

III. Psychoacoustic Masking

Psychoacoustic explains the subjective response to everything we hear. It seeks to reconcile acoustic stimuli scientific, objective and physical properties that surround them with the physiological and psychological responses evoked by them. In simple words it is the science that studies the statistical relationships between acoustical stimuli and hearing perception, in order to explain the auditory behavioral responses of human listeners, the abilities and limitations of the human ear and the auditory complex process that occur inside the brain. Because the process of embedding the watermark is required to be imperceptible, understanding the principles of psychoacoustic and making use of its perceptual properties is helpful for watermark implementation.[6]

Hearing involves a behavioral response to the physical attributes of sound including intensity, frequency and time based characteristics that permit auditory system to find clues that determine distance, direction, loudness, pitch and tones of many individual sounds simultaneously. It is the sense by which sound is perceived hand hearing of humans is performed by the auditory system: sound as pressure waves is detected by the ear and transduced into nerve impulses that are perceived by the brain. The range of human hearing is 20 Hz to 20 kHz, where human speech mainly falls between 100 Hz and 8 kHz. The ear plays an important role in human auditory system HAS[7]. Ear is sub divided into outer, middle and inner ear. Outer ear i.e. pinna captures sound and directs through the ear canal till tympanic membrane (eardrum). Middle ear processes the sound waves into mechanical vibrations through ossicles till oval window. These vibrations are transduced into neural impulses within the cochlea of inner ear. These then enter the brain via auditory nerve fibres. The Cochlea is the main organ. It's a snail shaped fluid filled chamber separated by basilar membrane. The basilar membrane is about 32mm long and the organ of Corti the receptor rests on it. The organ of Corti contains specialized cells called hair cells including inner and outer which translate fluid motion into electrical impulses for the auditory nerve.

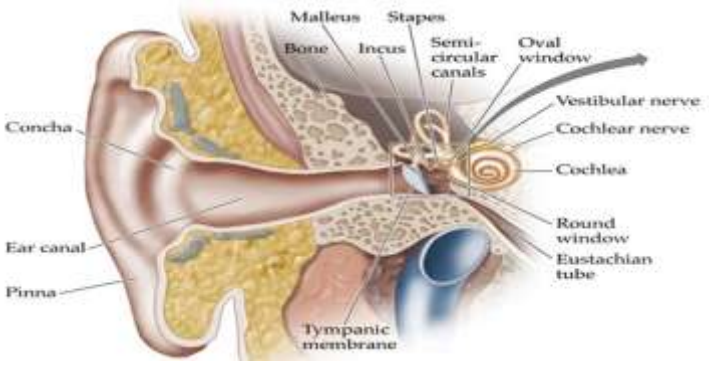

Fig 2. Diagram of Human hear

Experimental studies show that the basilar membrane is a resonant structure that has different resonant properties at different points along its length, acting as a spectral analyzer. Its motion is like travelling wave being the greatest at the point where the frequency of the incoming sound matches that of the movement of the membrane as shown figure given below. In the cochlea, low frequency signals will induce oscillations that reach maximum displacement at the apex of the BM, while high frequency signals induce oscillations that reach maximum displacement at the base of BM near oval window.

The perception of a sound is related to not only its own loudness and spectrum, but also to its neighbor components which is the effect of masking phenomenon. Masking is the phenomenon in which a very low but audible sound (known as the maskee) becomes inaudible in the presence of another loud audible sound (known as the masker). It plays an important role in hearing sensation. Masking is of two types: simultaneous and non simultaneous masking. Out these till now simultaneous masking is under study. It is also known as frequency masking which involves masking between two sounds with close frequencies but their loudness is different depicted by the figure below. There are two types of masking observed in the HAS –*frequency masking* and *temporal masking*. Watermarking procedure directly exploits both frequency and temporal masking characteristics to embed an inaudible and robust watermark

Frequency masking refers to masking between frequency components in the audio signal. If two signals, which occur simultaneously, are close together in frequency, the stronger masking signal may make the weaker signal inaudible. The masking threshold of a masker depends on the frequency, sound pressure level (SPL), and tone-like or noise like characteristics of both the masker and the masked signal . It is easier for a broadband noise to mask a tonal, than for a tonal signal to mask out a broadband noise. Moreover, higher frequency signals are more easily masked. The human ear acts as a frequency analyzer and can detect sounds with frequencies which vary from 10 to 20000 Hz. The HAS can be modeled by a set of 26 band-pass filters with bandwidths that increase with increasing frequency. The 26 bands are known as the critical bands. The critical bands are defined around a center frequency in which the noise bandwidth is increased until there is a just noticeable difference in the tone at the center frequency. Thus, if a faint tone lies in the critical band of a louder tone, the faint tone will not be perceptible. Frequency masking models are readily obtained from the current generation of high-quality audio codes.

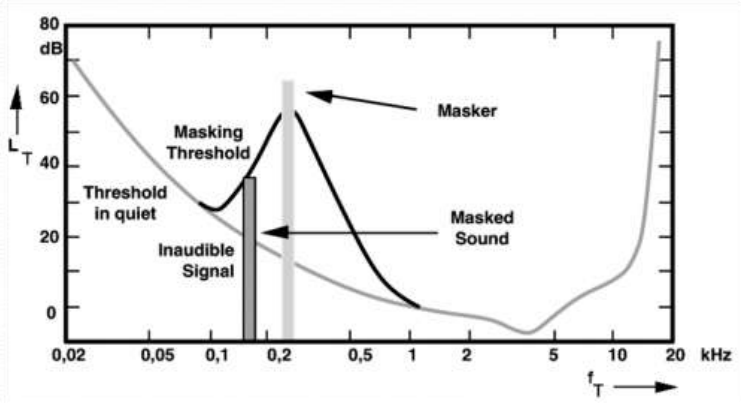

Fig. 3 Masking Effect

IV. Psychoacoustic Modeling

The implementation of the psychoacoustic model is flexible, depending on the required accuracy and the intended applications**.** The following steps describe the implementation of psychoacoustic model using ISO MPEG Audio Psychoacoustic Model 1 for Layer 1 [1]

- Calculate the power spectrum of the test signal
- Each input frame x(n) with 512 points (N) is weighted by hanning-window, h(n)[8]

$$h(n) = \frac{\sqrt{\frac{8}{3}}}{2}\left[1 - \cos\left(2\pi\frac{n}{N}\right)\right] \quad \text{.....Equation 1}$$

- So its power spectrum X(k) is computed and normalized to a reference [8]

$$X(k) = 10\log_{10}\left\{\frac{1}{N}\left[\sum_{n=0}^{N-1} x(n)h(n)\exp\left(-j2\pi\frac{nk}{N}\right)\right]^2\right\} \quad \text{.....Equation 2}$$

- Find the tone masker. Once found note the power found one index before and after and combine with the power at k to create a tone maker approximation since the tone may actually be between the frequency

samples. Determining whether a frequency component is a tone requires knowing whether it has been held constant for a period of time, as well as whether it is a sharp peak in the frequency spectrum, which indicates that it is above the ambient noise of the signal. Tonal components are special local maxima of the power spectrum. A local maximum refers to a spectral point satisfying $X(k)>X(k+1)$ and $X(k)\geq X(k-1)$. We take a local maximum as a tonal component if neighbors X(k) and X(k+j) are at greater than or equal to 7dB difference where

$j= -2,+2$ for $(2< k<63)$
$j= -3,-2,+2,+3$ for $(63< k<127)$
$j= -6,\ldots,-2,+2,\ldots,+6$ for $(127< k<250)$ ……..Equation 3

The sound pressure level of every tonal component is calculated

- Find noise-maskers and their locations within each critical band. If a signal is not a tone, it must be noise. Thus, one can take all frequency components that are not part of a tone's neighborhood and treat them like noise. Since humans have difficulty discerning signals within a critical band, the noise found within each of the bands can be combined to form one mask. Thus, the idea is to take all frequency components within a critical band that do not fit within tone neighborhoods, add them together, and place them at the geometric mean location within the critical band. Repeat this for all critical bands.
- If a masker is below the absolute threshold of hearing, it may be discarded. If two maskers are within a critical bandwidth of each other, the weaker of the two may be thrown out as well.
- Calculate the masking threshold of each mask. Sum the masking thresholds to get the overall masking threshold for all frequencies in this signal frame. The maskers which have been determined affect not only the frequencies within a critical band, but also in surrounding bands. The spreading can be described as a function that depends on the maskee location I, the masker location j, the power spectrum Ptm at j, and the difference between the masker and maskee locations in Barks (deltaz=z(i)-z(j)):

$$SF(I,j) = \begin{cases} 17deltaz - 0.4Ptm(j)+11 & -3 <= deltaz < -1 \\ (0.4Ptm(j)+6)deltaz & -1 <= deltaz < 0 \\ -17deltaz & 0 <= deltaz < 1 \\ (0.15Ptm(j)-17)deltaz - 0.15Ptm(j) & 1 <= deltaz < 8 \end{cases}$$

.....….Equation 4

There is a slight difference in the resulting mask that depends on whether the mask is a tone or noise. As a result, the masks can be modeled by the following equations, with the same variables as described above:

For tones: $Ttm(i,j) = Ptm(j) - 0.275z(j) + SF(i,j) - 6.025$ (dB SPL)……...…Equation 5
For noise: $Tnm(i,j) = Pnm(j) - 0.175z(j) + SF(i,j) - 2.025$ (dB SPL)…….……Equation 6

V. Spread Spectrum

As an effective anti-jamming technique, spread spectrum (SS) modulation has become one of the most important approaches to robust watermarking. SS-based watermarking will exhibit low embedding distortions and improved transparency when combined with a perceptual model. Spread spectrum techniques embed a narrow-band signal (the watermark) into a wide-band channel (the audio file). The process can protect watermark privacy by using a secret key to control a pseudorandom sequence generator. Pseudorandom numbers are binary numbers that have specific statistical properties such as correlations which can be used for watermark detection purpose.

Spread spectrum is a means of transmission in which the signal occupies a bandwidth in excess of the minimum necessary to send the information; the band spread is accomplished by means of a code which is independent of the data, and a synchronized reception with the code at the receiver is used for de-spreading and subsequent data recovery [9]. The process of watermark embedding can be viewed as intention jamming of the watermark signal with the music or the audio signal. In this case the signal (watermark) has much less power than the jammer (music). It is one of the problems to be overcome at the receiver end. The following chapter expresses the process of watermark generation in spread spectrum terminology. The approach selected in this algorithm is the Direct Sequence Spreading. The system modulates input the data bit-stream with the help of Pseudorandom Number sequence and modulator signal which is usually a cosine function of time with some centre frequency.

VI. Watermark Embedding and Extraction

A bit stream that represents the watermark information is used to generate a noise-like audio signal using a set of known parameters to control the spreading. These known parameters are the data bits per second *Rd*, *m* the repetition code factor, *N* is the spreading factor. $Rb = Rd*m$ is the coded bits per second, $Tb = 1/Rb$ is

the time of each coded bit, *Rc =N*Rb* is the PN sequence bits per second, *Tc=Tb/N* is the time of each PN bit or "chip".[2]

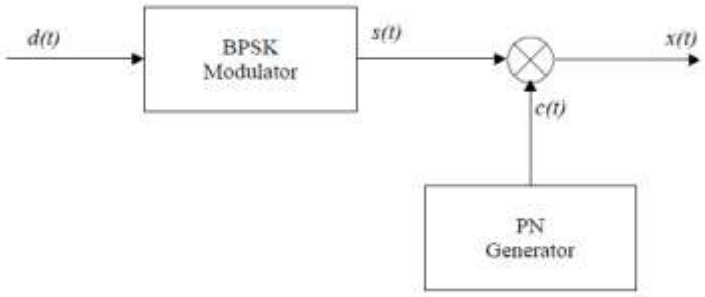

Fig 4 Watermark generation

- Select a data bit sequence some length for eg bit sequence 0110 of length 4 is is selected
- A PN sequence is also selected for spreading
- Multiplying the bit with PN sequence
- Modulating the hopped signal

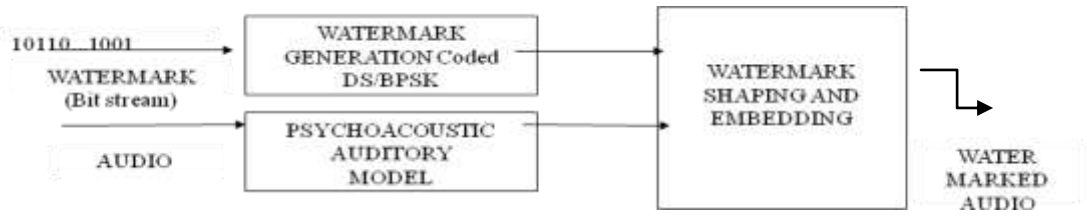

Fig 5 Watermark Embedding

At the same time, the audio (i.e. music/speech) is analyzed using a psychoacoustic auditory model. The final masking threshold information will be used to shape the watermark and embed it into the audio. The output will be a watermarked version of the original audio that can be stored.[2]

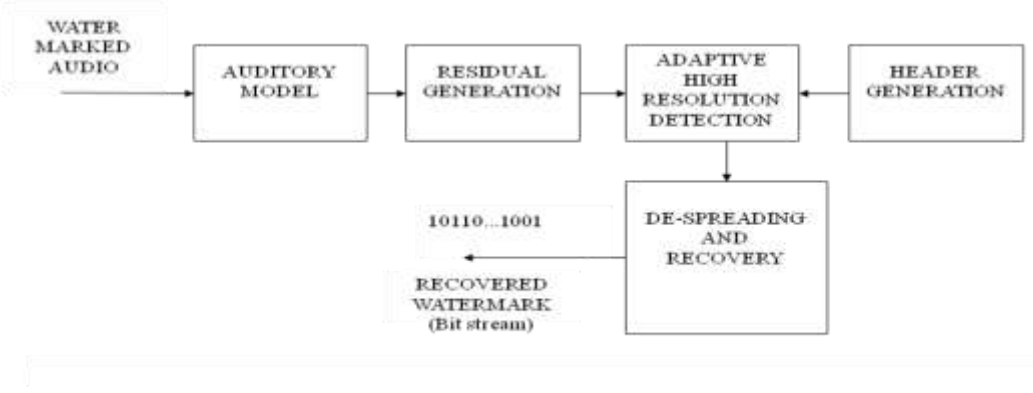

Fig 6 Watermark extraction

VII. CONCLUSION

Due to the great sensitive property of the human auditory system (HAS) compared to the human visual system (HVS), embedding watermarks into audio signal is much more challenging than inserting watermarks into image or video signals. To make the embedding process transparent and provide enjoyable high quality watermarked audio to the listeners, a psychoacoustic model is indispensable to most of the digital audio watermarking systems.

The watermarking algorithm analyzed mixes the psychoacoustic auditory model and the spread spectrum communication technique to achieve its objective. It comprised of two main steps: first, the watermark generation and embedding and second, the watermark recovery. Its robustness can be checked on the basis on

common attacks like cropping, filtering. This robustness checking is a wide area of research which can be explored further.

REFERENCES

Journal Papers:

[1] Pranab Kumar Dhar, Jong-Myon Kim, "*Digital Watermarking Scheme Based on Fast Fourier Transformation for Audio Copyright Protection",* International Journal of Security and Its Applications Vol. 5 No. 2,.pp33- 48, April, 2011

[2] Wahid Barkouti, Lotfi Salhi and Adnan Chérif, "*Digital audio watermarking using Psychoacoustic model and CDMA modulation*", Signal & Image Processing : An International Journal (SIPIJ) Vol.2, No.2, June 2011

[3] Nedeljko Cvejic, "Algorithms for audio watermarking and steganography", University of Oulu., 2004, ISBN 951-42-7383-4.

Books:

[4] Wu M & Liu B, *Multimedia Data Hiding*. Springer Verlag, 2003, New York, NY.

[5] Arnold M, Wolthusen S & Schmucker M, "Techniques and Applications of Digital Watermarking and Content Protection" 2003 Artech House, Norwood, MA.

[6] David Martin Howard, James A. S. Angus, *Acoustics and Psychoacoustics* (4th Edition, illustrated, Focal 2009).

[7] Hugo Fastl, Eberhard Zwicker, *Psychoacoustics: Facts and Models* (Edition 3, Springer 2007).

[8] Oppenheim, A.V., and Schafer, R.W., *Discrete-Time Signal Processing* (Englewood Cliffs, 1989, NJ: Prentice-Hall).

[9] Taub and Schilling, *Principles of Communication System* (Tata McGraw-Hill, New Delhi, 1995).