Voice Athentication System by Embedding Watermarks

J. A. Rios Chavez[1], C. E. Aguilar Meza[2], C. Aquino Ruiz[3]

Telecommunication Academy, National Polytechnic Institute ESIME CU
Santa Ana Avenue #1000 Coyoacan, México Federal District
[1]jriosc0500@ipn.mx; [2]caguilarm@ipn.mx ; [3]caquino@ipn.mx

Abstract-**In current times, extortion and / or telephone frauds are more common than they seem. For example, anyone can call, say that it is the bank and you need personal data. To solve this problem we designed a voice authentication system, which by embedding watermark is detected if the authentic voice of the person who should be. Algorithm was used watermarking embedding through modulation echo since the degree of fragility of this front subjecting digital attacks is considerable and this action is able to authenticate the processed voice.**

Keywords- Voice Authentication; Watermark; Embeeding; Algorithm.

I. INTRODUCTION

Today people want to protect your personal information as bank transactions, account statements, company information to which he holds, and that such information is confidential and very important, just as it is very vulnerable to anyone has access to it thus causing identity theft, data theft, extortion, identity theft, misuse of generating them.

It should be mentioned that the methods of protection of information available nowadays are efficient and keep information secure and ranging from a security code to fingerprint recognition, iris and voice.

Only this time used watermarks for copyright[1] purposes or copyright from being easy to download and share media such as photos, images, pdf files, music and more. This is why it is necessary to design a system for inserting digital watermarks for voice authentication applications to help solve security problems in various fields of information. Watermarks are series of binary digits can be a code word or even an audio file, label act as media files that you want to cover.

As an answer to the problems that exist when you violate copyright in the audio to share music without restriction, and the misuse of identities and confidential data theft arise, in addition to the security schemes, marks water.

Note that in this paper a new search techniques for the protection of digital information that can complement those that already exist today. Although the digital watermarks are little known, there is documentation showing good results of applying them in the various existing digital media, claiming that they are a viable proposal for the protection of information and also have a large field application.

So make an algorithm is proposed to have a certain degree of fragility and strength in order to use the watermark as an integrity checker so you can use it in the solution to security issues such as bank situations, data checks, speaker identification, evidence before courts or confidential data.

The aim is to design a system in which to enter the voice input is processed online and out to obtain an authentication response to identify the speaker.

The digital watermarking is a recent technology that is being used in various fields, with the importance of marking media files to generate a security right has become a study that is so global. The development of algorithms and systems that enable a high degree of inaudibility, a high degree of robustness or otherwise highly fragile digital processing attacks, has led to investigations by various authors obtained different results regarding these characteristics. The techniques of digital watermarking derived from a science called steganography.

The word "stenography"[2] from the Greek, meaning "writing protected" is used to refer to the science of hiding information in other information. This term is also used to refer to the procedures developed as a result of this science. The purpose of steganography is to establish a secret communication channel between two parties so that third portion located between both is unable to detect the existence of such communication.

In contrast to cryptography, focusing on the transformation of the original messages in other unintelligible to unauthorized persons may intercept, the essence lies in steganography devise undetectable methods to hide the messages.

Throughout history men have devised many ways, sometimes curious and surprising, to hide posts or other objects. Thus, there are numerous examples of steganography applications in the military and intelligence fields. Some of these methods though perhaps with greater sophistication have come down to us, so that in daily life is often their implementation. In this connection may be cited the complicated designs that are printed on the banknotes to prevent counterfeiting (some with ink only visible under special lighting) or holograms are recorded with certain object, such as credit cards.

The hiding algorithm for traditional echo delays are inserted to signal the end of embedding ones and zeros. The echo kernel choice determines the fidelity and robustness of watermarking, for this reason, bipolar extension proposes echo core having improved audio fidelity marked without affecting the detection of embedded watermark.

The chosen algorithm is proposed a bilateral extension of time in the echo core[3] to improve robustness against malicious attacks simultaneously increasing the value of cepstrum peak of the core to optimize the detection, this in order to increase the robustness the algorithm to only be fragile processes that will help make the authentication. Time spread echo hiding (TS) is a strong approach compared to conventional echo hiding systems. This system uses the many echoes spreading in time with small amplitude for each segment as in real room (that includes echoes).

The system consists of two processes[4], the embedder that is hiding the information bits in the speech signal and the detector which is responsible for verifying and extract the bits that are hidden in the processed signal in order to carry out the authentication that is, if the detected bits are be the same as the embedded are talking voice actually emitted is correct, otherwise, if the detected bits are different from the embedded voice of another person emitted is being impersonating the original. In some applications require the opposite of robustness, the watermark can be destroyed, as mentioned in the application, one example is that through knowledge and detection of the watermark can be destroyed to know the type of attack was exposed and the technique used to destroy it. The fragility of the watermark should be adjustable to allow certain attacks and resist other.

The article is composed as mentioned: Section 1 shows the introduction, section 2 shows the design of the system, in section 3 can be its implementation, the results of this implementation are reflected in section 4 and finally the sections 5 and 6 show the conclusions and references consulted respectively.

II. DEVELOPMENT

The proposed system is based on the algorithm of embedding watermarks using echo hiding. This technique is based on adding host echo signal. The parameters are handled in this technique the distance of the echo signal with respect to the original (offset) and the amplitude of the echo. We used two different displacement or offset to represent different values.

The echo hiding method[5] inserts a watermark in an audio signal by adding echo to it. Data is hidden varying echo parameters: amplitude and offset (or displacement). The offset between the original signal and the echo can be decreased in order to combine both signals being offset small enough that the human ear cannot perceive but only as a resonance.

In the coding process uses two travel times, one to provide a "binary 1" (offset) and one representing "0 binary" (offset + delta).The encoding process is based on two cores (or kernel) which is inserted as an echo signal depending on the amplitude indicated by the cores as well as the displacement and which can be observed in Figure 1, to be encoded with multi-bit signal is divided into parts. Each of the parts can be encoded with the desired bit considering each part as a separate signal. The encoded signal is a combination of all parts of the signal.

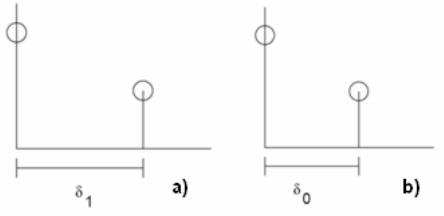

Fig. 1 Echoes for '1' and '0' binary digits

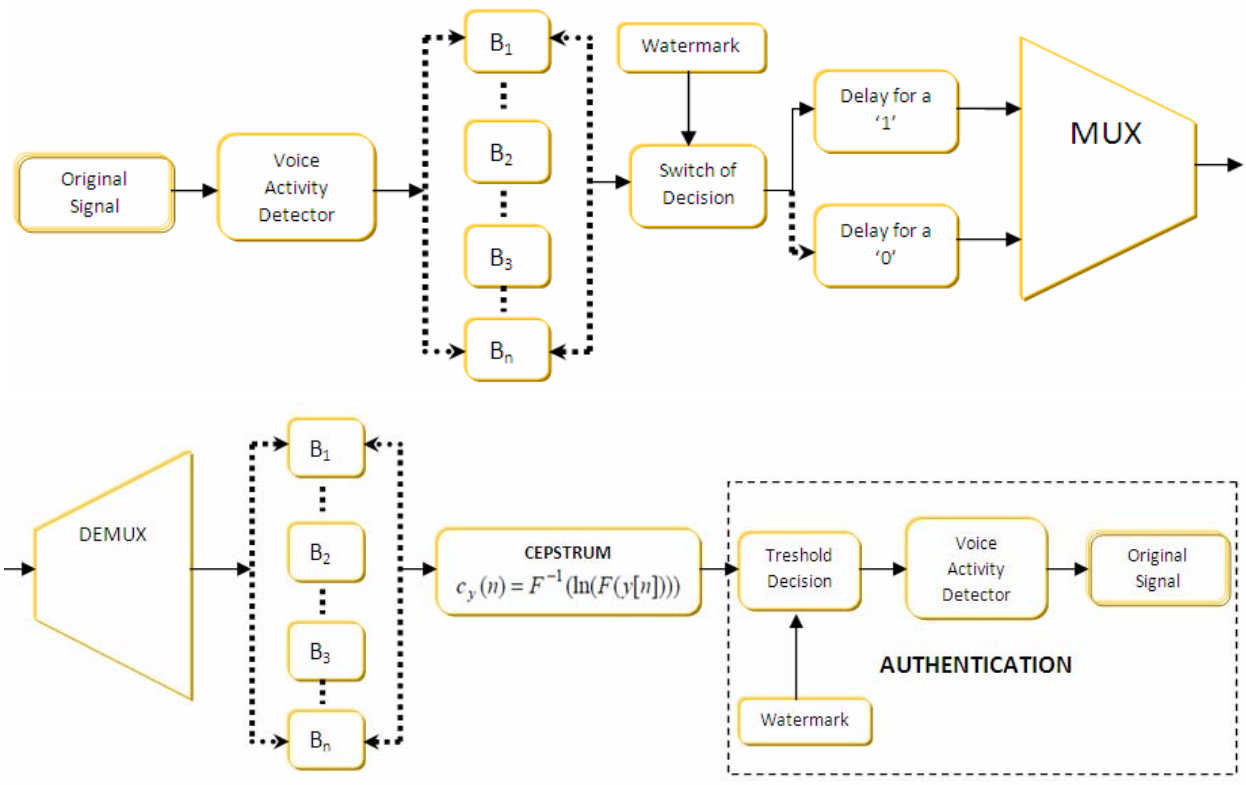

Fig. 2 Embedding and detection system of the watermark

Figure 2 shows the full system proposed for embedding the watermark in the original signal (voice). Having been embedded undergoes a process detcción speech activity, this process detects as being active voice and avoid that the watermark is embedded in and silences probable that this is detrimental to the time that the human ear can distinguish a

coded signal that is not. The next step is cut the total original audio samples "N" blocks "r" samples each block, where N is the number of bits of the string used as watermark. Subsequently, taking into account the embedded watermark, each block is subjected to a switch in the decision that depending on the bit that will be hosted decide if you inject a delay 1 or delay 0. And finally watermarked blocks are concatenated through a multiplexer (multiplexers are circuits that are composed of several inputs and data output and has controls that can select one of the entrances to transmit the result to the only exit. Selected input is given by the combination of zeros and ones that have the controls. They are also known with the name of MUX) with the purpose of regenerating the complete audio samples but this time with the embedded watermark.

The detection process involves subjecting the audio watermarked a demultiplexer (It is the opposite of a multiplexer that consists of one input and multiple outputs, which takes in input a number distributed in several outlets, has the feature of choosing the output, based on input. We call also the name of DEMUX) for the full audio again be severed and then analyze each audio block by calculating the cepstrum (The method involves applying cepstrum two Fourier transforms of the sound. The FFT spectrum is called inverse Fourier Transform and its result is the cepstrum. This allows determining the fundamental frequency from the frequency of the harmonic component of a sound, represented by the peak cepstral upper region of the cepstrum. By the Cepstrum is possible to identify features that allow to evaluate the quality of the voice. The spectral richness can be quantified by means of the amplitude and width corresponding to the pitch cepstral component. If a peak with considerable breadth, is signaling the presence of energy in the harmonic component, remain a feature of great vocal quality voices) and thus verify the echo that is embedded in each block. The cepstrum is controlled by a decision threshold that will indicate if it was embedded in the selected block was a '1 'or a '0' bit, allowing a generation of embedded watermark and so the system will be able making a decision and authenticate the signal obtained with the original.

Mathematically in the process we have the following equation

$$y(n) = x(n) + ax(n - d_1) + ax(n - d_2) \quad (1)$$

Where;

$y(n) = watermarked\ signal$

$x(n) = orignal\ signal$

$a = modulation\ coefficient$

$n = audio\ samples$

$d_1 and\ d_2 = delays\ (echoes)$

The next section shows the implementation of the algorithm, it is noteworthy that the software used was MATLAB ©

III. IMPLEMENTATION

1. The watermarked audio is sectioned in “N” blocks of size “r”

a) Embedder System

Following are the steps for the process of embedding watermark

1. For easy implementation is convenient to transform the audio to WAV format, for this we use the function wavwrite and then we read wavread function in order to generate a one-dimensional vector whose elements are samples
2. The original signal is divided into many segments as bits have the watermark

```
N = length(watermark); %Number of
segments in which the signal is
divided
L = floor(length(audio)/N); % Length
of each segment of the original
signal
```

3. Two delay functions are generated, one for the digit '1 'binary and a delay function for the digit '0' binary

```
for k=1:N
    zero(k,:) =
filter(kernelzero,1,aux(k,:));
    one(k,:) =
filter(kernelone,1,aux(k,:));
end
```

4. Decision signals are created, one for the '1 'and another for the '0', note that these signals are composed of segments of the same length as the original signal

```
for k=1:N
    if marca(k)==0
        for i=1:L
            decision_zero(k,i)=1;
            decision_ones(k,i)=0;
        end
    else
        for i=1:L
            decision_zero(k,i)=0;
            decision_ones(k,i)=1;
        end
    end
end
```

5. And finally step marked blocks are concatenated to generate signal watermarked

b) Detection System

For the process of detecting the watermark remains with the following methodology:

2. For each block is performed cepstrum analysis and we will result in the distance, in number of samples, which

is between the echo signal and the original signal (if there is an echo)

```
zero = abs(cepstrum(delay1));
ones = abs(cepstrum(delay2));
```

3. The bit added to each segment is estimated as one in which the magnitude of the cepstrum is greater

```
if a > b
    watermark_detected(k) = 0;
else
    watermark_detected(k) = 1;
end
```

4. Finally built the watermark bits uniting all estimated and proceeds to do the authentication by the following equation

$$recovery\ rate = \frac{(number\ of\ bits\ correctly\ decoded) * 100}{number\ of\ bits\ placed} \quad (2)$$

Figure 3 shows the physical implementation scheme for the process performed embedding / detection

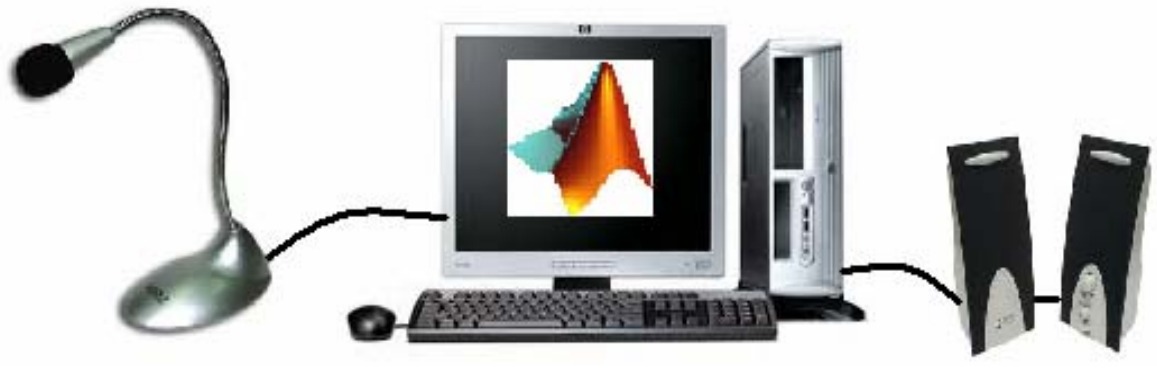

Fig. 3 Physical Implementation Scheme

In section 4 shows the results of the previous implementation

IV. RESULTS

a) Embedder System

The algorithm of embedding the watermark was implemented in a hundred audio clips in order to get an average of all, Figure 5 shows the time comparison of an audio clip and the same but with the watermark embedded.

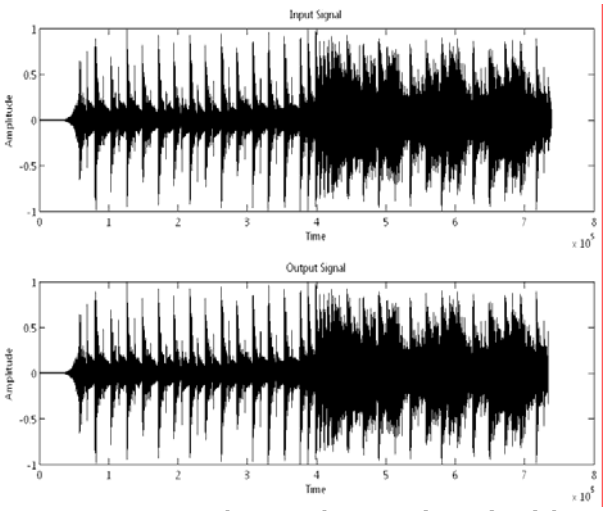

Fig. 5 Comparison between the original signal and the signal watermarked

Then in Figure 6 shows a comparison between the same file (original and marking) but in the frequency domain, ie. the frequency spectrogram

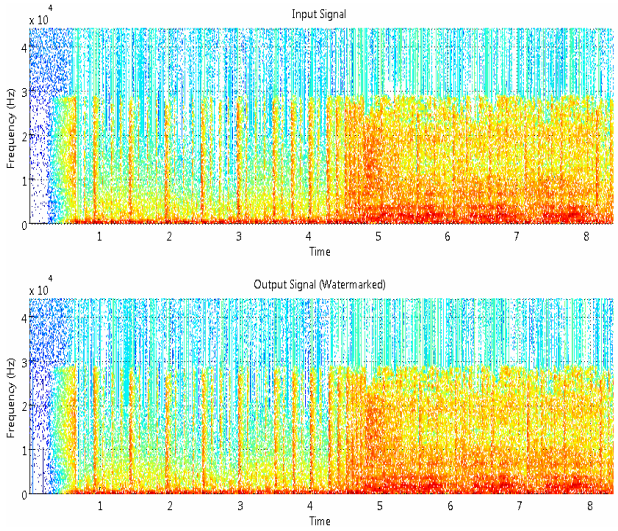

Fig. 6 Comparison of signals in the frequency domain

b) Detection System

For the detection process performed cepstrum calculation and Figure 7 we see an echo to '1 'and an echo embedded to a '0' embedded.

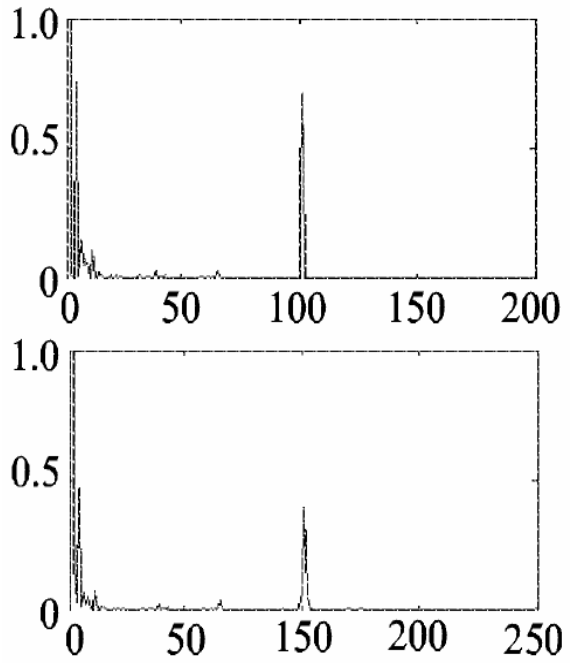

Fig. 7 Echo for '1' (top), Echo for '0' (bottom)

In deciding on the number of delayed samples was based on post-masking technique that tells us that for a delay is inaudible to human auditory system (essential feature of the watermarking algorithms), it must be no more than 50ms, otherwise the user can distinguish the original file and marking.

To verify the degree of fragility of the selected algorithm used the audiostirmark database[6], software that allows us to simulate attacks or digital signal processing marked in order to show whether the algorithm is fragile or robust to the presence of the same. The following table shows the results obtained.

Table 1. AudioStirmark Database for Echo Hiding Algorithm

Echo Hiding	
Robustness	Fragile
Nothing	AWGN
Amplify	Remuestreo Compressor
ExtraStereo	AddDynNoise
TimeStrech	BitChanger
	CopySample
	CutSamples
	LSBZero
	RC_HighPass
	RC_LowPass
	ReplaceSamples
	Invert
	Zeros (ZeroCross, Zerolength1 y 2, ZeroRemove)

V. CONCLUSIONS

The algorithm is a significant vulnerability, so we can conclude that an algorithm is fragile and viable for the intended application, authentication signal. Contrary to the provisions of the algorithm 2 is benefited only a slight increase in strength, but this is almost imperceptible and try to increase this parameter is high computational complexity that attracts a variation resulting in the extraction process.

VI. REFERENCES

[1] MASOUD NOSRATI, ET. AL., *AN INTRODUCTION TO STEGANOGRAPHY METHODS*, VOL. 1 ISSUE 3, WORLD APPLIED PROGRAMMING, AUGUST 2011.

[2] W. BENDER, ET. AL, *TECHNIQUES FOR DATA HIDING*, IBM SYSTEMS JOURNALS, INFORMATION TECHNOLOGY, VOL. 35,NO. 5, NOV. 1996.

[3] NEIL JENKINS, ET. AL., *STEGANOGRAPHY IN AUDIO*, ANAIS DO IX SIMPÓSIO BRASILEIRO EM SEGURANÇA DA INFORMAÇÃO E DE SISTEMAS COMPUTACIONAIS, 2007

[4] J. A. RIOS, ET. AL, *PROGRAMMABLE LOGIC IMPLEMENTATION OF ECHO HIDING FOR AUDIO WATERMARKING*, VOL. 47 ISSUE 2., JOURNAL OF THEORETICAL AND APPLIED INFORMATION TECHNOLOGY (JATIT), ISLAMABAD PAKISTAN, 2012.

[5] J. A. RIOS, ET. AL, *DIGITAL AUDIO WATERMARKING BY DILATED AUDIO SAMPLES*, INTERNATIONAL JOURNAL OF COMPUTER SCIENCE, INFORMATION TECHNOLOGY, & SECURITY, VOL. 2,NO. 5, NOV. 2012.

[6] AUDIOSTRIMARK DATABASE

AUTHOR PROFILES:

Master´s Degree Juan Antonio Rios Chavez. Born in Mexico DF on May 7, 1988, holds a degree in Communications and Electronics from the National Polytechnic Institute in 2009 and earned his Master of Science in Engineering with specialization in Microelectronics Digital Signal Processing by the National Polytechnic Institute in 2011, his research interests include digital audio processing for robotic applications and the hardware implementation of general purpose DSP (Digital Signal Processor) and FPGA (Field Programmable Gate Array)

Ing. Celedonio Enrique Meza Aguilar Born in México on March 3, 1959 and holds a degree in Communications and Electronics from the National Polytechnic Institute, is currently research professor and development areas are the processing and handling, coding and compression, safety and efficient transmission.

Ing. Carlos Aquino Ruiz Born in México on April 2, 1973 is in Communications and Electronics Engineering from the National Polytechnic Institute, a research professor and areas of development are applied computing, data networking and security.