Watermarking of speech signals in the time-frequency domain

Mahmood Al-khassaweneh
Computer Engineering Department
Yarmouk University, Irbid 21163, Jordan
e-mail: khassaweneh@yu.edu.jo

Hussein Al-Zoubi
Computer Engineering Department
Yarmouk University, Irbid 21163, Jordan
e-mail: halzoubi@yu.edu.jo

Selin Aviyente
Department of Electrical and Computer Engineering
Michigan State University, East Lansing, MI 48823
e-mail: aviyente@egr.msu.edu

Abstract—In this paper, we introduce a new multi-bit watermarking method for speech signals in the joint time-frequency domain. The Wigner distribution for the speech signal is computed and the time-frequency cells in the Wigner domain that are robust to attacks are used for watermark embedding. The error introduced by the inversion of the watermarked distribution is quantized and used as a key for watermark extraction. The performance of the proposed watermarking algorithm under possible attacks is illustrated.

I. INTRODUCTION

With the improved technology and the rapid expansion of the Internet, the availability and the ease of redistribution of digital multimedia data such as audio, images and videos to the public has increased. In order to protect the copy rights, researcher start looking to means that provide such protection. Watermarking techniques were developed to provide copyright protection for digital data. In this paper, we focus on speech watermarking for protecting the ownership of digital speech signals. This application requires the watermark to be imperceptible such that the Human Hearing System (HHS) can not distinguish between the original and the watermarked data.

The idea of watermarking in the joint time-frequency domain is relatively new. Most work in this area concentrates on using Wigner distribution as the transform domain. In [1], the authors used a two-dimensional chirp signal with a variable spatial frequency to carry the watermark. The watermark is characterized by a linear frequency change and can be detected by using 2-D space/spatial-frequency distributions. The projections of the 2-D Wigner distribution and the 2-D Radon-Wigner distribution, are used in order to emphasize the watermark detection process. Although the authors where able to detect the watermark efficiently under most attacks, there was no analysis for the probability of error.

In [2], the Wigner distribution is used for watermarking. The watermark is embedded in a subset of the transformed cells. These cells are selected such that the watermark will survive the JPEG compression. Since the resultant watermarked distribution is not a valid Wigner distribution, the time signal that has the closest distribution in the mean square error sense is found. Although, this algorithm detects the presence of the watermark under JPEG attacks, there were no experimental results for other types of attacks.

In [3], a fragile image watermarking method is presented. The watermark is an FM modulated signal which is embedded in the diagonal elements of the image. Particular features of this signal in the time-frequency domain are used to identify the watermark. The Wigner distribution is used to analyze the content of the extracted watermark not for the watermarking itself.

Most of the work in watermarking, in general, is focused on images and few work has been done for audio signals [4], [5]. In this paper, a new time-frequency based watermarking method for speech signals will be introduced using Wigner distribution.
II. BACKGROUND

For a 1-D discrete signal, $s(n)$, the discrete Wigner distribution is given by [6],

$$WD(n, \omega) = 2 \sum_{m} s(n+m)s^*(n-m)e^{-j2\omega m}.$$  (1)

where $n$ and $\omega = 2\pi k/N$ are the discrete time and the frequency variables, respectively.

The Wigner distribution has some important properties that make it a good choice in watermarking applications; it is always real, satisfies the marginals, invertible and symmetric. Invertibility is especially important in watermarking applications where detecting or extracting the watermark is one of the primary goals.

In the case where we have positive and real valued signal, the Wigner distribution is symmetric and the original signal, $s(n)$, can be retrieved completely from its Wigner distribution as,

$$s(n) = \sqrt{\sum_{\omega} WD(n, \omega)}.$$  (2)

Equation (2) implies that for a positive real valued signal, the original signal can be retrieved from its Wigner distribution by taking the inverse Fourier transform of the Wigner distribution evaluated at $m = 0$ and taking the square root of each element in this row. This result will simplify the embedding and the detection of the watermark.

III. WATERMARK EMBEDDING

The watermark embedding algorithm can be summarized as follows:

1) Find the absolute value of the speech signal and save the sign in a key $K_s$. Randomly choose $N$ points from the signal and place them in a vector $E$ and save the locations of these points in a key $K_p$.

2) Compute the Wigner distribution for the vector $E$,

$$WD_E(x, \omega_x) = 2 \sum_{m} E(x+m)E(x-m)e^{-j2\omega_x m}.$$  (3)

3) Select $N$ time-frequency points form $WD_E(x, \omega_x)$ and use them as a mask for watermark embedding. The location for these points is stored in a key $K_c$. The mask $A_E(x, \omega_x)$ equals to zero if the corresponding $WD_E(x, \omega_x)$ is not used for watermark embedding and is equal to $WD_E(x, \omega_x)$ otherwise.

4) Embed the watermark $W$ into the time-frequency points found in the last step,

$$WD_E(x, \omega_x) = WD_E(x, \omega_x) + A_E(x, \omega_x) \circ W_m, $$  (4)

where $W_m$ is the multi-bit watermark re-ordered in a matrix form such that it has a value of zero if the corresponding time-frequency point is not chosen for watermarking and a value of $W(t)$ if the corresponding time-frequency point is used for watermark embedding. $A_E(x, \omega_x) \circ W_m$ is an element by element multiplication of the two matrices.

5) Take the inverse Fourier transform and use $K_s$ to obtain the watermarked signal,

$$\hat{E}(x) = \sqrt{\sum_{\omega_x} WD_E(x, \omega_x)}.$$  (5)

6) Compute the difference between the Wigner distribution for $\hat{E}(x)$, $WD_E(x, \omega_x)$, and $WD_E(x, \omega_x)$. Save the difference between these two distributions at the locations that correspond to the key $K_c$ in a key $K_d$. This be can be encrypted and sent to the receiver for watermark extraction.

7) Using the key $K_p$ from step 1, modify the signal as,

$$S(x, y) = \hat{E}(x, y) \quad \text{if} \quad (x, y) \in K_p$$

$$S(x, y) = S(x, y) \quad \text{otherwise}.$$  (6)

IV. WATERMARK DETECTION

In many data hiding applications, it is important to detect or retrieve the watermark even after the watermarked image is attacked.

The extraction algorithm, assuming we have access to the original signal, can be summarized as follows:

1) Find the Wigner distributions for the absolute value of the points at locations $K_p$ for the original and received signals. Name them $WD_R(x, \omega_x)$ and $WD_R(x, \omega_x)$ respectively.

2) Find the watermarked cells from $WD_R(x, \omega_x)$ using the key $K_c$ and add to them the difference in key $K_d$.

3) Extract the watermark matrix according to,

$$\hat{W}_m = \text{sgn} \left( \frac{WD_R(x, \omega_x) - WD_E(x, \omega_x)}{A_E(x, \omega_x)} \right),$$  (7)

where the division in (7) is element by element division for the non-zero elements in $A_E(x, \omega_x)$. If the element is zero in $A_E(x, \omega_x)$, the $W_m$ is set to zero.

4) Get the watermark $W$ from $\hat{W}_m$.

The extracted watermark is compared with the original one to determine the correlation between them,
\[ \rho(\hat{W}, W) = \frac{\sum \hat{W}(n)W(n)}{\sqrt{\sum \hat{W}^2} \sqrt{\sum W^2}} = \frac{\sum \hat{W}(n)W(n)}{N}. \]  

V. Simulation Results

The watermark embedding algorithm proposed in this paper has been applied to a speech signal of the word "one". The multi-bit watermark is a sequence of 1 and \(-1\). The performance of the proposed algorithm has been tested under different types of attacks including additive white Gaussian noise (AWGN), Median filtering, and cropping attacks. The detector for the algorithm described by (8) is tested for a series of watermarks. Figs. 1 and 2 show the original and the watermarked speech signals respectively. There are no perceptible differences between the original and the watermarked data.

![Fig. 1. The original speech signal.](image1)

![Fig. 2. The watermarked signal (PSNR=35.21db).](image2)

Fig. 3 shows that the normalized correlation function under AWGN attack, for the extracted watermark and a sequence of possible randomly generated watermarks, reaches its maximum value when the embedded watermark and the tested one are the same and is close to zero for all other watermarks.

![Fig. 3. The normalized correlation detector response for the proposed embedding method under AWGN attack, a. AWGN (PSNR=12db), b. AWGN (PSNR=9db) c. AWGN (PSNR=6.8db).](image3)

The algorithm has been tested under cropping as well, where a random number of samples are removed. Fig. 4 shows the cropped signal with PSNR=14.1db, while Fig. 4 shows that the maximum correlation occurs with the real watermark. In Fig. 6, a median filter of size \(4 \times 1\) has been applied and as its clear form the figure, the real watermark is detectable.

![Fig. 4. The cropped signal (PSNR=14.1db).](image4)
VI. CONCLUSION

In this paper, we introduced a new watermarking algorithm for speech signals based on using Wigner distribution. It has been shown that for positive and real signals, the signal can be retrieved from its Wigner distribution without error. The Wigner distribution of samples of the speech signals is found and the watermark is embedded into selected time-frequency locations. Although the watermarked distribution may not be a valid Wigner distribution, a simple inversion method is carried out to find the synthesized signal. The error introduced by this simplification is analyzed and used as a key for watermark extraction. The proposed method is shown to be transparent and robust under different types of attacks.

REFERENCES
