An Approach to Digital Watermarking of Speech Signals in the Time-Frequency Domain

Srdjan Stanković, Irena Orović, Nikola Žarić, Cornel Ioana[1]
Faculty of Electrical Engineering, University of Montenegro, (srdjan, irenao, zaric)@cg.ac.yu
[1] ENSIETA, Rue Francois Verny 2, Brest, France, ioanaco@ensieta.fr
Corresponding author e-mail: srdjan@cg.ac.yu

Abstract *A watermarking scheme in the time-frequency domain is proposed. The region for watermark embedding is selected by using the S-method based on the time-frequency representation. Time-varying filter scheme is used to map the watermark sequence from the time-frequency domain to the time domain. Watermark detection is performed in the time-frequency domain. Theory is illustrated by the example.*

Keywords – *Digital watermarking, Speech signals,Time-frequency analysis.*

1. INTRODUCTION

The rapid development of multimedia data applications creates a demand to protect their contents. Digital watermarking is a technique that provides data copyright protection and the ownership proof. These watermarking schemes have to satisfy two important, but conflicting requirements: imperceptibility and robustness. Nowadays, there is a large amount of proposed watermarking techniques for audio and speech signals. Watermark can be inserted in the time domain or in the transform domain. One of the common algorithms for speech watermarking is based on spread-spectrum method [1], [2]. According to H.J.Kim [1], it implies additive embedding of pseudo-random sequence as watermark, while detection is performed by using a linear correlation. On the other hand, self-marking methods for watermark embedding have also been proposed in [3], [4], and [5]. As a specific scheme that belongs to this category, Mansour and Tewfik [4], [5] have used the time-scale method.

In this paper a time-frequency based approach for speech watermark embedding and detection is introduced. The S-method, as a time-frequency distribution, is used to determine the region in which watermark should be embedded. In order to create pseudo-random sequence that corresponds to the specified time-frequency region, the time-varying filtering procedure is used. Detection is performed by using correlation in time-frequency domain.

2. THEORETICAL BACKGROUND

The proposed watermarking scheme is based on the time-varying filtering technique. Since, time-frequency representations play a crucial role in this filtering technique, they will be considered in the next subsection.

2.1. Time-frequency representation of speech signals

Speech signals are of a highly nonstationary and multicomponent nature. Time-frequency distributions have been used in order to analyze frequency components of nonstationary signals in time. The spectrogram (square module of the Short Time Fourier Transform - STFT) is the oldest and commonly used time-frequency distribution. The spectrogram is defined by:

$$SPEC(t,\omega) = \left|STFT(t,\omega)\right|^2 \qquad (1)$$

where,

$$STFT(t,\omega) = \int_{-\infty}^{\infty} f(t+\tau)w(\tau)e^{-j\omega\tau}d\tau, \qquad (2)$$

and $w(t)$ is the lag window. Note that there is a trade-off between the time and frequency resolution, meaning that high resolution in time and in frequency domain, cannot be achieved simultaneously.

Quadratic time-frequency distributions are introduced to provide better resolution in the time-frequency plane. The best auto-terms concentration is obtained by using the Wigner distribution. At the same time, the Wigner distribution of multicomponent signals produces a large amount of cross-terms. In order to preserve auto-terms concentration as in the Wigner distribution, and to reduce the presence of cross-terms, the S-method (SM) has been introduced [6].
The SM is defined by:

$$SM(t,\omega) = \int_{-\infty}^{\infty} P(\theta)STFT(t,\omega+\theta)STFT^*(t,\omega-\theta)d\theta \qquad (3)$$

where $P(\theta)$ represents a finite frequence domain window function.
Discrete version of the SM is given by:

$$SM(n,k)=\sum_{l=-L}^{L}P(l)STFT(n,k+l)STFT^{*}(n,k-l), \quad (4)$$

where n and k are the discrete time and frequency variables, $P(l)$ is the window of the length 2L+1.

By taking the rectangular window, the discrete SM can be written as:

$$SM(n,k)=\left|STFT(n,k)\right|^{2}+ +2\,\mathrm{Re}\left\{\sum_{l=1}^{L}STFT(n,k+l)STFT^{*}(n,k-l)\right\}. \quad (5)$$

2.2. Time-varying filtering

Since the signal cannot be recovered directly from its time-frequency domain, time-varying filtering will be used for signal reconstruction from its time frequency representation.

Time-varying filtering procedure is introduced for noisy signal [7], [8], [9]:

$$Hx(t)=\int_{-\infty}^{\infty}h(t+\frac{\tau}{2},t-\frac{\tau}{2})x(t+\tau)d\tau, \quad (6)$$

where $x(t)$ is the signal, while $h(t+\tau/2,t-\tau/2)$ represents the impuls response of the time-varying filter.

The pseudo form of the relation (6), which is more appropriate for numerical realizations can be written as:

$$Hx(t)=\int_{-\infty}^{\infty}h(t+\frac{\tau}{2},t-\frac{\tau}{2})w(\tau)x(t+\tau)d\tau, \quad (7)$$

where $w(\tau)$ is a lag window.

The support function, that plays the main role in filtering, can be defined as:

$$L_{\cdot}(t,\omega)=\int_{-\infty}^{\infty}h(t+\frac{\tau}{2},t-\frac{\tau}{2})e^{-j\omega\tau}d\tau. \quad (8)$$

Having in mind that $w(\tau)$ does not affect the output signal if $w(0)=1$ [8], the signal can be retrived by using the support function as follows [7]:

$$Hx(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}L_{H}(t,\omega)STFT(t,\omega)d\omega \quad (9)$$

The discrete form of the relation (9) can be written as:

$$Hx(n)=\frac{1}{N}\sum_{k=-N/2}^{N/2}L_{H}(n,k)STFT(n,k) \quad (10)$$

where N is the length of the signal.

Note that if the suport function L_H is obtained by using high resolution time-frequency representation, the original signal at the filter output will be obtained, since the lag window in the STFT does not influence the output signal [8].

3. WATERMARKING PROCEDURE USING TIME-FREQUENCY REPRESENTATION

In this section a new speech watermarking method is proposed. It provides imperceptibility and successful detection.

By using this method a pseudo-random sequence with specific time-frequency characteristics will be created in the time domain. Watermark embedding procedure will be done by:

$$x_{w}(t)=x(t)+\alpha * w_{key}(t), \quad (11)$$

where $w_{key}(t)$ is the watermark sequence, α is the parameter that controls the amplitude of watermark and x_w is the watermarked signal.

3.1. Watermark sequence derivation

In order to create an appropriate watermark sequence, let us start with the pseudo-random sequence p. The sequence p represents the input signal for the time-varying filtering procedure.

Assume that the SM of the speech signal lies in a region R^2 in the time-frequency plane [8]. Thus, the support function that will provide the retrieval of the original signal components is given by:

$$L_{H}(t,\omega)=\begin{cases}1, & for \quad (t,\omega)\in R^{2}\\ 0, & for \quad (t,\omega)\notin R^{2}\end{cases} \quad (12)$$

Therefore, function L_H produces the information about localization of speech signal components in the time-frequency plane. Thus, the main idea of this work is to use modified support function of the signal to create a watermark sequence with specified time-frequency caracteristics.

If we consider the region D in the time-frequency plane that is located between time instants t_1 and t_2, and frequencies ω_1 and ω_2:

$$\mathrm{D}=\left\{(\mathrm{t},\omega):\mathrm{t}_{1}<\mathrm{t}<\mathrm{t}_{2},\omega_{1}<\omega<\omega_{2}\right\}, \quad (13)$$

then the modified function L_H can be written as:

$$L_M(t,\omega)=\begin{cases}1, & \text{for} \quad (t,\omega)\in D\\ 0, & \text{for} \quad (t,\omega)\notin D\end{cases} \qquad (14)$$

The region $D\subset R^2$ is illustrated in Fig.1.

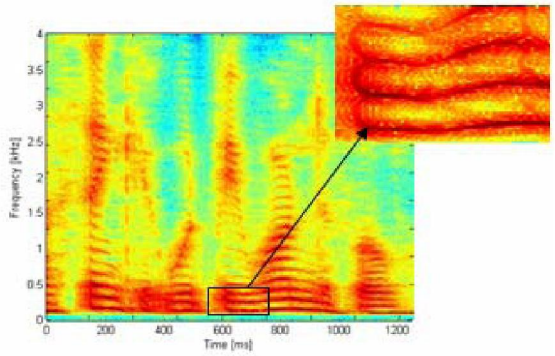

Fig.1. An illustration of the region D specification

Additionaly, if we only take into account the SM signal components above some threshold value, the modified L_H function becomes:

$$L_M(t,\omega)=\begin{cases}1, & \text{for} \quad (t,\omega)\in D \quad \text{and} \quad SM_x(t,\omega)>\beta\\ 0, & \text{for} \quad (t,\omega)\notin D \quad \text{and} \quad SM_x(t,\omega)\le\beta\end{cases} \qquad (15)$$

where $SM_x(t,\omega)$ represents the SM of speech signal x, and β is the threshold value. The obtained L_M function contains the information about signal components in the specified region of time-frequency plane. In order to use low frequency components, the threshold β can be determined as:

$$\beta \ge \overline{SM}_x(t,\omega), \qquad (16)$$

where $\overline{SM}_X(t,\omega)$ is average value of speech signal SM.

Furthermore, the L_M function will be used for deriving the filtered version of the sequence p, that will be present only within specified region where the strong signal components exist. With the right choice of the region, we are able to avoid watermarking of sensitive parts of speech signal, such as speech pauses and high frequencies.

The watermark sequence, obtained by using fuction L_M, is given by:

$$w_{key}(n)=\frac{1}{N}\sum_{k=-N/2}^{N/2} L_M(n,k)*STFT_p(n,k) \qquad (17)$$

where $STFT_p(n,k)$ is the Discrete STFT of sequence p, while N is the length of the sequence.

3.2. Watermark detection

The successful watermark detection is obtained by using the correlation function in the time-frequency domain. Namely, the watermark signal $w_{key}(t)$, that represents the right key for detection, as well as watermarked signal $x_w(t)$, should be transformed to the time-frequency domain. At the same time, by taking another pseudo-random sequence q generated in the same manner, wrong key $w_{wrong}(t)$ is obtained. The next step would be to extract only the region D from the time-frequency plane. Having in mind that the watermarked signal already contains the right key, the detector response for any wrong trial i=1,2,3, etc., should satisfy the relation:

$$\sum_D STFT_{x_w}(t,\omega)*STFT_{w_{key}}(t,\omega) > \\ > \sum_D STFT_{x_w}(t,\omega)*STFT_{wwrong_i}(t,\omega), \qquad (18)$$

where $STFTx_w(t,\omega)$, $STFTw_{key}(t,\omega)$ and $STFTw_{wrongi}(t,\omega)$ represent the STFTs of watermarked signal, right key and wrong key, respectively.

4. EXAMPLE

In the implementation we consider the speech signal with maximal frequency f_{max}=4kHz. The Gaussian white noise is used as the sequence p. The STFT was calculated using rectangular window with 256 samples for time-varying filtering, and 1024 samples for preview. Zero padding up to 1024 samples was carried out, and the parameter L=5 is used in SM calculation. The region D is selected to cover first three formants of the speech signal, Fig.2.

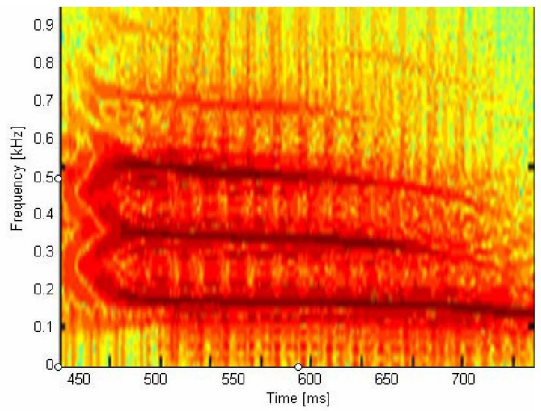

Fig.2. Region D of analyzed speech signal

The modified support function L_M is created using threshold value β=10*($\overline{SM}_x(t,\omega)$). The L_M function is shown in Fig.3.

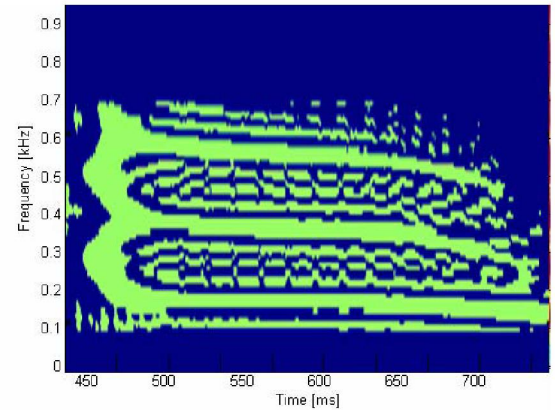

Fig.3. Modified support function L_M

The detection procedure is performed by using 100 trials with wrong keys, created in the same manner as the right key. We present the results of watermark detection in Fig.4. It is obvious that the watermark detector response has significantly higher value in the case of right key when compared with the responses of wrong keys.

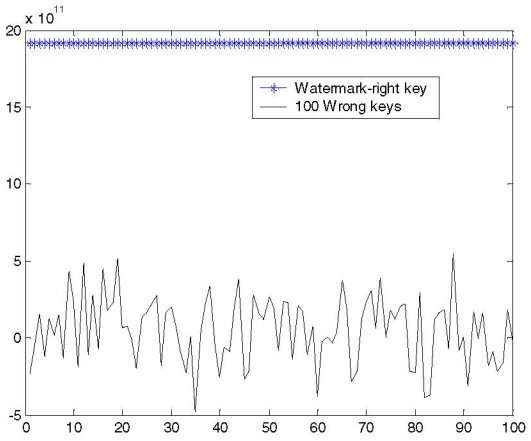

Fig.4. Watermark detector responses

5. CONCLUSION

An approach for watermarking procedure of speech signals by using time-varying filtering is presented. It has been shown that this approach provides modeling of pseudo-random sequence in time-frequency domain. The presented results demonstrate that the procedure assures promising watermark detection.

ACKNOWLEDGEMENT

This work is supported by the joint Montenegrin-French "Pelikan" project. A part of this paper is supported by the project of Ministry for Science and Education of Montenegro.

REFERENCES

[1] H.J.Kim, "Audio Watermarking Techniques", *Pacific Rim Workshop on Digital Steganography*, Kyushu Institute of Technology, Kitakyushu, Japan, July, 2003.

[2] I.J.Cox, J.Kolian, F.T.Leigton, T.Shamoon, "Secure Spectrum Watermarking for Multimedia"(1996), IEEE *Trans.Image Processing*, Vol.6,pp.1673-1687.

[3] C.P.Wu,P.C.Su, C.C.J.Kuo, "Robust and Efficient Digital Audio Watermarking Using Audio Content Analysis"(2000), *Security and Watermarking of Multimedia Contents,* SPIE, Vol. 3971, pp. 382-392.

[4] M.F.Mansour, A.H.Tewfik, "Audio Watermarking by Time-Scale Modification", *International Conference on Acoustic, Speech, and Signal Processing,* 2001, Vol. 3, pp. 1353-1356.

[5] M.F.Mansour, A.H.Tewfik, "Time-Scale Invariant Audio Data Embedding", *International Conference on Multimedia and Expo,* 2001.

[6] LJ.Stanković, "A method for Time-Frequency Signal Analysis", IEEE *Transaction on Signal Processing,* Vol.42, No.1, January 1994.

[7] S.Stanković, "About Time-Variant Filtering of Speech Signals with Time-Frequency Distributions for Hands-Free Telephone Systems", *Signal processing,* Vol.80, No.9, 2000.

[8] LJ.Stanković, "On the Time-Trequency Analysis Based Filtering", *Annales des telecommunications,* Vol.54, No.5/6, May/June, 2000.

[9] G.Matz, F.Hlawatsch, W.Kozek, "Generalized Evolutionary Spectral Analysis and the Weyl Spectrum of Nonstationary Random Processes", IEEE *Transaction on Signal Processing,* Vol.45, No.6, Jun 1997, pp.1520-1534.