Speech Signals Protection via Logo Watermarking based on the Time-Frequency Analysis

Irena Orović, Predrag Zogović, Nikola Žarić, Srdjan Stanković

Abstract— **A procedure for logo watermarking of speech signals is proposed. In order to take full advantage of both time and frequency domain, we have used joint time-frequency representation of speech signal, which provides a more comprehensive information compared to each domain individually. Time-frequency based selection of the Discrete Fourier Transform (DFT) coefficients is used for logo embedding. The proposed coefficients selection provides both the imperceptibility and robustness requirements. The binary logo is embedded by using the concept of bit-planes modification. The logo extraction procedure is provided, as well. The robustness of the proposed procedure is tested under various attacks such as compression, noise addition, filtering, re-sampling, requantization, time stretch, pitch scaling, etc.**

I. Introduction

Unauthorized usage and distribution of digital multimedia data creates a demand for their protection. The digital watermarking techniques are used for these purposes. The digital watermarking is based on embedding a signal (called watermark) in the original data [1], [2]. This procedure can be employed in various real applications. As an interesting example in the case of speech signals, the credibility of user's voice, in different on-line services (e.g. banking), could be proven by extracting user's signature. Also, as a kind of steganography, a secret message in the form of meaningful mark can be transmitted within the host signal. Watermarking procedure can be realized in the time domain [3], [4] or in the transform domain [5], [6]. However, it should not degrade the quality of the host signal, i.e. it should be imperceptible.

Various watermarking techniques for speech and audio signals have been developed. In [7], the pseudo-random sequence is perceptually shaped and embedded directly into the segments of an audio signal. The modulated complex lapped transform (MCLT) has been used in [6] to perform robust spread-spectrum watermarking method. Here, the shaped spread-spectrum sequence has been usually considered for speech and audio watermarking, since they are more robust. Generally, this sequence has to be quite large to provide low probability of detection error. However, it significantly increases the detection complexity. In order to provide more convincing, faster and easier extraction of embedded data than in the case of numerical mark, a meaningful binary logo image has been used [8]-[14]. Namely, the visual perception of meaningful mark improves reliability of signal identification especially for the proof of the ownership in front of non-technical arbitrators [12], [13]. Thus, the logo watermark can be considered as "virtually" more robust [12]. Most of the existing logo watermarking procedures are related to images, but only a few of them deal with audio signals. An interesting logo watermarking algorithm for audio signals based on the usage of support vector machine is given in [8]. Based on the human auditory model, the advantages of the wavelet transform (WT) and the complex cepstrum transform (CCT) for audio watermarking are explored in [9]. Also, a procedure based on the usage of neural networks has been proposed in [10]. The most of the mentioned procedures exploit characteristics of the HAS via frequency-masking to provide the imperceptible watermark.

The usually applied frequency domain masking cannot always assure inaudible water-

Annals of Telecommunication, Vol. 63, No. 5-6, May 2008.

mark. It is based on the fixed length Fourier transform that does not provide exact information about time localization, and thus the watermark will be spread over the entire analysis window. If the signal energy is concentrated in the interval shorter than the length of analysis window, the watermark is not masked outside this interval. In other words, the watermark can be present in the time-frequency regions where signal components do not exist. It causes watermark audibility.

In this paper, a time-frequency based approach for speech watermarking is proposed. Time-frequency analysis enables both temporal and frequency information about signal behavior. Thus, the time-frequency representation is used to identify robust components that have good masking properties. Also, this function enables avoiding noisy areas and sensitive transients of speech signal. According to the time-frequency analysis, the set of suitable DFT coefficients are selected for embedding a meaningful binary logo image. It will be shown that coefficients selection based on the time-frequency analysis provides flexibility and robustness even for de-synchronization attacks such as time stretch [14]. The concept of bit-planes is used for logo insertion [15]. The S-method is used for time-frequency representation of speech signals. A simple, fast and blind procedure for logo extraction is also provided. Apart from the ability to use human visual system to verify the logo extraction results, the bit error rate (BER) as a statistical measure of extraction efficiency is calculated. The robustness has been tested for various common attacks, such as compression, noise addition, filtering, re-sampling, re-quantization, echo, time stretch, pitch scaling, etc. Additionally, through an example, we show that the proposed method can be realized even as fragile or semi-fragile watermarking approach, which can be a topic of future work.

II. Theoretical Background

The most significant formants of speech signal are considered in order to ensure the complete inaudibility of the logo watermark. These formants are of high energy and small changes will not influence signal quality. Since the formants are time-variant frequency components, they can be adequately analyzed only in the joint time-frequency domain. Otherwise, by using either time or frequency domain separately, the precise location of formants cannot be determined. In watermarking, this causes the presence of watermark in the time-frequency regions where speech components do not exist. Consequently, it can result in watermark audibility. Since the time-frequency analysis is used for formants selection, we will give a brief overview of commonly used time-frequency distributions.

The spectrogram is the simplest and commonly used time-frequency representation. It is defined as a square module of the short time Fourier Transform:

$$SPEC(n,k) = |STFT(n,k)|^2 = \\ = \left| \sum_{m=-N/2}^{N/2-1} f(n+m)w(m)e^{-j2\pi mk/N} \right|^2, \quad (1)$$

where f is a signal, w is a window function, n and k are the discrete time and frequency variables, while N is the signal length. In order to overcome the existing trade-off between time and frequency resolution in the case of spectrogram, quadratic time-frequency distributions were introduced. The highly concentrated auto-terms could be obtained by using the Wigner distribution:

$$WD(n,k) = \\ = 2 \sum_{m=-N/2}^{N/2} f(n+m)f^*(n-m)e^{-j2\pi 2mk/N}. \quad (2)$$

Thus, it provides a representation with highly concentrated speech components that can be precisely localized in the time-frequency plane. However, since speech is a multicomponent signal, its Wigner distribution contains large amount of cross-terms [16]. Therefore, the S-method (SM) that combines the advantages of spectrogram and Wigner distribution is used. It preserves auto-terms concentration as in the Wigner distribution,

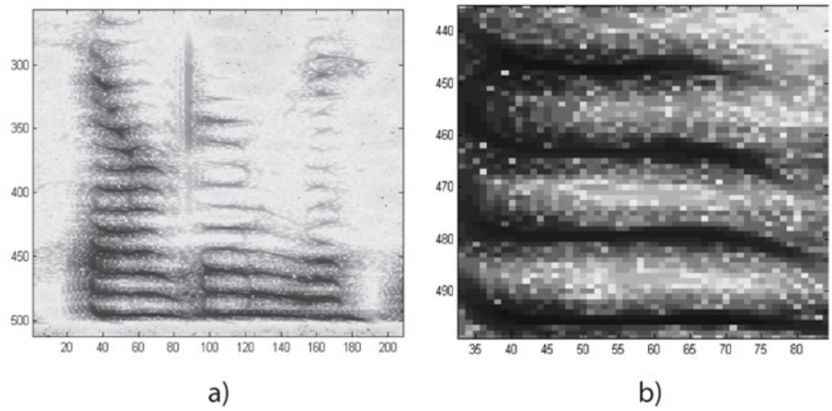

Fig. 1. a) The S-method of a speech signal, b) the considered region D

and removes or significantly reduces the presence of cross-terms [17]. The discrete form of the SM is defined as:

$$SM(n,k) =$$

$$= \sum_{l=-LL}^{LL} P(l) STFT(n,k+l) STFT^{*}(n,k-l), \tag{3}$$

where $P(l)$ is a finite frequency domain window with the length $2 \cdot LL+1$.

Furthermore, the SM is used to determine the location of formants components. Consider the region D in the time-frequency plane, i.e., $D = \{(t,\omega) : t \in (t_1,t_2), \omega \in (\omega_1,\omega_2)\}$, so that the formants exist within the time interval (t_1,t_2). The frequency interval $\omega \in (\omega_1,\omega_2)$, selects the low frequency formants. For a signal sampled at 8 KHz, the interval $\omega \in (100\text{Hz}, 750\text{Hz})$is considered. The time instances t_1 and t_2 correspond to the start and the end of one voiced speech part, which is selected by using Teager energy based word endpoint detector. The SM of a speech signal is presented in Figure 1.a, while the SM within the region D is shown in Figure 1.b.

In order to select the strongest formants in D, a threshold value T is introduced. It is defined as: $T > \overline{|SM(t,\omega)|}$, where $\overline{|SM(t,\omega)|}$ is the mean absolute value of the SM computed over D. Applying the threshold T, the region D becomes:

$$D_T = \{(t,\omega) : t \in (t_1,t_2), \omega \in (\omega_1,\omega_2) \text{ and } \mathrm{SM(t,\omega) > T}\} \tag{4}$$

Thus, the region D_T provides the information about significant speech components. The DFT coefficients that correspond to this region will be used in watermarking procedure. In order to provide certain level of security as well as better performance of the proposed procedure, the multiple logo embedding is considered. Thus, more than one region of speech signal will be used for logo embedding.

III. Logo Embedding Procedure

Consider a set of DFT coefficients selected by the region D_T, here named the DFT_{DT} coefficients. The binary logo will be embedded in several specific bit-planes of the DFT_{DT} coefficients. Since the logo is a 2D signal, the DFT_{DT} coefficients are reordered in an arbitrary 2D matrix. The logo is decomposed into sub-images that are embedded in the bit-planes of the DFT_{DT} matrix by using a unique secret key to ensure security of the procedure.

Assume that the number of DFT_{DT} coefficients is n. The DFT_{DT} matrix is of size MxN, where M and N are chosen to satisfy: $M \cdot N = n$. The binary representation of coefficients within the DFT_{DT} matrix is used. Thus, one bit-plane contains the bits on the same quantization level. An illustration of bit-plane is shown in Figure 2, where white and

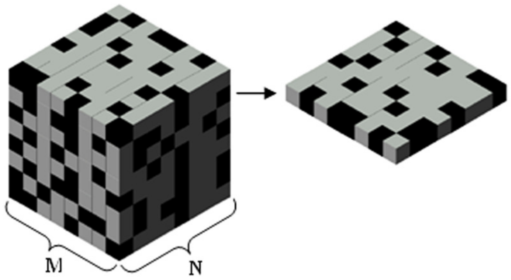

Fig. 2. Illustration of a bit plane

black fields represent the bit values "0" and "1", respectively.

In order to retain the quality of speech signal, only a few bit-planes of DFT_{DT} matrix should be used for watermarking. The bit-planes selection depends on specific constraints imposed by the watermarking application scenario. In the case of robust watermarking procedure, the mid-level bit-planes should be considered (the most significant bits should be omitted since their modifications could influence the signal quality). The least significant bits should be considered for the fragile watermarking technique, so that any modification of watermarked DFT_{DT} coefficients results in altered logo information. It is important to note that the bit-planes should be selected in relation to the most significant one. In this way, the positions and values of the watermarked bit-planes will remain unchanged under amplitude scaling. For example, if the amplitudes are divided by 2, the bit-planes (Figure 2) will be shifted toward the least significant one, but their distance from the most significant bit-plane will not be changed.

Logo is created as a binary image: $L(i,j)$, i=1,...,P, j=1,...Q, where $P \leq M$ and $Q \leq N$. In order to simplify the procedure, P=M and Q=N is considered. In the sequel, the unique procedure for generating logo sub-images is provided. The number of sub-images is equal to the number of selected bit-planes. Note that if logo is embedded in only one bit-plane, without using logo sub-images, it can be visible and easily removed. As unique key, a 2D random matrix Z (of size MxN), with the values in the range (ψ_d, ψ_g), is used. The logo sub-images are generated according to:

$$F_k(i,j) = \begin{cases} 1 & \text{if} \quad \psi_{k-1} \leq Z(i,j) < \psi_k \\ & \quad \wedge \; L(i,j) = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

$$k = 1,\ 2...\ ,\ \text{q}, \qquad \text{i} = 1,\ 2, ...,\ \text{M},$$

$$j = 1,\ 2, ...,\ \text{N},$$

where F_k is the k-th logo sub-image, q is the number of selected bit-planes of the DFT_{DT} matrix, while $\wedge$ represents the logical AND operation. The threshold values ψ_k applied to the matrix Z are obtained as:

$$\psi_{k-1} = \psi_d + (k-1)\frac{\psi_g - \psi_d}{q}. \tag{6}$$

Note that thresholds ψ_k are determined by using an equidistant rule. However, by using some other secret rule for threshold determination, an additional key for watermarking procedure can be obtained.

The logo sub-images F_k are embedded into the bit-planes B_k of the DFT_{DT} matrix according to:

$$B_k(i,j) = \begin{cases} 1 & \text{if} \quad F_k(i,j) = 1 \\ B_k(i,j) & \text{otherwise.} \end{cases} \tag{7}$$

In the case of multiple logo embedding, the described procedure is done for each selected region of speech signal and its corresponding set of the DFT coefficients.

IV. Logo Extraction Procedure

In order to retrieve the logo from a watermarked signal, the steps of logo embedding procedure are repeated. Namely, the watermarked speech region is determined by using time-frequency representation. The corresponding matrix of watermarked DFT_{DT} coefficients is obtained identically as in the case of original DFT_{DT} matrix. Instead of embedding, logo sub-images are extracted from the watermarked bit-planes of DFT_{DT} matrix according to:

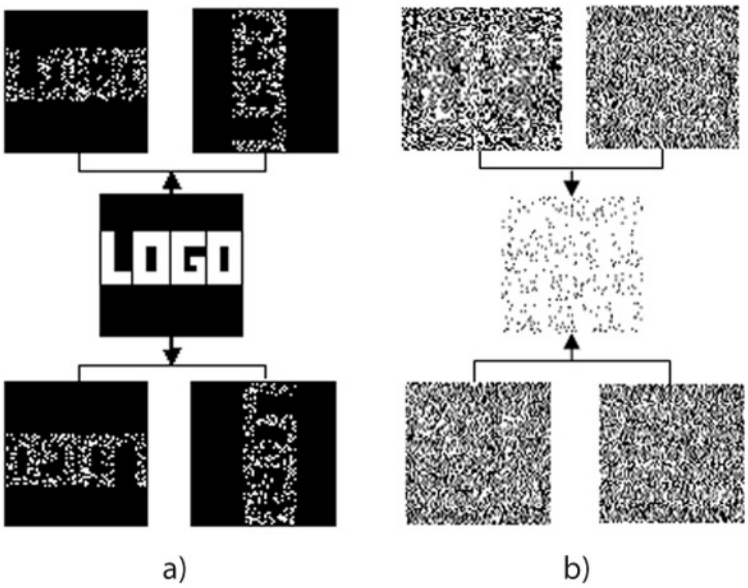

Fig. 3. a) Logo and rotated sub-images, b) watermarked bit-planes and its summation

$$F_k(i,j) = \begin{cases} B_k(i,j) & \text{if} \quad \psi_{k-1} \leq Z(i,j) < \psi_k \\ & \quad \wedge \; L(i,j) = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

The logo sub-images are not noticeable within the bit-planes of watermarked DFT_{DT} matrix. Thus, the logo sub-images cannot be extracted if the random matrix Z (key) is not available, since it provides information about positions of embedded sub-images pixels. By adding extracted sub-images, the original logo is obtained.

V. Examples

Speech signals with maximal frequency 4 kHz are considered. In order to select a suitable time-frequency region, the SM with parameter LL=5 is used. The STFT is calculated by using rectangular window with 1024 samples. The selected DFT_{DT} coefficients are arranged in a square matrix. Only the real parts of the DFT_{DT} coefficients are considered. All coefficients are represented by 16 bits. Four bit-planes are used for logo sub-images embedding. Thus the logo is decomposed into four sub-images (Figure 3.a).

In order to improve the security of watermarking procedure, the sub-images are rotated for angles $k\alpha$ (α=90°, k=0,1,2,3) before embedding them in the bit-planes of the DFT_{DT} matrix. The watermarked bit-planes are shown in Figure 3.b. Note that the logo sub-images are not visible within the watermarked bit-planes, or within the sum of bit-planes (Figure 3.b. center).

In the logo extraction procedure, the random matrix Z is used as a key, in order to determine the positions of sub-images pixels. Only if Z is available, the logo sub-images can be retrieved. The original logo is obtained by adding the extracted sub-images rotated for $-k\alpha$ (α=90°, k=0,1,2,3).

A. Example 1

In order to demonstrate the robustness of the proposed procedure, the low frequency formants belonging to the voiced speech parts are selected according to (4). The same logo is embedded into the three different DFT_{DT} matrixes that correspond to the selected regions. Further, four mid-level bit-planes of the DFT_{DT} matrixes have been used. One original and the corresponding watermarked

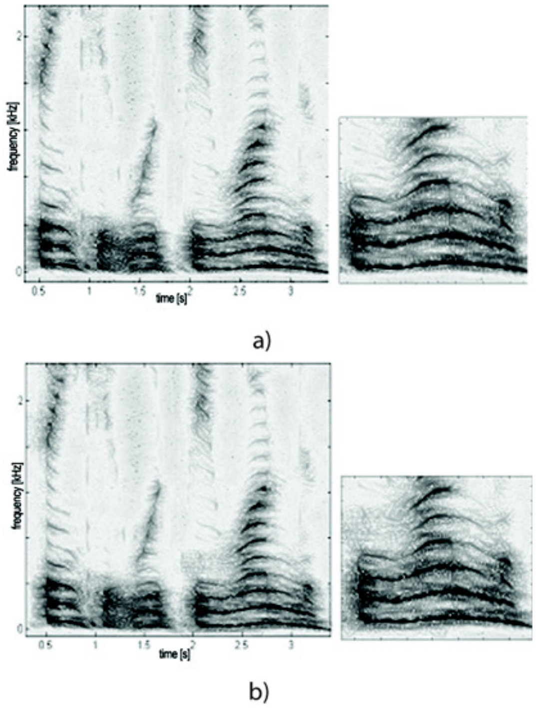

Fig. 4. a) Part of the original speech signal with zoomed non-watermarked region, b) Part of the watermarked speech signal with zoomed watermarked region

time-frequency region are shown in Figure 4.a and Figure 4.b, respectively. The differences between original and watermarked signal are negligible.

The obtained signal to watermark ratio (SWR) is around 37 dB, and the watermark is not audible, as it will be explained in the sequel. The watermark imperceptibility has been proven by using the ABX listening test, where A, B and X are the original, watermarked, and original or watermarked signal, respectively. The listener listens to A and B. Then listener listens to X and decides whether X is A or B. Three female and five male listeners with normal hearing participated in the listening test. The test was performed five times, and from the obtained statistics it was concluded that the listeners cannot positively distinguish watermarked and original signal (the signal was correctly recognized in only 43% of cases, approximately). Additionally, a kind of quality rating listening test is done. We

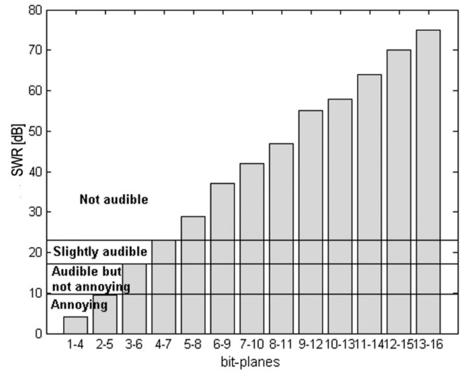

Fig. 5. SWR in terms of selected bit-planes for logo embedding

have performed logo embedding within different level bit-planes of DFT_{DT} matrix (from the first bit plane (MSB) until the last one (LSB), considering four consecutive bit-planes in each test). The values of SWR for each tested set of bit-planes are calculated (Figure 5). Here, the bit-planes denoted as 1-4 are the four most significant bit-planes, while 13-16 are the four least significant bit-planes. In the cases when the watermark was audible, the listener had to characterize the distortion as follows: annoying, audible but not annoying or slightly audible. For the SWR bellow 10 dB the distortion was declared as annoying. However, in the case 10dB$<$SWR$<$18dB, although the values of SWR are still very low, the distortion was not annoying. Namely, the watermarked coefficients correspond to the strongest voiced formants that mask the intensity of distortion, while all the sensitive speech parts are avoided. Slightly audible category, for 18dB$<$SWR$<$23dB, means that the original and watermarked signal are correctly recognized in 60% of cases, approximately. Finally, for SWR higher than 23 dB, the watermark has been declared as not audible, since the listeners were not able to distinguish the original from watermarked signal.

In the logo extraction procedure, all extracted logo sub-images are composed to obtain the original readable mark. Suppose that W is the original watermark (logo) and W' is the extracted watermark. In order to evaluate

TABLE I
LOGO EXTRACTION RESULTS FOR VARIOUS ATTACKS

	No attack	**Adding noise (Gaussian)**	**Low pass filter**	**Median filter**	**MP3 128 kbps**
Extracted logo	LOGO	LOGO	LOGO	LOGO	LOGO
	NC =1	NC=0.992	NC=1	NC=0.975	NC=1
	BER =0	BER=0.008	BER=0	BER=0.025	BER=0
	MP3 64 kbps	**Mp3 70-120 kbps**	**Mp3 40-50 kbps**	**Echo (100 ms)**	**Delay 100 ms, mixing 5%**
Extracted logo	LOGO	LOGO	LOGO	LOGO	LOGO
	NC=1	NC=1	NC=1	NC=1	NC=0.93
	BER=0	BER=0	BER=0	BER=0	BER=0.07
	Pitch scaling (5%)	**Amplify +1.5dB**	**Amplify +0.5dB**	**Amplify -0.5dB**	**Amplify -1.5dB**
Extracted logo	LOGO	LOGO	LOGO	LOGO	LOGO
	NC=0.91	NC=0.9	NC=0.93	NC=0.942	NC=0.91
	BER=0.09	BER=0.1	BER=0.07	BER=0.058	BER=0.09
	Time stretch 10%	**Re-quantization, 8b**	**Re-sampling 11.025 kHz**	**Re-sampling 22.050 kHz**	**Re-sampling 32 kHz, 8b**
Extracted logo	LOGO	LOGO	LOGO	LOGO	LOGO
	NC=0.971	NC=0.98	NC=1	NC=1	NC=0.974
	BER=0.029	BER=0.02	BER=0	BER=0	BER=0.026
	Wow	**Bright mono flutter**	**MP3 64 kbps + Low pass filter**	**MP3 64 kbps + noise**	**MP3 64 kbps + Re-quantization**
Extracted logo	LOGO	LOGO	LOGO	LOGO	LOGO
	NC=0.93	NC=0.97	NC=0.998	NC=0.99	NC=0.99
	BER=0.07	BER=0.03	BER=0.002	BER=0.01	BER=0.01

the efficiency of logo extraction, the bit error rate (BER) has been calculated as follows [18]:

$$BER(W, W') = \frac{D}{N},$$

where N is the number of bits within the watermark, while D is the number of successfully recovered watermarking bits. BER is calculated as a mean value on the three embedding regions.

Additionally, in order to compare our procedure with State-of-The-Art, the normalized correlation coefficient is used:

$$NC(W, W') = \frac{\sum WW'}{\sum W_i^2}.$$

The obtained values for NC and BER are given in Table I. The procedure is tested on the following attacks: additive Gaussian noise (SWR 32 dB), low-pass filtering (with cutoff frequency at 1.6 kHz), median filtering (with window size=5), mp3 compression with constant bit rate 128 kbps and 64 kbps (wav-mp3-wav), mp3 with variable bit rates 70-120 kbps,

TABLE II
Comparison with previously reported experimental data

Attack	Proposed procedure	Tsai et all [8]	Jian et all [10]	Tang et all [9]	Xu et all [11]
Low pass*	NC=1 BER=0	NC=0.64	NC=0.947	NC=0.86	NC=0.958 BER=0.018
		NC=0.998			
mp3 128Kbps	NC=1 BER=0	NC=0.968	-	-	NC=0.962 BER=0.016
mp3 64Kbps	NC=1 BER=0	NC=0.24	NC=0.993	NC=0.856	-
Re-sampling 11.025kHz	NC=1 BER=0	-	NC=0.95	NC=0.88	-
Re-sampling 32kHz	NC=1 BER=0	-	-	-	NC=0.882 BER=0.05
Re-sampling 32kHz, 8b	NC=0.974 BER=0.026	-	-	-	NC=0.857 BER=0.057
Echo adding	NC=1 BER=0	-	NC=0.997	-	NC=1 BER=0
Median (window length =5)	NC=0.975	NC=0.836	-	-	-
mp3 64Kbps + Low pass	NC=0.998	NC=0.93	-	-	-

*Note that, in the proposed procedure, the results for NC and BER in the case of low pass filtering are obtained at cutoff frequency 1.6 kHz, for speech signal sampled at 8 kHz. Within the existing procedures, the results for low pass at cutoff frequency 16 kHz (NC=0.64) and 22 kHz (NC=0.998) in [8], 4 kHz in [10], 9 kHz in [9] and 10 kHz in [11], for audio signals sampled at 44.1 kHz are given.

mp3 compression with variable bit rates 40-50 kbps, echo adding (100 ms), Bright mono flutter, wow, delay 100 ms with mixing 5%, pitch scaling (5%), amplifying (+1.5 dB, +0.5 dB, -0.5 dB, -1.5 dB), Re-quantization (16-bit to 8-bit and back to 16-bit), Re-sampling (watermarked signal sampled at 8kHz is up-sampled to 11.025 kHz, 22.050 kHz, 32 kHz and then down-sampled back to 8 kHz). Some of the attacks are realized by using audio processing software Cool Edit Pro v2.0, while the rest of processing is done in Matlab 7. It can be seen from Table I that for most attacks the procedure guaranties high level of robustness, which makes this method suitable for ownership protection applications.

The proposed algorithm is compared with the existing algorithms [8]-[11], that perform logo embedding in audio signals and use the same quality measures for logo extraction results. In Table II, the values of NC (or both NC and BER) are given for attacks that are tested within the mentioned procedures (the slash denotes that the results for the considered attack are not reported). It can be seen that the proposed procedure provides better results than the other considered procedures.

B. *Example 2*

As an alternative, in this example we have aimed to show that by using slight modifications in the logo embedding procedure, the semi-fragile and fragile scenario for data authentication can be obtained. Here, two types of authentication can be distinguished: exact authentication and selective authentica-

TABLE III
The results of logo extraction under attacks (fragile scenario)

Attack	Extracted Logo under attack
Gaussian noise	NC=0.43 BER=0.57
Low-pass Filter	NC=0.45 BER=0.55
Median filter	NC=0.42 BER=0.58
mp3 128 kbps	NC=0.5 BER=0.5

TABLE IV
The results of logo extraction under attacks (semi-fragile scenario)

Attack	Extracted Logo under licit attacks	Attack	Extracted Logo under stronger attacks
No attack	NC=1 BER=0	**Median filter**	NC=0.783 BER=0.217
Low-pass Filter	NC=0.995 BER=0.005	**Echo 100 ms**	NC=0.79 BER=0.21
Adding noise (37dB)	NC=0.968 BER=0.032	**Delay 100ms, mixing 5%**	NC=0.716 BER=0.284
mp3 128kbps	NC=1 BER=0	**Pitch scale 5%**	NC=0.687 BER=0.313

tion [19].

If the four least significant bit-planes are used, logo is fragile even under licit attacks (Table III). Thus, it could be used for traditional data authentication called exact authentication [19], in applications that do not accept any data alteration (for example when the data are used as a proof in a court of law [1]).

Semi-fragile watermarking for selective authentication is obtained by using higher bit-planes (less significant) than in the previous example (providing SWR=45 dB). The coefficients within the DFT_{DT} matrix are selected corresponding to low-middle frequency formants. The logo, which is embedded twice within the signal, can be almost perfectly recovered after some licit attacks [20] (low-pass with cutoff frequency at 2 kHz, adding noise, mp3 compression at 128 kbps). However, after some stronger attacks, the bit error rate significantly increases, causing non-reliable results (Table IV). The semi-fragile approach will be further developed and included within the future work.

VI. Conclusion

A procedure for logo watermarking of speech signals is proposed. Time-frequency based selection of the DFT coefficients is used to ensure good trade-off between robustness and imperceptibility. The speech regions suitable for watermarking are selected by using the S-method. The DFT coefficients that correspond to the selected regions are arranged into a 2D matrix. The binary logo is decomposed into sub-images that are embedded in the bit-planes of DFT matrix. The embedded logo does not influence the quality of watermarked speech signal, and it is not visible inside the bit-planes of the DFT matrix. The logo sub-images can be extracted from the watermarked signal only if the watermarking key is available. The robustness of the proposed procedure has been tested against various attacks. A comparison with other procedures that deal with logo watermarking of audio signals is given. It has been shown that the proposed approach provides high level of robustness and can be efficiently used for ownership protection applications. Additionally, with slight modifications, the proposed algorithm can also be used for data authentication, as illustrated in the last example.

Acknowledgement

The authors are thankful to the anonymous reviewers for useful comments that helped to make the paper more complete.

References

[1] Barni (M.) and Bartolini (F.), Watermarking Systems Engeneering, Marcel Dekker, Inc., 2004.

[2] Hartung (F.), Kutter (M.), Multimedia Watermarking Techniques, *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1079-1107, July 1999.

[3] Bassia (P.), Pitas (I.), Robust Audio Watermarking in the Time Domain, *IEEE Trans. on Multimedia*, vol. 3, No. 2, pp. 232-241, June 2001.

[4] Ko (B.S.), Nishimura (R.), Suzuki (Y.), Time-Spread Echo Method for Digital Audio Watermarking, *IEEE Trans. on Multimedia*, vol. 7, No. 212-221, April 2005.

[5] Cvejić (N.), Algorithms for Audio Watermarking and Steganography, University of Oulu, 2004.

[6] Kirovski (D.), Malvar (H.S.), Spread-Spectrum Watermarking of Audio Signals, *IEEE Transactions on Signal Processing,* vol. 51, No. 4, pp. 1020-1033, April 2003.

[7] Swanson (M. D.), Zhu (B.), Twefik (A. H.), and Boney (L.), Robust Audio Watermarking Using Perceptual Masking, *Signal Processing*, vol. 66, No. 3, pp. 337-355, May 1998.

[8] Jian (W.), Fu-zhong (L.), Digital Audio Watermarking Based on Support Vector Machine (SVM), J. Comput. Res. Devel., vol. 42, No. 9, pp. 1605-1611, 2005.

[9] Tang (X.), Niu (Y.), Yue (H.), Yin (Z.), A Digital Audio Watermark Embedding Algorithm, *International Journal of Information Technology*, vol. 11, No. 12, pp. 24-31, 2005.

[10] Tsai (H. H.), Cheng (J. S.), Yu (P. T.), Audio Watermarking Based on HAS and Neural Networks in DCT domain, *EURASIP Journal on Applied Signal Processing*, vol. 2003, No. 3, pp. 252-263.

[11] Xu (X.), Peng (H.) and He (C.), DWT-Based Audio Watermarking Using Support Vector Regression and Subsampling, *Proc. of the 7th International Workshop on Fuzzy Logic and Applications*, WILF 2007, Camogli, Italy, pp.136-144, July 2007.

[12] Kundur (D.), Hatzinakos (D.), Toward robust Logo watermarking Using Multiresolution Image Fusion, *IEEE Trans. on Multimedia*, vol. 6, No. 1, pp. 185-198, Feb. 2004.

[13] Braudaway (G. W.), Protecting publicly-available images with an invisible image watermark, in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Oct. 1997, pp. 524–527.

[14] Arnold (O.), Wolthusen (S.), Schmucker (M.), Techniques and Applications of Digital Watermarking and Content Protection, Artech House Publishers, July 2003.

[15] Su (P.C.), Kuo (C.C.J.), Blind watermarking for cartoon and map images, *Proc. SPIE, Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 296-306, Jan. 1999.

[16] Boashash (B.), Time-Frequency Signal Analysis and Processing, *Elsevier* 2003.

[17] Stanković (LJ.), A method for Time-Frequency Signal Analysis, *IEEE Transaction on Signal Processing,* vol. 42, No. 1, Jan. 1994.

[18] Potdar (V.M.), Han (S.), Chang (E.), A survey of digital image watermarking techniques, Proc. of IEEE's 3rd International Conference on Industrial Informatics**,** INDIN'05, August 2005, pp. 709-716.

[19] Alomari (R.S.) and Al-Jaber (A.), A Fragile Watermarking Algorithm for Content Authentication, *International Journal of Computing and Information Sciences*, vol. 2, No. 1, pp. 27-37, April 2004.

[20] Altun (O.), Sharma (G.), Celik (M. U.) and Bocko (F. M.), A Set Theoretic Framework for Watermarking and Its Application to Semifragile Tamper Detection, *IEEE Trans. on Information Forensics and Security*, vol. 1, No. 4, pp. 479-492, Dec. 2006.