

AN AUDIO WATERMARKING ALGORITHM VIA ZERO ASSIGNED FILTER BANKS

Zeynep Yücel and A. Bülent Özgüler

Electrical and Electronics Engineering Department,
Bilkent University, 06800, Bilkent, Ankara, Turkey
phone: + (90) 312 266 4307 fax: + (90) 312 266 4192, email: zeynep@ee.bilkent.edu.tr, ozguler@ee.bilkent.edu.tr
web: www.ee.bilkent.edu.tr

ABSTRACT

In order to identify the owner and distributor of digital data, a watermarking scheme for audio files is proposed in frequency domain. The scheme satisfies the imperceptibility and persistence requirements and is robust against additive noise. It consists of a few stages of wavelet decomposition of several frames of the original signal using special zero assigned filter banks. By assigning zeros to filters on the high frequency portion of the spectrum, filter banks with frequency selective response is obtained. Text information is then inserted in the wavelet-decomposed and compressed signal. Several robustness tests are performed on male voice, female voice, and music files.

I. INTRODUCTION

Due to rapid developments in information technology, digital data became easily accessible through the multimedia services on the Internet. This raises issues of copyright and intellectual property protection. A popular approach employed to embed imperceptible structures into digital media files in order to hinder unauthorized copying and distribution is digital watermarking.

Watermarking may be performed in spatial domain or in frequency domain. In this paper we deal with watermarking audio signals in frequency domain via wavelets and zero assigned filter banks. Previous works in frequency domain watermarking are addressed in [1], [2], [3].

Wang *et al.* discuss the practical requirements for watermarking systems [3]. For standardized algorithms storing watermarks, original or marked signals and secret keys may introduce excessive memory requirements and a great deal of financial burden for registration of all those by the legal authority. A good marking scheme should meet several requirements as explained in detail by Swanson *et al.* [4]. Embedding data must not violate the perceptual quality of the host signal. The mark should be easily detectable. It is also a desired property that the recovery of data does not use the original signal to decode the embedded watermark. Furthermore it must be robust against modifications and manipulations such as compression, filtering, and additive noise.

The marking procedure must be able to resolve rightful ownership when multiple ownership claims are made. A pirate may modify the marked signal in a way that if his fake original signal is used in detection process, both claimers may gather equal evidence for ownership, [5]. This situation is called the deadlock problem [4]. The importance of decoding without the original signal arises here. On the other hand, the author should provide secret keys in order to obtain a more secure encryption technique that allows only the authorized detections of the watermark with the help of proper keys.

The watermarking algorithm proposed here consists of a few stages of wavelet decomposition of several frames of the original signal using special zero assigned quadrature mirror filter (QMF) banks. By assigning zeros to filters on the high frequency portion of the spectrum, filter banks with frequency selective response is obtained. Text information is then inserted in the wavelet-decomposed and compressed signal. Two perfect reconstruction quadrature mirror filter banks with different assigned zeros are built. Then each frame is processed with one of the filter banks depending on the bit to be embedded in the frame of interest. The insertion process is run as much as the length of input signal allows and this gives the opportunity of checking reliability of the extracted sequence. In decoding, signal is divided into frames and each frame is tested with both filter banks in order to determine the assigned zero and, hence, the embedded bit.

The original signal is not used for detecting the embedded text. As our approach accounts for the features of Human Auditory System (HAS) during the design of the filter banks, their frequency responses are adjusted to match the characteristics of HAS and perceptual transparency condition is thus satisfied. Detection procedure requires the storage and transmission of the stage number, frame size, and values of assigned zeros to QMF's. As multiple keys are used in designing the filter banks, the watermarking scheme is secure against pirates. Furthermore, since the conditions in [5] are satisfied, a deadlock problem does not arise under these circumstances. Simulations show that even under high channel noise rates when the signal itself is hardly intelligible, the watermark can still be extracted with a bit reliability of more than 95%. Thus robustness against channel noise is obtained up to a considerable level.

2. ZERO ASSIGNMENT

In a perfect reconstruction (PR) QMF filter bank design [6], synthesis filters are completely determined by the analysis filters. Thus, the construction of the filter bank reduces to the construction of the analysis filters. The zero assignment in our method refers to the construction of FIR, QM, and minimal length analysis filters having assigned zeros at desired locations with respect to the unit circle (or at desired frequencies), [7]. We now summarize the PR, FIR, QM, and minimal length filter bank construction method of [7].

Suppose a permitted odd filter bank delay of n_0 is given. Further suppose that $G_1(z)$ and $G_2(z)$ are two FIR transfer functions of order (number of zeros) k each whose zeros coincide with the desired zeros of the analysis low-pass filter $H_1(z)$ and high-pass filter $H_2(z)$, respectively. Thus, the analysis filters will contain desired zeros if and only if

$$\begin{aligned} H_1(z) &= \hat{H}_1(z)G_1(z) \\ H_2(z) &= \hat{H}_2(z)G_2(z) \end{aligned}$$

The PR condition implies that

$$H_1(z)H_2(-z) - H_1(-z)H_2(z) = 2z^{-n_0}$$

or

$$\hat{G}_1(z)\hat{G}_2(-z)\hat{H}(z)\hat{H}_2(-z) - \hat{G}_1(-z)\hat{G}_2(z)\hat{H}(-z)\hat{H}_2(z) = 2z^{-n_0}$$

where n_0 is the filter bank delay. For this equation to have a solution, it is necessary that the greatest common divisor of $(G_1(z)G_2(-z), G_1(-z)G_2(z))$ is of the form z^m , i.e., it should be a pure delay. Define $G(z) = G_1(z)G_2(-z)$ and for simplicity assume that $(G(z), G(-z))$ is co-prime. It is shown in [7] that a minimal length FIR solution $\hat{H}(z)$ of

$$G(z)\hat{H}(z) - G(-z)\hat{H}(-z) = 2z^{-n_0} \quad (1)$$

exists and is unique whenever $n_0 < 4k$ and has order at most $2k-2$. The analysis filters are obtained by a factorization

$$\hat{H}(z) = \hat{H}_1(z)\hat{H}_2(-z),$$

and are in general non-unique. A hand-rule is to select the left half plane zeros in the low-pass filter and right half plane zeros in the high-pass filter, [7]. As an example, consider the desired zeros to be $-1, -0.97 + 0.2431i$ and $-0.97 - 0.2431i$ for the low pass filter and $1, 0.8 + 0.6i, 0.8 - 0.6i$ for the high-pass filter. Suppose that the duration of the allowable delay is $n_0=5$. Under these circumstances, minimal order solution to (1) and its factorization according to the hand rule described above produces the following high-pass and low-pass filters of order four each with frequency responses given in Figure 1.

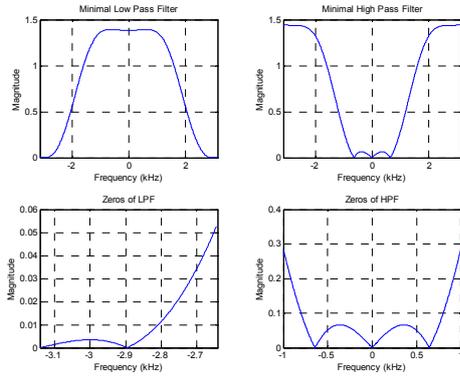


Figure 1. Frequency responses of the designed filters (a) Low-pass filter (b) High-pass filter (c) Zoomed image around assigned zero for LPF (d) Zoomed image around assigned zero for HPF

3. WATERMARKING SCHEME

3.1 Proposed Algorithm

Human auditory and visual systems are imperfect detectors [4]. Human ear usually cannot detect sounds at very high frequencies. It is thus proposed to construct analysis filters, which have assigned zeros at the high frequency portion of the spectrum and to build a several stage filter bank.

Encoding

Step 1: Design two filter banks with different assigned zeros by using the scheme explained in Section 2. Each filter bank will correspond to one of the bits 0 or 1.

Step 2: Divide the input signal into a fixed number of frames.

Step 3: Fix a stage number to be used in the wavelet decomposition using the cascade algorithm, [9].

Step 4: Obtain the wavelet decomposition of each frame via the cascade algorithm using the number of stages fixed in Step 3.

Step 4: Obtain a compressed version of the each frame by zeroing the highest stage detail coefficients obtained in Step 4.

Step 5: The watermarked signal is the concatenation of the compressed signals of Step 4. The watermark consists of the information embedded in each frame as a bit 0 or 1 depending on the filter bank that has been used to obtain the wavelet decomposition for that particular frame. The sequence of 1's and 0's embedded in consecutive frames allows us to encode text information as the watermark.

Decoding:

Step 1: Partition the watermarked signal into frames of a fixed size, which is provided as one of the keys.

Step 2: Obtain the wavelet decomposition of each frame with both filter banks using the cascade algorithm. Here, the construction of the analysis filters, and hence the filter banks, require the information consisting of n_0 and the assigned zeros. This information constitutes the second key. The number of stages of the cascade algorithm must also be provided as the third key.

Step 3: Extract the bit information embedded in a frame from the two wavelet decompositions of that frame by a comparison of their highest stage detail coefficients.

Step 4: Verify the ownership by identifying the correct sequence of 1's and 0's in the consecutive frames.

The selection of the zeros to be assigned to the analysis filters should be such that frequency suppression provided by the filters does not affect the quality of sound. In this application, the following procedure is used to determine the zeros to be assigned. A frequency value f is determined by examining the spectra of all frames of the original audio signal. This value f should be a high frequency value at which each frame should have a nonzero component. The determination of f fixes one of the zeros. In order to emphasize the suppression at that frequency, usually two or three copies of the same zero is incorporated into the low-pass filter. Once $G_1(z)$ is so determined, the zero assignment procedure in Section 2 determines the analysis filters uniquely. The number of stages to be employed in the cascade algorithm must generally be low in order for the decoding process to succeed. However, it can not be a single stage as this may violate the perceptual transparency requirement for watermarking. In order to get closer to a pure tone, the filtering process is thus carried out for two or three stages along the high frequency branch of the filter bank. After filtering the input with the high-pass decomposition filter $H_2(z)$ several times, components of frequency f is accumulated on the lower most branch. All the coefficients on that section are set to zero and the reconstruction operation is then done. The following illustrates the whole process for a 2-stage decomposition and reconstruction.

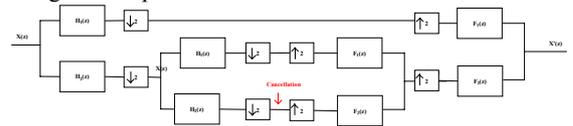


Figure 2. Two stage implementation of the cascade algorithm

3.2 Watermarking with Zero Assigned Filters

In this application, we embed text data in male voice, female voice, male voice with pauses, and music signal. A set of two filter banks are obtained by assigning f_1 or f_2 as frequencies to be suppressed by the low-pass analysis filters. Say the filter bank with a zero assigned at f_1 is FB_1 and the one with a zero assigned at f_2 is FB_2 . Each of the frames is processed with either FB_1 or FB_2 depending on the bit to be embedded in that frame according to the cascade algorithm of Fig. 2.

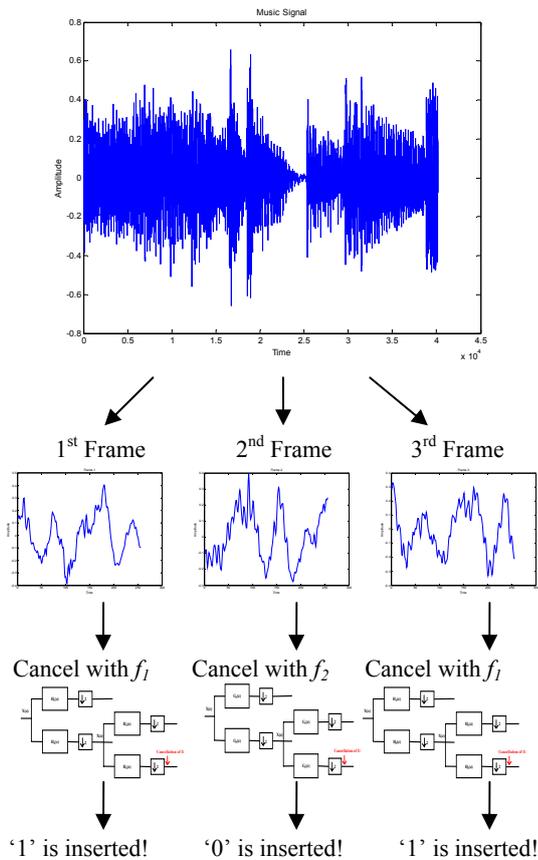


Figure 3. Insertion of bit sequence '101' into a music signal

The watermark consists of the information embedded in each frame as a bit 0 or 1 depending on the filter bank that has been used to obtain the wavelet decomposition for that particular frame. The sequence of 1's and 0's embedded in consecutive frames allows us to encode text information as the watermark. Simple reconstruction operation is carried out with corresponding synthesis filters according to the cascade algorithm. The knowledge of filter bank delay n_0 , frequencies f_1 and f_2 , stage number, frame size, and the watermark sequence are the keys to be provided to the authority for storage. A careful choice of the watermark may increase the robustness of the algorithm against false alarms during the detection process. Note that, fixing the frame size also fixes the number of bits that can be embedded in the watermarked signal. For instance, our input signals are sampled at 22 kHz and each sample is represented in 8 bits. Duration of the experimented audio signals is around 1-3 seconds. In the experiments, the text to be embedded is chosen to be composed of a word with several letters, generally 4-7 letters. For a 6-letter word, 42 bits must be embedded as each letter is represented in 7 bits in the ASCII table. For each bit a frame of 128, 256 or 512 samples is taken. This allows us, e.g., to embed about 12 copies of a 6-letter word in a 3-second signal if we use frames of size 128. This would tolerate one or two false diagnosis of the text in say noisy environments.

3.3. Watermark Detection

Although the filter banks of the cascade algorithm are determined for perfect reconstruction, the reconstructed signal would not be the same as the original one in the cascade

algorithm of Figure 2. This is because a compressed version of the original signal is fed into the synthesis part. Because of imaging that results from the upsampling in the reconstruction phase, components from the preserved regions are introduced onto the frequencies f_1 and f_2 . In the decoding phase, an authority checks which frequency is suppressed in each frame by constructing FB_1 and FB_2 . In order to understand whether the highest stage detail signal is an image or not, a detection rule must be used. One possibility, followed in our application, is to compute and compare the L_1 norms of the detail coefficients obtained via FB_1 and FB_2 . In our case this method of comparison has been very successful. However, for different applications and depending on the attack or add-on noise in the transmission channel, alternate detection schemes may be employed. Note that there is a trade-off between the number of stages of the cascade algorithm and the sound quality of the watermarked audio data. As the number of stages increases, the number and the size of coefficients of the highest stage detail signal are getting smaller. Setting coefficients carrying less information to zero will yield less distortion in the watermarked signal. However, it will now be harder for the authority to detect which frequency is suppressed, and hence which bit information is embedded, since the result of the comparison of the detail coefficients obtained via FB_1 and FB_2 will now be more sensitive to noise and other sources of disturbance. We thus made experiments with 2 and 3 stage filter banks only and obtained good results. The probability of extracting wrong information is quite low especially in the noise free medium.

The distinction between the filters FB_1 and FB_2 , and hence the distinction between the detail coefficients obtained by these different filters, is solely dependent on the choice of f_1 and f_2 . Obviously, closer values for f_1 and f_2 will give rise to difficulties in the detection scheme. While they should both be on the high frequency portion for perceptual transparency, they being placed too close will cause false alarms to occur more often. In our experiments, we worked with zeros that are separated by 1% to 23% of the whole spectrum and got good results in several media. We note that, zeros may even be placed around the mid-frequency band while watermarking in noisy environments. As the noise on the watermarked signal increases, the distortion resulting from the inserted data becomes less perceptible. Thus, the add-on-noise acts like a mask for the watermark. As a result, filter banks with assigned zeros on 23% of the spectrum can work with at least 95% bit reliability at decoding under a white Gaussian noise of 22dB.

4. Experimental Results

We have performed robustness experiments on audio files with different sound characteristics. These files were the recordings of male and female voices and a music file together with a male voice recording with pauses. All are sampled at 22 kHz and represented in 8 bits. Robustness issues in mainly three situations are considered: noise free medium, noisy medium and attacks. Details of the experimental results reported below are provided in [10].

4.1. Experiments in Noise Free Medium

Effects of stage number selection, frame size selection and placement of zeros are examined in noise free environment. One of the zeros is fixed at 10% vicinity of 2π and the other is located at %3, %5, %7 or %9 of 2π , respectively. The effect of these alternative zero configurations on bit reliability is examined. Our method works with 100% bit reliability even when the zeros are placed 2% percent apart on the unit circle. Only when one zero is on 9% of 2π and the other is on 10% of 2π , the bit reliability decreases to 0.964. But as our watermarking scheme allows us to insert multiple copies of the text into the audio signal, missing out a few bits in the detection stage still

provides many correct sequence of text to be detected. On the other hand, it should be noted that as long as the perceptual transparency condition is satisfied, there is no absolute necessity to place zeros too close to each other.

4.2. Experiments in Noisy Medium

Typical SNR values for testing the robustness of an audio watermark with the zero locations defined above were chosen to be as the ones in [11].

As expected, bit reliability increases with decreasing SNR. It reaches up to 80% with SNR = 50 dB while it is around 45% with SNR = 20dB. Choosing frame size to be 128 and stage number to be 2 produces better results. On the other hand performance of the method in male voice with pauses was not as good as continuous utterances and music.

As channel noise, we used a recording of sound on a voiceless wireless telephone channel. Such a channel noise is added on top of the watermarked signal and decoding performance is examined. It is observed that only when zeros are too close, i.e. 1% – 3% apart on the unit circle bit reliability decreases to 71% - 96%, otherwise it is 100% for any stage number or frame size selection.

In order to determine the tolerable SNR values, white Gaussian noise is added on top of the marked signal and values that lead to at least 95% bit reliability are assumed to be safe.

It is observed that, as the frame size increases, the detection procedure gets more sensitive to add-on noise when the zeros are located around the lower frequency values. On the other hand, zeros that are around the higher frequencies lead to a detection scheme that is less sensitive to channel noise. Especially in case of smaller frame sizes, the zeros should be placed apart from each other in order to increase bit reliability.

4.3. Experiments with Signals under Attack

Watermarked signals are converted to MP3 format at 192 Kbps and regular detection scheme is employed. It is observed that compression causes a fair decrease in bit reliability. Bit reliability changes between 40% and 50% for all cases. Best results are obtained with frame size equal to 256 and with a stage number selection of 2.

Estimation type of attacks assume that the watermark can be estimated without prior knowledge of the embedding rule or embedding keys, the watermark is considered to be noise and a denoising scheme is employed [12].

Low pass filtering seems a common approach for denoising audio signals. Our proposed method is in fact fragile under low pass filtering attack and hence needs to be strengthened. One counter measure against low pass filtering is to employ a notch filter on a low frequency component in wavelet decomposed signal. A notch filter of narrow stop band can be shown not to degrade the perceptual quality of the signal. Moreover, the watermark embedded by the notch filter can be detected efficiently. Our watermark extraction experiment with different types of speech signals indicate about 90% against low pass filtering bit reliability provided the stop band of the notch filter is determined taking into account the frequency characteristics of the set of signals to be watermarked.

5. Conclusions

Our method outperforms the previously proposed techniques in several respects. First of all, it is obvious that memory requirement is much lower than the algorithms in which the

storage of the watermark, the original signal, and several security keys is required. Our approach requires only the storage of frequencies f_1 and f_2 , the stage number, the frame size, and the watermark, which consists of a text of 4-7 letters. Decoding is done with the help of these five parameters only. Since the last stage detail coefficients of the wavelet-decomposed signal are cancelled during encoding, there is a built-in compression in our watermarking scheme. Security keys, i.e., f_1 and f_2 are impossible to predict from the marked signal even if the stage number and the frame size somehow became available to the pirate. Finally, the recovery of the original from the watermarked signal is impossible even if the security keys are available. If a pirate is aware of the watermarking scheme and inserts his information on to a marked signal, there is no way he can erase the effect of the first watermark. Hence, in our scheme, deadlock issue is not a concern.

Using wavelets in watermarking is not a new idea. However, usually signals that are of the same kind as the input signal are used as watermark. The method proposed here is fundamentally different from these as the watermark itself consists of using different filters in synthesizing the watermarked signal. The extension of this method to image watermarking seems straightforward and is currently under investigation.

Acknowledgements: The authors would like to thank Dr. E. Çetin for suggesting this watermarking application to zero-assigned filter banks.

References

- [1] M. D. Swanson, B. Zhu, A.H. Tewfik, and L. Boney, "Robust Audio Watermarking Using Perceptual Masking", *Signal Processing*, vol. 66, pp. 337-355, 1998
- [2] I. J. Cox, J. Killian, F. T. Thomson, T. Shamoan, "Secure Spread Spectrum Watermarking for Multimedia", *IEEE Transactions on Image Processing*, vol. 6, No. 12, 1997
- [3] Y. Wang, J. Doherty, and R.E. Van Dyck, "A Wavelet Based Algorithm for Ownership Verification of Digital Images", *IEEE Transactions on Image Processing*, vol. 11, No. 2, Feb. 2002
- [4] M. D. Swanson, M. Kobayashi, and A.H. Tewfik, "Multimedia Data-Embedding and Watermarking Technologies", *Proceedings of the IEEE*, vol. 86, No. 6, 1998
- [5] K. Ratakonda, R. Dugad, and N. Ahuja, "Digital Image Watermarking: Issues in Resolving Rightful Ownership", *International Conference on Image Processing ICIP 98. Proceedings 1998*, vol.2, pp. 414-418. 1998, Chicago, Illinois, USA
- [6] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, Englewood Cliffs, NJ, 1993.
- [7] M. Akbaş, "Zero-assigned Filter Banks and Wavelet", M. Sc. Thesis, Electrical and Electronics Engineering Department, Bilkent University, Bilkent, Ankara, 06800 Turkey, 2001.
- [8] M. J. Lai., "On the Digital Filter Associated with Daubechies Wavelet", *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2203-2205, 1995
- [9] I. Daubechies, 'Ten Lectures on Wavelets', Philadelphia, PA., SIAM, 1992
- [10] Z. Yücel and A. B. Özgüler, "Speech Watermarking via Zero Assigned Filter Banks", Report, Electrical and Electronics Engineering Department, Bilkent University, Ankara, Turkey, 2005.
- [11] L. Xueyao, M. Zhang, R. Zhang, "A New Adaptive Audio Watermarking Algorithm", *Proceedings of the 5th World Congress on Intelligent Control and Automation*, June 15-19, 2004, Hang Zhou, P.R. China
- [12] S. Voloshynovskiy, S. Pereira, T. Pun, J. Eggers, J. K. Su, 'Attacks on Digital Watermarks: Classification, Estimation-Based Attacks, and Benchmarks', *IEEE Communications Magazine*, vol.39, issue 8, pp. 118-126, August 2001