Adaptive Speech Watermarking in Wavelet Domain based on Logarithm

Mehdi Fallahpour[1], David Megias[1] and Hossein Najaf-Zadeh[2]
[1]Estudis d'Informàtica, Multimèdia i Telecomunicació, Internet Interdisciplinary Institute (IN3),
Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018 Barcelona, Spain
[2]Advanced Audio Systems, Communications Research Centre Canada (CRC),
3701 Carling Ave., Ottawa, K2H 8S2, Ontario, Canada

Keywords: Speech Watermarking, Data Hiding, Wavelet Transform, Logarithm.

Abstract: Considering the fact that the human auditory system requires more precision at low amplitudes, the use of a logarithmic quantization algorithm is an appropriate design strategy. Logarithmic quantization is used for the approximation coefficients of a wavelet transform to embed the secret bits. To improve robustness, the approximation coefficients are packed into frames and each secret bit is embedded into a frame. The experimental results show that the distortion caused by the embedding algorithm is adjustable and lower than that introduced by a standard ITU-T G.723.1 codec. Therefore, the marked signal has high quality (PESQ-MOS score around 4.0) and the watermarking scheme is transparent. The capacity is adjustable and ranges from very low bit-rates to 4000 bits per second. The scheme is shown to be robust against different attacks such as ITU-T G.711 (a-law and u-law companding), amplification and low-pass RC filters.

1 INTRODUCTION

Practical applications for digital watermarking vary from copy prevention and traitor tracing to broadcast monitoring or archiving, among others. Watermarking can be used to identify the license information, owners, or other information related to the digital object carrying the watermark. Imperceptibility, robustness and capacity are three important conflicting properties which are used to evaluate watermarking systems.

Speech watermarking systems usually embed watermarks in inaudible parts of the speech signals. Many speech watermarking and information embedding schemes have been proposed. These methods can be classified into seven approaches: least significant bit, phase coding, echo hiding analysis by synthesis-based, spectrum techniques in the transform domain, feature-based and watermarking combined with coding frames.

Several quantization-based methods have been proposed in the recent years by using discrete Hartley transform coefficients (Sagi and Malah, 2007), autoregressive model parameters (Chen and Leung, 2007), and the pitch period (Celik et al., 2005). The payload range of those systems is from a few bits to a few hundred bits per second, with varying robustness against different types of attacks.

The proposed method takes advantage of the wavelet transform, which divides the signal into low and high frequency bands. The signal-to-noise ratio (SNR) values in the low frequency region are in the range of 15 to 20 dB, and gradually decreases to zero as the frequency increases. To achieve robustness, using the low frequency bands is more advisable, whereas embedding the watermark in high frequency bands leads to better transparency.

The human auditory system requires more precision (in terms of absolute errors) at low-energy audible amplitudes (*i.e.*, audible soft sounds), but is less sensitive at higher amplitudes. Considering this fact, and making use of the logarithm, a logarithmic quantization algorithm is used for approximating the *cA* coefficients of a wavelet transform to embed the secret bits. To improve robustness, the *cA* samples are grouped into frames and a single secret bit is embedded into the corresponding frame. Increasing the frame size decreases the embedding capacity and increases the robustness.

The experimental results show that the distortion caused by the embedding algorithm is adjustable and lower than that caused by the ITU-T G.723.1 speech

Fallahpour M., Megias D. and Najaf-Zadeh H..
Adaptive Speech Watermarking in Wavelet Domain based on Logarithm.
DOI: 10.5220/0004091704120415
In *Proceedings of the International Conference on Security and Cryptography* (SECRYPT-2012), pages 412-415
ISBN: 978-989-8565-24-2

codec (ITU-T, Recommendation G.723). G.723 is a standard speech codec that guarantees the quality of compressed speech and thus, it is evident that the marked signal has high quality (PESQ-MOS around 4, *i.e.*, near transparent). The embedding rate is adjustable and ranges from very low bit-rates to 4000 bits per second (bps).

The rest of the paper is organized as follows. Section 2 introduces the proposed method. In Section 3, a discussion on the transparency and robustness of the proposed scheme is provided and the experimental results are presented. Finally, Section 4 summarizes our work with a conclusion.

2 PROPOSED SCHEME

The proposed scheme includes two methods, *i.e.*, the embedding and the extracting processes.

2.1 Embedding Process

To embed secret information into the approximation coefficients of the wavelet transform (*cA*), these coefficients are divided into frames and each single secret bit is embedded into the corresponding frame. Each wavelet coefficient is mapped into the logarithm domain and changed depending on the secret bit. The embedding steps are described below:

1. Compute the first level Daubechies wavelet transform (db10) of the original signal.

2. Divide the *cA* samples into frames of a given size (*f*), where *cA* is the approximation of the input signal (*i.e.*, the output of the low-pass filter of the wavelet transform).

3. Assume that w_j is the *j*th secret bit embedded into the *j*th frame of *cA*. Then let

$$m_i = \begin{cases} \operatorname{sign}(c_i)10^{\delta\left\lfloor\frac{(\log_{10}|c_i|)}{\delta}\right\rfloor}, & \text{if } w_j = 0, \\ \operatorname{sign}(c_i)10^{\delta\left\lfloor\frac{(\log_{10}|c_i|)}{\delta}\right\rfloor+\frac{\delta}{2}}, & \text{if } w_j = 1. \end{cases}$$

m_i is the marked coefficient, c_i represents a *cA* coefficient, δ is the quantization value and $j = \lfloor (i-1)/f \rfloor$. For example when $j = 1$ and $f = 4$, w_1 (the first secret bit) is embedded into c_1, c_2, c_3 and c_4, then the second secret bit (w_2) is embedded into the second frame (c_5, c_6, c_7 and c_8). This process is repeated for the remaining secret bits. Finally, the inverse DWT is applied to the marked wavelet coefficients to obtain the marked audio signal in the time domain.

2.2 Extracting Process

In the extracting process, we first obtain the wavelet coefficients of the marked signal. After that, the difference between the marked samples and the rounded marked samples is computed to extract a secret bit. The extracting steps are listed below:

1. Compute the first level Daubechies wavelet transform (db10) of the marked signal.

2. Extract the bit embedded into each marked wavelet coefficient based on the following equation:

$$z_i = \begin{cases} 0, & \text{if } t(\delta \log_{10}|m_i|) < 0.25\,, \\ 1, & \text{otherwise,} \end{cases}$$

where $t(x) = \big||x| - \operatorname{round}(x)\big|$.

To make a decision about an embedded bit in each frame, a voting algorithm is used. It means that for each frame, we count the number of zeros and ones in the extracted values (*i.e.*, $z_i's$). The exact embedded secret bit in that frame will be the value with the highest count. For instance, if the frame size is $f = 5$ and the number of $z_i's$ with a value of "1" is more than two, then the embedded bit in the frame is "1"; otherwise it is "0".

3 EXPERIMENTAL RESULTS AND DISCUSSION

Four male speech files sp01.wav – sp04.wav and five female files sp11.wav – sp15.wav taken from the Noizeus speech corpus (Hu and Loizou, 2007) have been selected for our experiments. The sampling frequency is 8000 Hz and each sample is represented with 32 bits. In a watermarking system, we have different properties and, among them, capacity, transparency, and robustness are the most relevant ones. In this section, we discuss these properties for the proposed watermarking scheme.

3.1 Capacity

Capacity is defined by the number of bits embedded in one second of the speech file. For different applications, different ranges of capacity are demanded. For example for copyright protection, we just need to embed a short identification code and therefore the required capacity is about 100 bps, however in this kind of application robustness against various attacks is essential.

The capacity of the proposed scheme can be modified adaptively according to the requirements. The embedded capacity can be adjusted using the frame size *f*. This is a relevant feature of the proposed scheme. When the frame size is one, the embedded capacity is very high (*i.e.*, 4000 bps). For instance, when the sampling rate of a speech file is 8000 samples per second, there are 4000 coefficients available in the approximation part (*cA*) for embedding the secret bits.

3.2 Transparency

In general, it is difficult to prove that the distortion caused by coding, watermarking or other processing operations is imperceptible. In fact, as the perceptual properties of the human auditory system are quite complex, it is difficult to map them into linear equations and prove transparency.

In the proposed scheme, speech samples are changed based on a logarithmic function, which is a common technique used by several speech codecs. We have obtained some results showing that the distortion introduced by watermarking is lower than the coding distortion introduced by a G.723 codec (the corresponding plots results are omitted here for space limitations). As G.723 is a standard speech codec with guaranteed high quality, the perceptual transparency of this scheme is guaranteed as well. Experimental results using the Perceptual Evaluation of Speech Quality (PESQ) objective measurement (ITU-T, Recommendation P.861) show that the PESQ-MOS is around 4. PESQ results principally model mean opinion scores (MOS), which cover a scale from 1 (bad) to 5 (excellent).

3.3 Robustness

A trade-off between capacity and robustness is always a challenge for audio watermarking systems. High capacity usually results in a very fragile method and, conversely, robust schemes lead to very low capacity. Repeating a single bit is a simple but effective idea to increase robustness. Consider that we face a situation where a burst error destroys 10 ms of the speech signal. If we just embedded the secret information in the missing part, we would loose a relevant part of important information. On the other hand, if the secret information were repeated in other places, we would not loose all the secret bits due to the burst error.

In the proposed scheme, we just embed each single bit into a frame, *i.e.*, the same secret bit is embedded into all the coefficients in that frame. Thus, if we were not able to extract the secret bit from some coefficients, this bit may be extractable from other coefficients in the same frame. Finally, a voting technique leads us to extract the embedded bit in the frame. As mentioned above, considering a trade-off between capacity and robustness is necessary. For instance, if the frame size is $f = 4$, capacity is decreased by a factor of 4 (compared to $f = 1$) but, in return, robustness is increased. Thus, the parameters of the scheme should be chosen according to the demands and the specific application.

Table 1 shows the capacity and transparency for the test files with different parameters. By increasing δ (quantized value), the distortion decreases. In all the results, PESQ-MOS is around 4, which means that the quality of the marked signal is high. In addition, changing the frame size affects the capacity (800–4000 bps).

Table 2 shows the robustness of the proposed method against different attacks. The results were obtained with two values of δ; for $\delta = 5$, the PESQ-MOS is 3.5 and for $\delta = 3$, the PESQ-MOS is 3.1; in both cases frame size is $f = 4$ and thus, the capacity is 1000 bps. The scheme has been tested against the attacks provided in the Stirmark Benchmark (Lang, A., Stirmark Benchmark for Audio), where each attack has different parameters defined in the benchmark's website. As expected, decreasing δ increases robustness (and distortion). This table illustrates that this technique is robust against the G.711 codec.

Table 1: Results of 2 signals (robust against table 2 attacks).

Audio File	δ	Frame size (*f*)	PESQ-MOS of marked	Payload (*bps*)
GH (Sp01–Sp04)	5	1	3.4	4000
		4	3.4	1000
	10	1	4.1	4000
		4	4.1	1000
JE (Sp11–Sp15)	6	2	3.5	2000
		5	3.5	800
	12	2	4.0	2000
		5	4.0	800

Table 2: Robustness test results for selected files.

	$\delta = 5$		$\delta = 3$	
Attack name	**Params.**	**BER**	**Params.**	**BER**
Amplify	60–170	0.00	10–250	0.00
AddDynNoise	10	0.05	20	0.05
AddNoise	10	0.05	30	0.05
FFT_Invert	1024	0.00	1024	0.00
RC Low pass filter	2000	0.09	2000	0.07
RC High pass filter	50	0.05	50	0.03
G.711 a-law	–	0.05	–	0.02
G.711 u-law	–	0.01	–	0.01

In Table 3, we compare the performance of the proposed watermarking algorithm and several recent speech watermarking strategies. In (Sagi and Malah, 2007), the MOS of the narrow band (NB) speech is 3.7 and the MOS of the NB speech with embedded data is 3.625. The small difference between the MOS results demonstrates the transparency of the proposed data-embedding scheme. In simulations, the embedding data rate is 600 information bits/second. The method of (Celik et al., 2005) allows a relatively low embedding capacity (about 3 bps), which is suitable for metadata tagging and authentication applications. However, (Celik *et al.,* 2005) is robust with low data-rate (5-8 kbps) speech coders. The focus of (Gurijala and Deller, 2007) is on the robustness performance of linear prediction embedded speech watermarking. The technique is robust to a wide range of attacks including noise addition, cropping, compression, and filtering, but the achieved capacity is low.

Table 3: Comparison of different speech watermarking algorithms.

Algorithm	**SNR (dB)**	**PESQ-MOS**	**Payload (bps)**
(Sagi and Malah, 2007)	35	3.6	600
(Celik et al., 2005)	–	–	3
(Girin and Marchand, 2004)	High	–	200
(Gurijala and Deller, 2007)	–	–	24
Proposed	30–40	~ 4	800–4000

4 CONCLUSIONS

Using the wavelet transform and a logarithmic quantization results in an adaptive speech watermarking scheme. Considering the fact that the human auditory system requires more precision at low amplitudes (soft sounds) and taking advantage of the logarithm, a logarithmic quantization algorithm is used to quantize the approximation coefficients of the wavelet transform (*cA*) to embed the secret bits. To improve robustness, the *cA* samples are split into frames and each single secret bit is embedded into all the samples in the corresponding frame. Increasing the frame size decreases the embedding capacity and increases the robustness.

The experimental results show that the distortion caused by the embedding algorithm is adjustable and lower than that introduced by the G.723 speech codec. Therefore, the marked signal has high quality (PESQ-MOS around 4), *i.e.* the proposed watermarking scheme is transparent. The embedding rate is adjustable and can start from very low bit-rates to 4000 bps, depending on the application. The scheme is shown to be robust against some attacks such as ITU-T G.711 compression (a-law and u-law companding), amplification and RC filters.

ACKNOWLEDGEMENTS

This work was partially funded by the Spanish Government through projects TSI2007-65406-C03-03 E-AEGIS, TIN2011-27076-C03-02 CO-PRIVACY and CONSOLIDER INGENIO 2010 CSD2007-0004 ARES.

REFERENCES

Celik, M., Sharma, G., Tekalp, A. M., 2005. Pitch and duration modification for speech watermarking. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2, pp. 17–20.

Chen, S., Leung, H., 2007. Speech bandwidth extension by data hiding and phonetic classification. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 4, pp. 593–596.

Girin, L., Marchand, S., 2004. Watermarking of speech signals using the sinusoidal model and frequency modulation of the partials. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 1, pp. 633–636.

Gurijala, A., Deller, J., 2007. On the robustness of parametric watermarking of speech. In *Multimedia Content Analysis and Mining, ser. Lecture Notes in Computer Science*, Springer, vol. 4577/2007, pp. 501–510.

Hu, Y., Loizou, P., 2007. Subjective evaluation and comparison of speech enhancement algorithms. *Speech Communication*, 49, 588-601.

ITU-T, Recommendation P.861. http://www.itu.int/rec/T-REC-P.861/en (accessed on June 22nd, 2012).

ITU-T, Recommendation G.711. http://www.itu.int/rec/T-REC-G.711/en (accessed on June 22nd, 2012).

ITU-T, Recommendation G.723. http://www.itu.int/rec/T-REC-G.723/en (accessed on June 22nd, 2012).

Lang, A., Stirmark Benchmark for Audio. http://wwwiti.cs.uni-magdeburg.de/~alang/smba.php (accessed on June 22nd, 2012).

Sagi, A., Malah, D., 2007. Bandwidth extension of telephone speech aided by data embedding. *EURASIP J. Adv. Signal Process.*, vol. 2007, article ID 64921.

Salomon, D., 2007. *Data Compression: the Complete Reference*. Springer.