International Journal of Advanced Science and Technology
Vol.112 (2018), pp.57-66
http://dx.doi.org/10.14257/ijast.2018.112.06

Application of QIM Based Audio Watermarking for Synchronizing Detection in the Air Environment

Donghwan Shin[1] and Youngseok Lee[2*]

[1]*MarkAny Research Institute, MarkAny*
[2*]*Dept. of Electronics, Chungwoon University*
[1]*dhshin@markany.com,* [2*]*yslee@chungwoon.ac.kr*

Abstract

Automatic recognition and accurate synchronization between TV programming and interactive applications running on smart TVs and second screen devices - Significantly improve your viewing experience on your TV. ACR is bringing new possibilities to the broadcasting industry. In this paper, we have confirmed through experiments that an audio watermarking algorithm using QIM method can be applied in ACR environment where watermark is extracted from audio signal in the air. From the experimental results, it can be confirmed that the normalization of the audio signal size must be preceded when the watermark is inserted and extracted in order to use the QIM method in the ACR environment.

Keywords*: Audio watermarking, Automatic content recognition, Quantization index modulation*

1. Introduction

In recent years, many media devices have been regularly in our living rooms. This space was primarily used for watching television, but it evolved into a shared space using other devices such as laptops, tablets, smart phones or gaming handhelds. Automatic content recognition facilitates content consumption in the living room. And it improves interaction, monetization and creativity. The way people watch TV is changing rapidly. Previously it was a passive experience, but now it is an interactive and engaging experience.

Consumers use a second screen device such as a smartphone, tablet, or laptop to view, interact, and engage with a variety of secondary screen applications, websites, and communities while watching TV. Next-generation TVs, like the second screen device, are connected to access a wide variety of first screen applications. According to a recent study by Deloitte, almost half of the 16-24 year olds are using messaging, email, Facebook or Twitter to discuss what they are watching on TV. In fact, 24% of all respondents use the second screen. This desire for consumer interaction opens up opportunities for broadcasters, content owners, and ad agencies to deepen their relationships with consumers. Adding interactive applications to companion devices and smart TVs will meet the needs of today's media consumers and provide new revenue opportunities for targeted applications and advertisings. Automatic Content Recognition (ACR) supports most of these next generation interactive applications.

Automatic recognition and accurate synchronization between TV programming and interactive applications running on smart TVs and second screen devices - Significantly improve your viewing experience on your TV. ACR is bringing new

Received (October 9, 2017), Review Result (December 19, 2017), Accepted (January 26, 2018)

ISSN: 2005-4238 IJAST

possibilities to the broadcasting industry. The use of ACR can be divided into the following three categories.

Content Identification: Audiences can easily find information about content they've viewed using ACR technology. For applications with Smart TV and ACR technology, viewers can see the name of the song being played or a description of the movie they watched. In addition, identified video and music content can be linked to Internet content providers for on-demand viewing, and third party or complementary media for additional background information.

The first for broadcast Monitoring, It's important for advertisers and content owners to know when and where their content plays. Traditionally, an agency or advertiser must manually audit a presentation. On the scale, only statistical sampling methods are available. ACR technology automatically monitors content played on your TV. Information such as playback time, duration, and frequency can be done without manual intervention.

The second for content Enhancement, because the device can "see" the content being watched or received, the second screen device can provide users with more complementary content than what is shown on the main viewing screen. ACR technology not only identifies the content, but also identifies the exact location within the content. Thus, additional information may be presented to the user. ACR can use a variety of interactive features such as polls, coupons, lotteries, or purchases based on timestamps.

The last for measure your Audience Measurement, You can now apply ACR technology to mobile devices such as smart TVs and set-top boxes, smart phones and tables to implement real-time audience measurement metrics. This metric is critical to quantifying audience spending to set up advertising pricing.

Audio-based ACR is commonly used in the market. The two main methodologies are acoustic fingerprinting and watermarking. There is an alternative approach that focuses on video fingerprinting but improves accuracy and scalability with other content-aware solutions running in parallel and continuously.

Acoustic fingerprints generate unique fingerprints from the content itself. Fingerprinting technology works regardless of content type, codec, bit rate, and compression technology. Available via network and channel. Therefore, it is widely used in the field of interactive TV, second screen application and content monitoring. Popular apps like Civolution, Digimarc and Facebook. We chat and Weibo uses the audio fingerprinting methodology to recognize content played on the TV and launch additional features such as voting, lotteries, topics or purchases.

Unlike fingerprinting, digital watermarking requires that you insert a digital tag in the content itself that contains information about the content before distribution. For example, a broadcast encoder can insert watermarks every few seconds that it can use to identify broadcast channels, program IDs, and timestamps. Watermarks are not generally audible or visible to the user. Terminal devices, such as cell phones and tablets, read the watermark instead of actually seeing what's playing. Watermarking technology is used to track where piracy occurs in the field of media protection. Next / Market Insights expects 2.5 billion devices to integrate with ACR technology to deliver real-time live video and on-demand video viewing experiences.

Since the position of the inserted information is arbitrary in the audio signal obtained from the ACR environment, it is most important to find the starting position where information is inserted in order to obtain meaningful information. The bit information that performs this role is called the synchronization information bits. In this paper, we describe how to insert and extract information related to synchronization when inserting and extracting information in audio in ACR environment.

2. Related Works

In the many researches, there are a few algorithms focusing on solving desynchronization problems. For cropping attacks (such as editing, signal interruption in wireless transmission, and data packet loss in IP network), researchers repeatedly embedded a template related with synchronization into different regions of the signal [1]-[7], such as synchronization code-based self-synchronization methods and feature- based methods [8]-[12] and the use of multiple redundant watermarks [13], [14].

Template-based watermarking can be confronted with cropping, but it cannot cope with TSM (Time Scale Modification) operations, even with a scaling amount of ± 1%. The audio watermarking community has TSM resilience watermarking strategies such as peak point based [15] - [18] and recent histogram based [19], [20].

The bits can be hidden by quantizing the length of each two adjacent peak points [16]. In [17], the watermark was repeatedly embedded at the edge of the audio signal by removing and adding a portion of the audio signal while preserving the pitch and viewing the pitch-invariant TSM as a special form of random cropping. In [18], the invariance of the binomial wavelet transform with linear scaling was used to design audio watermarking by modulating the wave shape.

Three main peak point based watermarking methods are resistant to TSM because peaks can still be detected before and after TSM operation. The histogram-based method is possible because the histogram form of the audio signal is invariant to time-linear scaling. The histogram is also independent of the position of the sample in the time domain.

Above watermark algorithms only consider watermark attacks under a digital environment. The effects of analogue transmission channels through DA / AD conversion are seldom mentioned.

In this article, this article proposes a solution for DA / AD conversion taking into account the degradation of the conversion (empirically proven by a combination of volume change, additional noise and small TSM). First, a relationship-based watermarking strategy is introduced for volume changes by modifying the relative energy relationship between groups of three consecutive DWT coefficient sections. Second, the watermark is embedded in the low-frequency sub-band for additive noise.

Third, the synchronization strategy by the interpolation processing operation through the search of the synchronization code is applied to the TSM. Experimental results show that the proposed watermarking algorithm is robust to DA / AD conversion, is robust to general audio processing operations, and is resistant to most attacks in ACR circumstance.

3. Proposed Synchronization Method

In order to embedding and extracting watermarks in an ACR environment it is important to process information related to synchronization. ACR environment can be modeled as compound attack of cropping attack and A/D and D/A attack among various watermarking attacks. Therefore, the watermark format for embedding can be expressed as shown in Figure 1, which is divided into a part for embedding synchronization information for cropping attack and a part for storing information related to underlying audio content.

The scheme of Figure 1 is to improve the robustness against cropping attack and detectability when it loses synchronization, audio segment is used for at first, and then, synchronization code and watermark embedded into each segment. Our embedding scheme follows in [21] except for normalization of amplitude at each frame. The embedding scheme of synchronization code is following as described in [21].

Quantization index modulation (QIM) methods, a class of nonlinear methods that we describe in this paper, reject this host-signal interference. As a result, these methods have very favorable performance characteristics in terms of their achievable trade-offs among the robustness of the embedding, the degradation to the host signal caused by the embedding, and the amount of data embedded.

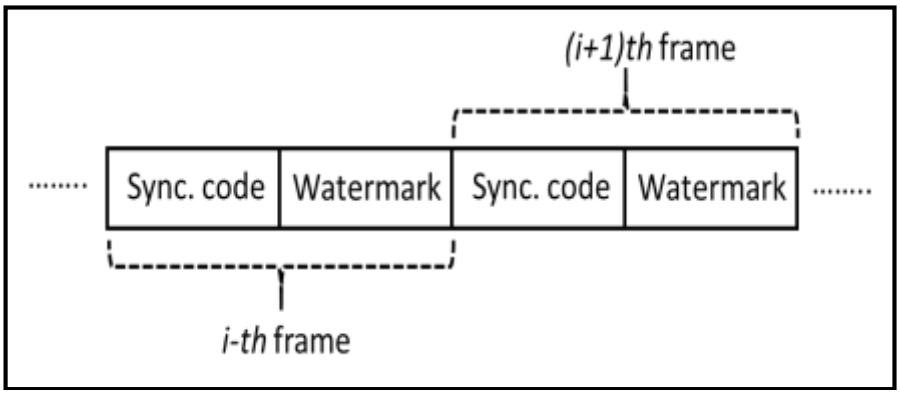

Figure 1. Watermark Format to Embedding Information for ACR

But in our method, amplitude normalization is performed at each frame and embeds synchronization code which is different from the method in [21]. Our method considers amplitude attenuation by propagation path in ACR environment. The ACR environment typically considered is airborne. In an ACR environment, we can assume the following in a scenario physical environment in which an audio signal is acquired, a watermark is extracted therefrom, and related information is acquired. The sound emitted through the multimedia device or the speaker is acquired through the microphone of the mobile phone and then processed. In the process of acquiring an audio signal, the power of the original audio signal decreases in inverse proportion to the square of the distance, and resampling occurs in the process of acquiring the sound. These effects in the air can be understood as causing of distortions such as resampling and power reduction of original audio signal as watermark carrier.

In embedding process, let A_1^0 is cut into L_{syn} audio segments, and each audio segment $P_tA_1^o(m)$ having *n* samples, where $L_1 = n \times L_{syn}$. The normalized form of audio segment $P_tA_1^o(m)$ having *n* samples is presented as (1)

$$\mathrm{P}A_1^0(m) = P_tA_1^0(m)/\max(\left|P_tA_1^0(m)\right|) \quad (1)$$

In (1), $\mathrm{P}A_1^0$ is described by sample by sample as

$$PA_1^0(\mathrm{m}) = \mathrm{pa}_1^0(m)(i) = \mathrm{a}_1^0(i + m \times n),\ 0 \le i < n, 0 \le m < L_{syn} \quad (2)$$

Notation of $PA_1^0(\mathrm{m})$ in (2) is also used in [21], while the notation of [21] represents the samples in the frame, in this paper it mean the samples in the normalized frame. At the second step, mean value $\overline{PA_1^0(\mathrm{m})}$ is calculates as following:

$$\overline{PA_1^0(\mathrm{m})} = \frac{1}{n}\sum_{i=0}^{n-1} pa_1^0(m)(i), (0 \le m < L_{syn}) \quad (3)$$

Using (1) and (2), the synchronization code is embedded into each $\mathrm{P}A_1^0(m)$ by quantization value of the mean value $\overline{PA_1^0(\mathrm{m})}$, the rule is given by

$$\mathrm{pa}_1^{0*}(m)(i) = \mathrm{pa}_1^0(m)(i) + (\overline{PA_1^{0*}(\mathrm{m})} - \overline{PA_1^0(\mathrm{m})}) \quad (4)$$

where $PA_1^0(\mathrm{m}) = \{\mathrm{pa}_1^0(m)(i), 0 \le i < n\}$ is original sample, $PA_1^0(\mathrm{m}) = \{\mathrm{pa}_1^0(m)(i), 0 \le i < n\}$ is modified sample, and

$$\overline{PA_1^0(\mathrm{m})} = \begin{cases} IQ\left(\overline{PA_1^0(\mathrm{m})}\right) \times S_1 + \frac{S_1}{2}, \ for\ Q\left(\overline{PA_1^0(\mathrm{m})}\right) = f(m) \\ IQ\left(\overline{PA_1^0(\mathrm{m})}\right) \times S_1 - \frac{S_1}{2}, \ for\ Q\left(\overline{PA_1^0(\mathrm{m})}\right) \neq f(m) \end{cases} \quad (5)$$

In (5), $IQ\left(\overline{PA_1^0(\mathrm{m})}\right)$ is defined as

$$IQ\left(\overline{PA_1^0(\mathrm{m})}\right) = \left\lfloor \frac{IQ\left(\overline{PA_1^0(\mathrm{m})}\right)}{S_1} \right\rfloor \quad (6)$$

And also $Q\left(\overline{PA_1^0(\mathrm{m})}\right)$ is defined ad

$$Q\left(\overline{PA_1^0(\mathrm{m})}\right) = mod\left(IQ\left(\overline{PA_1^0(\mathrm{m})}\right), 2\right), \quad (7)$$

where $mod(x, y)$ returns the remainder of division of x and y, S_1 is the quantization step.

In extraction process the unit of watermarked frame is calculated as following:

$$\mathrm{Watermark(m)} = \mathrm{mod}(\overline{PA_1^0(\mathrm{m})}, 2) \quad (8)$$

As in [21], we use Barker code to synchronize watermark audio. Barker code is a sort of binary pseudo random code that are commonly used for frame synchronization in digital communication. The characteristic of Barker code is that side-lobe of autocorrelation is small and has high correlation value. The side-lobe of correlation, C_k for a k –symbol shift of N-bit code sequence $\{x_j\}$ is expressed as

$$C_k = \sum_{j=1}^{N-k} x_j x_{j+k} \quad (9)$$

4. Experimental Results

The proposed method is applied to arbitrary audio. Figure 2 shows the original audio and watermarked audio. It can be observed that there is no difference from the visual point of view. We have found that the proposed method is influenced by the number of samples to which the quantization index is applied. Through experiments, we confirmed that the algorithm operates normally only when the number of samples constituting the audio segment of $PA_1^0(m)$ expressed in Equation (1) is fixed to 5.

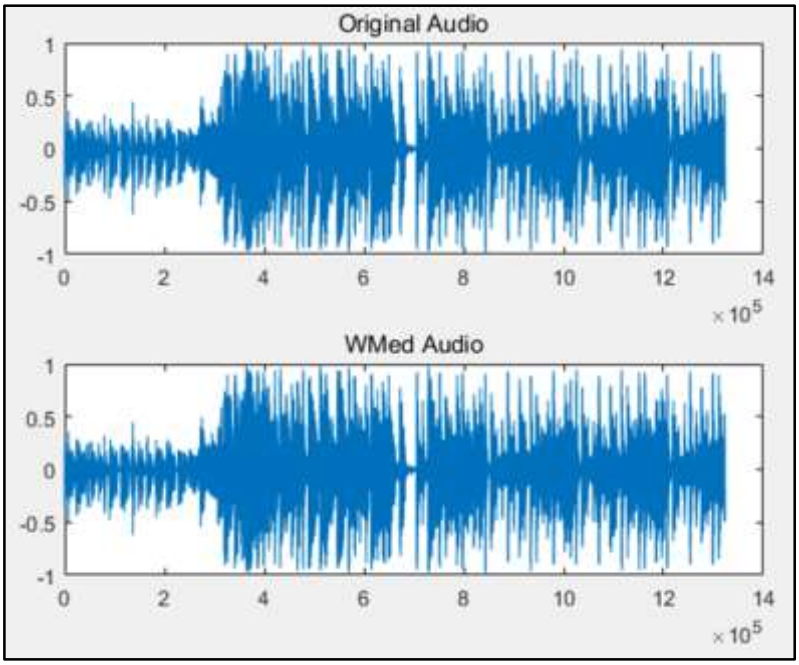

Figure 2. Comparison of Digital Original and Watermarked Audio

For example, when embedding and extracting a watermark with 9 samples per frame, we can observe that a very high bit error rate (BER) occurs as shown in Figure 3. The same result can be observed even when the number of samples per frame is different. The results in Figure 3 can be generalized according to Table 1. Table 1 shows the BER that occurs when a watermark is extracted by varying the number of samples per frame.

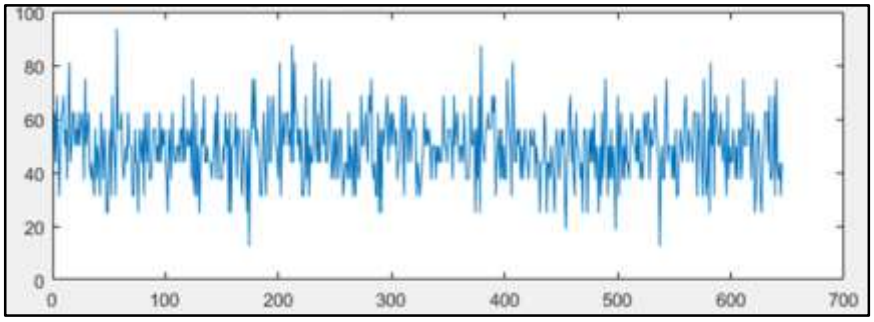

Figure 3. BER at Frames in Sample Number= 9

As shown in Table 1, except for the case where the number of samples per frame is 5, the BER of a watermark extracted from the number of samples per remaining frame represents an average of 50%. This result shows the typical characteristics of the QIM based watermarking algorithm.

Table 1. BER by Sample Number in Quantization Step Index

Sample Number	4	5	6	7	8	9
BER(%)	48.9	00.0	54.1	53.6	48.8	49.3

In order to apply the proposed method to the ACR environment in the air, the attenuation of the audio must be considered. In general, the power of a sound signal is known to be inversely proportional to the square of the distance. In this study, we attempted to extract the watermark from the audio signal whose audio power was reduced by $\sqrt{10}$ to assume that the audio signal was recorded at a distance of 10% of the volume of the sound.

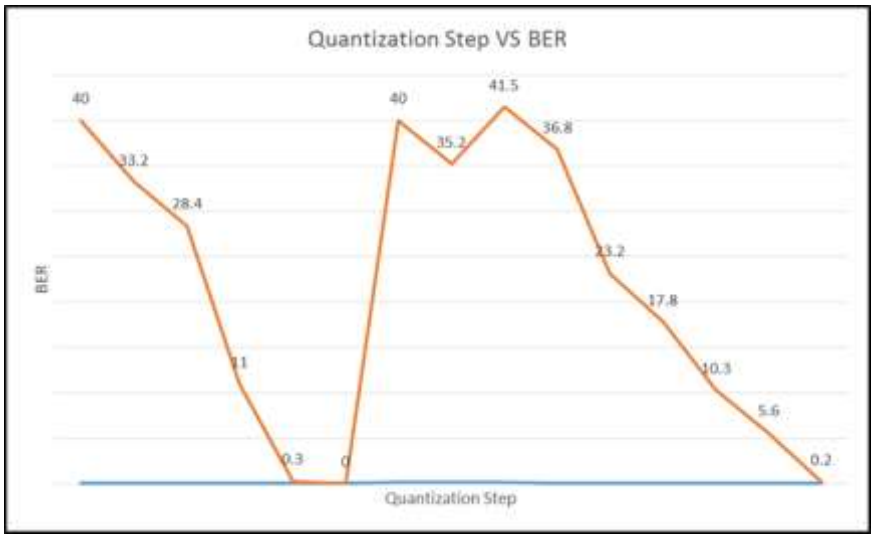

Figure 4. BER by Quantization Step Scale at 10% Attenuated Audio

In this experiment, the insertion strength of the watermark is fixed to 0.028, and the insertion strength of this degree is such that the watermarked audio signal can be heard without disturbance.

Table 2. BER by Quantization Step Scale at 15% Attenuated Audio

Quantization Step Scale										
0.145	**0.146**	**0.147**	**0.148**	**0.149**	**0.150**	**0.151**	**0.152**	**0.153**	**0.154**	**0.155**
BER (%)										
40.0	**0.15**	**00.3**	**00.0**	**00.0**	**00.0**	**00.0**	**00.0**	**00.2**	**01.3**	**02.8**

Figure 4 shows the BER of the watermark extracted from the attenuated audio signal. As shown in the figure, the quantization step size should be reduced by reducing the audio volume so that watermarks can be accurately extracted.

The same result can be obtained by applying the result of Figure 4 to the audio signal of which the size is reduced by 15%. Table 2 shows the BER of the extracted watermark by applying a 15% reduction in the quantization step size value and the surrounding values to the audio signal of which the size is reduced by 15%. Table 2 shows that the BER increases to 84.9%, 84.85, 85.1% and 85.2%, including 85.0% of the original quantization step size, but the BER increases in other quantization step sizes. In particular, we can observe that the BER increases gradually when the quantization step size becomes smaller.

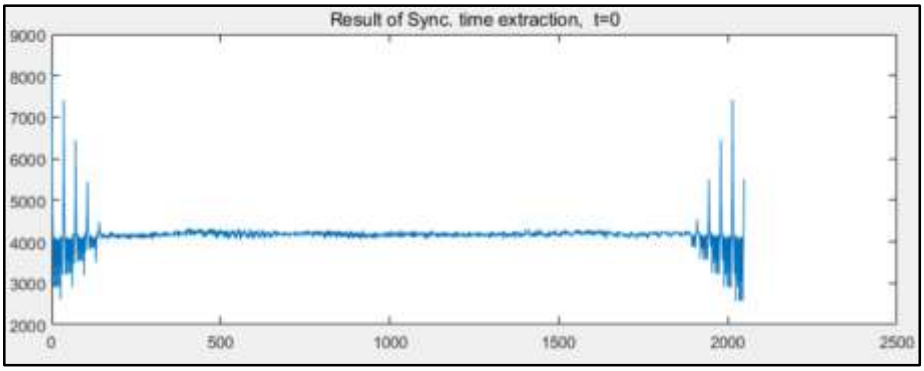

Figure 5. Fourier Transform of Correlation between Barker Code and Extracted Bits

Considering the cropping attack, we extracted watermark from arbitrary sample point of the watermarked audio and them compared Fourier transform of correlation between Barker code and extracted watermark. Figure 5 shows the result of Fourier transform of correlation between 16bits-Barker code and watermark that is extract from arbitrary position. We can observe peaks in Fourier domain which are typical property of Fourier representation of correlation of Barker codes. As a result, it can be confirmed that the attenuation of the volume of the audio signal needs to be carefully considered in order to extract the watermark by a given method in the case of the ACR environment in air as it is implemented in this study. Therefore, in order to apply this method to ACR applications that can be used in the air, the size of the audio signal is normalized frame by frame and the watermark is embedded. In the case of extraction, a watermark is extracted by normalizing the reduced audio signal in consideration of attenuation of the audio signal.

5. Conclusions

In this paper, we have confirmed through experiments that an audio watermarking algorithm using QIM method can be applied in ACR environment where watermark is extracted from audio signal in the air. From the experimental results, it can be confirmed that the normalization of the audio signal size must be preceded when the watermark is inserted and extracted in order to use the QIM method in the ACR environment.

Acknowledgments

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (2015-0-00219, Development of smart broadcast service platform based on semantic cluster to build an open media ecosystem)

This paper is a revised and expanded version of a paper entitled “Study on Synchronizing Detection of Audio by Variable Quantization Step” at GST 2017, Jeju National University, Jeju Island, Korea, December 1, 2017.

References

[1] S. Wu, J. Huang, D. R. Huang and Y. Q. Shi, “Efficiently self-synchronized audio watermarking for assured audio data transmission”, IEEE Trans Broadcast, vol. 51, no 1, **(2005)**, pp. 69-76.
[2] W. Bender, D. Gruhl and N. Morimoto, “Techniques for data hiding”, IBM System Journal, vol. 35, **(1996)**, pp. 313-336.
[3] J.W. Huang, Y. Wang and Y. Q. Shi, “A blind audio watermarking algorithm with self-synchronization”, Proceedings of IEEE Int. Symp. Circuits Syst., **(2002)**.
[4] W. E. Lie and L. C. Chang, “Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification”, IEEE Trans. Multimedia , vol. 8, no. 1, **(2006)**, pp. 46-59.

[5] P. Bassia, I. Pitas and N. Nikolaidis, "Robust audio watermarking in the time domain", IEEE Trans. Multimedia, vol. 3, no. 2, **(2001)**, pp. 232-241.
[6] D. Kirovski and H. Malvar, "Spread-spectrum watermarking of audio signals", IEEE Trans. Signal Processing, vol. 51, no. 4, **(2003)**, pp. 354-368.
[7] S. Xiang and J. Huang, "Histogram-based audio watermarking against time-scale modification and cropping attacks", IEEE Trans. Multimedia, vol. 9, no. 7, **(2007)**, pp. 1357-1372.
[8] M. Steinebach, A. Lang, J. Dittmann and C. Neubauer, "Audio watermarking quality evaluation: robustness to DA/AD processes", Proceedings of International Conference on Information Technology: Coding and Computing, **(2002)**.
[9] J. K. Lee and C. D. Yoo, "Wavelet speech enhancement based on voiced/unvoiced decision", Korea Advanced Institute of Science and Technology", Proceedings of the 32nd International Congress and Exposition on Noise Control Engineering, Jeju International Convention Center, Seogwipo, Korea, **(2003)**.
[10] E. B. Gouvêa and R. M. Stern, "Speaker Normalization through Formant-Based Warping of the Frequency Scale", Proceedings of 5th European Conference on Speech Communication and Technology, **(1997)**, pp. 1139-1142.
[11] L. Garcia, J. Segura, A. de la Torre, C. Benitez and A. Rubio, "Histogram equalization for robust speech recognition, Speech Recognition", Edited by France Mihelic and Janez Zibert, **(2008)**, pp. 248-276.
[12] U. Shrawankar and V. Thakare, "Adverse Conditions and ASR Techniques for Robust Speech User Interface", Proceedings of IJCSI International Journal of Computer Science Issues, vol. 8, issue 5, no 3, **(2011)**.
[13] A. Ricardo and G. Garcia, "Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum theory", Proceedings of the 107th AES Convention, New York, USA, **(1999)**.
[14] C. Xu, J. Wu, Q. Sun and X. Kai, "Applications of digital watermarking technology in audio signals", Journal of Audio Engineering Society, vol. 47, no. 10, **(1999)**, pp. 805–812.
[15] D. Gruhl, A. Lu and W. Bender, "Echo hiding", Proceedings of the Workshop on Information Hiding, number 1174 in Lecture Notes in Computer Science, Cambridge, England, **(1996)**.
[16] V. Licks, F. Ourique, R. Jordan and F. Pérez-González, "The effect of the random jitter attack on the bit error rate, performance of spatial domain image watermarking", Proc. IEEE Int. Conf. Image Processing (ICIP), vol. 2, Barcelona, Spain, **(2002)**, pp. 28–30.
[17] V. Licks, F. Ourique, R. Jordan and G. Heileman, "Performance of dirty paper codes for additive white Gaussian noise", Proceedings of the IEEE Workshop of Statistical Signal Processing (WSSP03), **(2003)**.
[18] P. Moulin, M. K. Mihcak and G.-I. A. Lin, "An information-theoretic model for image watermarking and data hiding", Proceedings of the IEEE Int. Conf. Image Processing, Vancouver, BC, Canada, **(2000)**.
[19] D. Kirovski and H. S. Malvar, "Spread spectrum watermarking of audio signals", IEEE Trans. Signal Process. (Special Issue on Data Hiding), vol. 51, no. 4, **(2003)**, pp. 1020–1033.
[20] R. Bäuml, J. J. Eggers and J. Huber, "A channel model for watermarks subject to desynchronization attacks", in Proceedings of Int'l. ITG Conference of Source Channel Coding, Berlin, Germany, **(2002)**.
[21] W. Xiang Yang and Z. Hong, "A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT", IEEE Trans. Signal Processing, vol. 54, no. 12, **(2006)**, pp. 4835-4840.

Authors

Donghwan Shin, he received the MS degree and Ph.D. in electronics engineering from University of Seoul, Korea. From 1992 to 1994, he was a member of the LG Electronics Inc. He worked as a senior researcher in the Korea Sports Science Institute from 1996 to 2000. He has been currently a chief manager of MarkAny Inc. from 2000. His research interests are in the areas of copyright protection, fingerprinting, watermarking and machine learning.

YoungSeok Lee, he received the Ph.D. degree from University of Seoul, major in signal processing in 1998. He is currently a professor of Dept. of Electronics at Chungwoon University in Korea since 1998. His research interests are in the area of image processing, human visual system modeling and biomedical engineering.

