Journal of Engineering and Applied Sciences 12 (Special Issue 7): 8127-8132, 2017
ISSN: 1816-949X

Spread Spectrum Audio Watermarking Technology for Audio Content Recognition as Second Screen

[1]Dong-Hwan-Shin, [2]Byung-Hoon Lee and [3]Young-Seok Lee
[1]Contents Solution Development Team, 04615 MarkAny, Korea
[2]MarkAny Research Institute, 04615 MarkAny, Korea
[3]Department of Electronic, Chungwoon University, 22100 Hongseong, Korea

Abstract: In recent years, interest in the broadcasting industry for second screen is increasing. This research focuses on ACR applications of audio watermarking at real world based on experimental results. In the ACR process, attenuation of a specific frequency band causes loss or distortion of the watermark embedded in the corresponding frequency band. Consequently, erroneous information is detected in the watermark detection. This study proposed an audio watermarking method based on experimental configuration related to the attenuation of the frequency band according to the distance obtained in the room where the TV speaker and the smartphone are installed. The proposed method is applied to 40 audio files in 5 sec to test imperceptibility and robustness. The experimental results show that the proposed method is robust to ADC and cropping attacks and imperceptible at human auditory system. The imperceptibility is proved by blind audience test at 92% and the watermark can be detectible from recorded audio signal above 90% at distance of 4 m and 60 db of signal intensity. And different segment length with the detection rate is compared. The longer the segment length, the greater the detection rate can be observed regardless types of music. The contribution of this research is that the proposed method was implemented on the experimental result of real world and simply computed while as high imperceptibility and robustness. Almost mobile and wearable devices in real world are required low computational complexity, low memory space and robustness in the analog world. Ultimately, the proposed method is suitable for the purposes of mobile and wearable devices in the real world. The experimental results show that the proposed method is robust to malicious attacks and imperceptible. The imperceptibility by blind audience test is 92% and detection rate is above 90%.

Key words: Automatic content recognition, audio watermarking technology, spread spectrum processing, second screen, interactive application, ACR environment

INTRODUCTION

In recent years, interest in the broadcasting industry for second screen is increasing. The concept of second screen is similar with inter-device collaboration. It means that smart phones and tablet PCs play a complementary role in watching content through TV, so that, TV and smart devices are linked in content (Shahadi *et al*., 2014).

In April 2010, Apple's launch of the iPad was the decisive event that sparked the activation of second screen services. As most of iPad owners began using the iPad while watching TV, the broadcast industry became interested in the second screen. IT professional online media Mashable classified the second screen applications into three categories according with how it works and service type such as audio recognition, camera angle adjustment, social TV application. Many second screen services properly mix and is provided in the form of a hybrid second screen applications (Lacy *et al*., 1998; Griffin and Lim, 1984; Arnold *et al*., 2002). The audio recognition application allows the second screen terminal to listen to the audio of a specific TV program streaming through the speakers of the TV terminal and then provides the contents pre-selected by the content creator through Automatic Content Recognition (ACR) technology. Since, the content provided by the audio recognition application is prepared and stored in advance it is possible to access the content even if the user watches recorded contents instead of the broadcast. In this study, we propose an audio watermarking system which is a basic technology for implementing second screen application for TV and audio broadcasting. The proposed method is constructed to select the frequency band to insert the watermark from the information analyzed by measuring the frequency amplitude of the

Corresponding Author: Young-Seok Lee, Department of Electronic, Chungwoon University, 22100 Hongseong, Korea

audio signal from recorded audio samples. The evaluation of the proposed watermarking technique was performed by measuring the detection rate of the synchronization signal according with the distance between the speaker and the smart phone and the BER (Bit Error Rate) of the detected information.

MATERIALS AND METHODS

Review of ACR technology: In its simplest definition, ACR technology provides the ability to view and listen to content on smart devices such as tablets, smartphones or smart TVs. The device can then provide supplemental, fully-synced content related to the program, movie or ad that you viewed. The viewer's experience automatically improves without manual intervention. Viewers can then use companion devices or smart TVs to access other apps and social networking tools (sharing experiences with friends, family and online communities) to increase viewership and attract customers to brand virtual space (Wang *et al*., 2007; Kabal, 2002).

As in Fig. 1, there are two ACR technologies: digital fingerprint recognition and digital audio watermarking. Watermarking is about inserting payload or specific channel/program ID and time stamp information into content that can be retrieved later. Fingerprinting is the process of analyzing and comparing content with a reference database. Both technologies are now commercially available. Many IT enterprises provide currently the only vendor to offer fingerprint and watermarking ACR solutions to providers such as content owners, PayTV providers, broadcasters and technology partners who integrate advertisers and technology into their platforms (Arnold, 2008; Haitsma *et al*., 2000).

In particular, audio watermarking technology involves dynamically inserting digital tags containing specific information such as program ID, channel ID and time code into the audio content before broadcast distribution. Watermark is not heard by the consumer and is strong in all kinds of signal processing (Chena *et al*., 2012).

The main conditions of audio watermarking for ACR application are robust even in noisy environments and quick search for audio watermarks or so called "payloads" using the appropriate watermark detection algorithm. This resource-efficient algorithm runs on end-user devices such as smart TVs, tablets, smartphones or laptops. In the detection stage of a watermarking system, the ACR watermark detector has a minimal footprint and is easy to integrate. This detector can be used for smart phone protocols such as iOS android and standard protocols that support most common web browsers. After a short

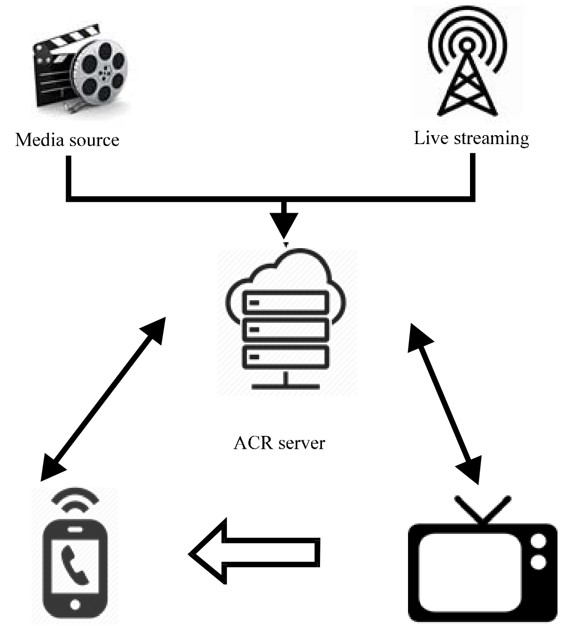

Fig. 1: Basic concept of ACR technology

initial synchronization, the frame of time stamp is accurate. Interactive applications can then apply an interactive task list in time, even for time-shifted views (Arnold *et al*., 2009).

RESULTS AND DISCUSSION

In this study, the experimental environment as shown in Fig. 2 is implemented to select the frequency band to embedding the watermark. In the room of 4, 5 and 2.5 m in length, width and height, respectively a TV with stereo speakers and a smartphone with recording function were installed.

The stereo speakers on the TV in the room emit sine waves at 100 Hz intervals from 100-4 kHz and smart phones in front of the TV record the sine, waves emitted. The frequency range from 100-4 kHz which is the human audible is a frequency band that is robust against compression attacks while it is sensitive to small distortions. From the frequency analysis using the Fast Fourier Transform (FFT), the attenuation of the frequency band according to the distance is observed. As shown in Fig. 3, we can observe the frequency spectrum of the sine wave emitted from the stereo speakers and the attenuation band of the frequency spectrum of the signal collected from the recording device of the smart phone. Frequency bands where significant attenuation occurs are observed in the 2700-3000 and 3500-3800 Hz bands. A typical watermark insertion process that does not consider attenuation due to distance can extract a watermark at the receiving end almost completely by transmission of

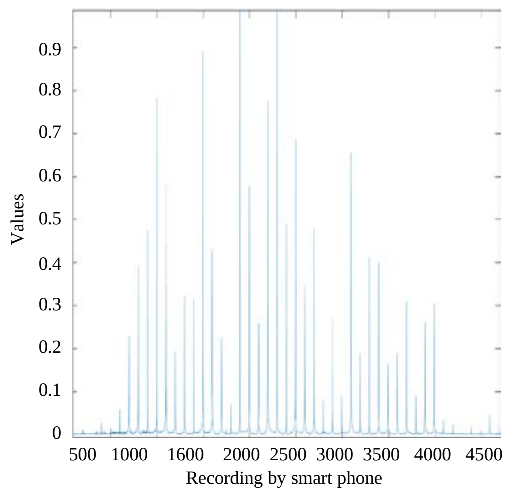

Fig. 2: Experimental configuration for audio recording by smart phone

Fig. 3: Experimental configuration for audio recording by smart phone

digitized audio data but it can cause a serious error in the ACR environment. The analog waves emitted from the speakers of the TV are converted back to digital signals from the recording device of the smartphone. This process is called Analog to Digital Conversion (ADC) attack in audio watermarking technology. The process of converting a digital signal back to a digital signal via. an analog signal is processed by a different sampling frequency from that of the original digital signal. In the ACR process, attenuation of a specific frequency band causes loss or distortion of the watermark embedded in the corresponding frequency band. Consequently, erroneous information is detected in the watermark detection process and the performance of the corresponding watermarking system is degraded. And as expected, it is observed that the low audio frequency band that is the 1-600 Hz frequency band is removed during most smartphone recording processes to solve limited memory space of smartphones. At the result of our experiment, we found from the experimental results that the frequency bands that should not insert the watermark for ACR applications are 900-1000 and 3500-3800 Hz.

The problem to be solved to apply audio watermarking technology to ACR applications is synchronization processing. As with digital audio signals, audio watermarking techniques for ACR applications are exposed to a 'cropping attack' in which a portion of the audio data is removed. A watermarked audio signal that has exposed a cropping attack loses synchronization information, subsequently the watermark detection process detects erroneous audio information that relies on synchronization information that is detected incorrectly. In this research, we propose a watermark embedding method using Spread Spectrum Processing (SSP) to solve this problem. The contribution of the proposed method is to embedding the watermark considering the frequency attenuation property according to the distance. The watermarking technique using a typical SSP embeds watermark information in the whole frequency band using pseudo random noise generated by a specific 'seed' and applies a psychoacoustic model considering the human auditory system in some cases. Based on our experiment, we have slightly modified the typical SSP method. The proposed method is that the watermark generated by pseudorandom number is embedded into the surviving frequency band. In addition, the embedded watermark is modulated according to the frequency band, so that, the human auditory system cannot identify the existence of it.

The process for embedding watermark in our method is shown in Fig. 4. The first step for embedding the watermark is to segment the original audio signal to insert a 1 bit watermark. In the next step, the power of the segmented audio signal is calculated and then the embedding strength of the watermark is determined. The determination of the embedding strength is determined by the following equation in consideration of the human auditory system. In Eq. 1, SWR is defined as the Signal to Watermark Ratio and means the power ratio of the signal to the watermark in each audio segment. In the third step, thewatermark is embedded into the corresponding frequency band using the SSP method. The embedding process repeats until the last segment:

$$\text{SWR} = \frac{10\log\left(\sum_{i=1}^{n}\frac{s(i)^2}{N}\right)}{\sum_{i=1}^{n}\frac{\omega(i)^2}{N}} < -7\,\text{db} \quad (1)$$

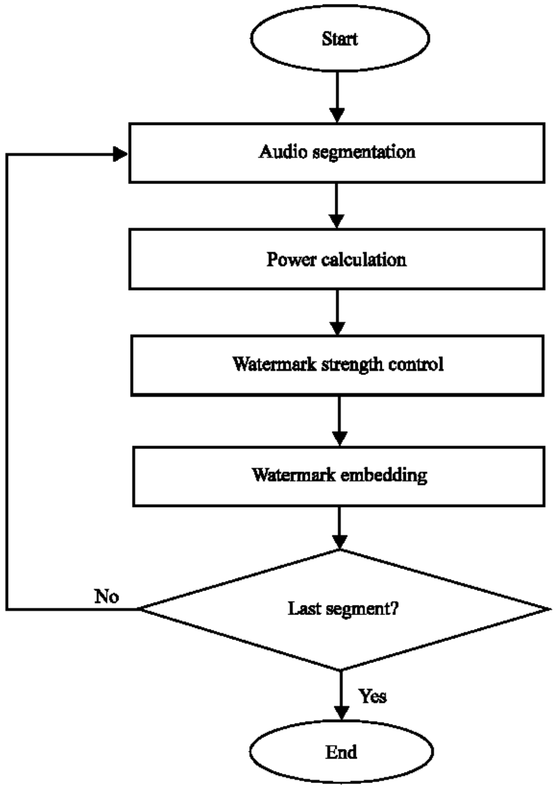

Fig. 4: Watermark embedding scheme

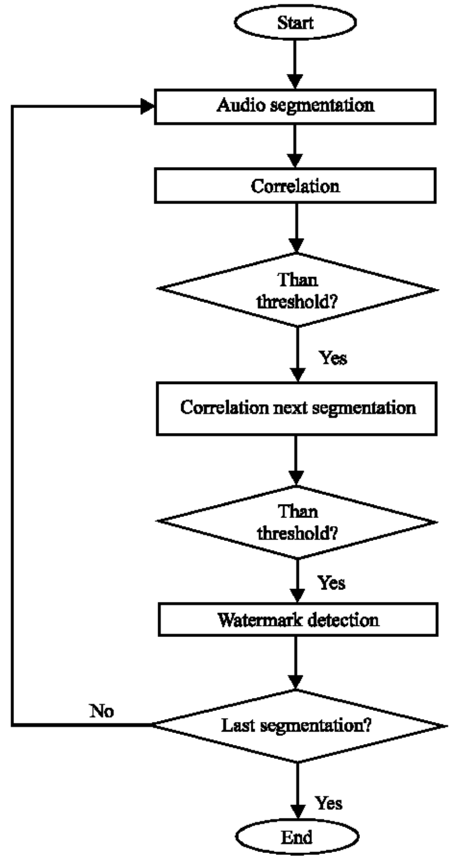

Fig. 5: Watermark detecting scheme

The process for detecting watermark in our method is shown in Fig 5. The watermark detection process starts with segmenting the audio signal into the same length as the embedding process. It is impossible to detect the watermark at the start point of the watermarked audio signal since the watermarked audio signal regards that the synchronization information has already been lost due to the cropping attack. A typical method is to detect the in Fig. 5. In our method, since, the proposed method inserts a watermark into a specific frequency band, a typical correlation method should be performed in the corresponding band. Subsequently, it is necessary to perform a band pass filtering process to extract the signal of the corresponding frequency band before using the correlation method. The next step is the same as the typical correlation process and calculation as in Eq. 2:

$$CRF(\tau) = \int_{i=1}^{N} R_{BPF}(t) S_{WM}(t-\tau) \quad (2)$$

synchronization information using the correlation method. In Eq. 2, CRF(τ) represents the correlation value between the band pass filtered pseudorandom noise, $R_{BPF}(t)$ and the watermarked audio segment, $S_{WM}(t)$. For the experiment, 50 audio files of 5 sec with 44.1 kHz of sampling frequency are used. The file can be divided into two categories: the first class is audio files consisting of extremely or moderately quiet music and the second class is an audio file consisting of extremely loud rock music or moderate dance music. An audio watermark is embedded in each audio file by the proposed method. When embedding a watermark, a watermark having a different segment length is embedded to compare the watermark length with the detection rate. A subjective blind audience test was performed to test the imperceptibility of the embedded watermark. A total of 10 test participants randomly listened watermarked and non-watermarked audio and recorded the watermarked audio signal. The results of the blind audience test showed that 92% of the participants could not distinguish watermarked audio signals. These results mean that the proposed method cannot be perceived in the human auditory system.

The watermarked audio signal is emitted through the TV speaker in Fig. 2 and the smartphone records audio signal emitted sound with signal intensity of 60 db at 4 m. The watermark is extracted from the recorded audio file through the detection process and the detection rates are

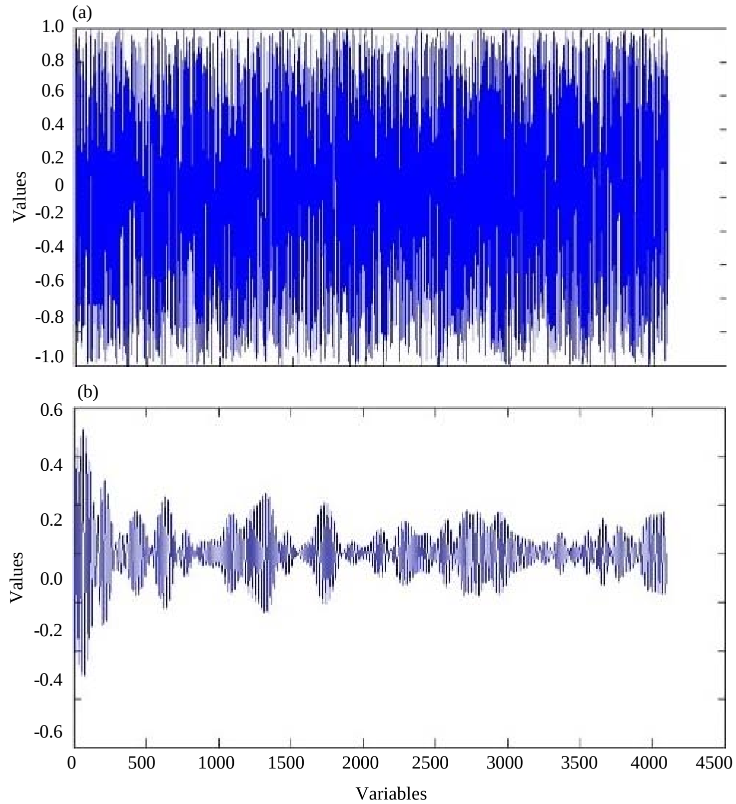

Fig. 6: Pseudorandom noise and its modulated version

compared. All watermark embedding and detecting process are implemented by M file and computed in MATLAB environment. Figure 6 shows the random noise and its bandpass filtered version used as a watermark and the correlation between them. It can be observed that the watermark represents a typical amplitude modulated signal with a constant interval of oscillation and the correlation between them shows a high peak at 0 delay and low values at another delay.

Segments of 256, 512, 1024 and 2048 samples and detection rates are shown in Table 1 to compare the watermark detection rate with the length of the audio segment. The detection rate is different depending on the kind of music but the longer the segment length, the greater the detection rate can be observed. Audio files containing mainly loud music such as file 1 and 2 show a high detection rate even in the segment of small samples. On the other hand, audio files containing primarily quiet music such as file 1 and 2 can be observed relatively low detection rates in segments with 256 and 512 samples while >90% in segments above 1024 samples.

Table 1: Watermark detection rate by segment length

File No.	Segmentation length (samples)	Detection rate (%)
1	256	65.3
2		66.1
3		33.4
4		26.5
1	512	70.1
2		80.6
3		50.6
4		76.5
1	1024	90.4
2		92.1
3		78.5
4		98.4
1	2048	95.8
2		97.3
3		88.8
4		99.9

Figure 7 shows the correlation between the watermarked audio signal and the watermark. It inserts 1 bit of information per segment and extracts 1024 samples as one segment. The high peak indicates the position of the synchronization signal for extracting the watermark. Since, the same watermark is repeatedly inserted every 1024 samples it can be observed that peaks occur at regular intervals.

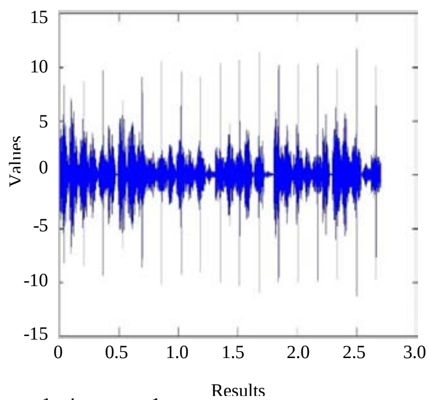

Fig. 7: Correlation results

CONCLUSION

This study proposed an audio watermarking method based on experimental configuration which is applicable to ACR applications. Based on the experimental results related to the attenuation of the frequency band according to the distance obtained in the room where the TV speaker and the smartphone are installed, a watermark embedding method are implemented using the SSP in a frequency band in which the less attenuation of the frequency magnitude.

The proposed method is applied to 40 audio files in 5 sec to test imperceptibility and robustness. The experimental results show that the proposed method is robust to ADC and cropping attacks and imperceptible at human auditory system. The imperceptibility is proved by blind audience test at 92% and the watermark can be detectible from recorded audio signal above 90% at distance of 4 m and 60 dB of signal intensity. And different segment length with the detection rate is compared. The longer the segment length, the greater the detection rate can be observed regardless types of music. The contribution of this research is that the proposed method was implemented on the experimental result of real world and simply computed while as high imperceptibility and robustness. Almost mobile and wearable devices in real world are required low computational complexity, low memory space and robustness in the analog world. Ultimately, the proposed method is suitable for the purposes of mobile and wearable devices in the real world.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Advanced Technology Center R&D Program funded by the Ministry of Trade, Industry and Energy of Korea (10042252).

REFERENCES

Arnold, M., 2008. Subjective and Objective Quality Evaluation of Watermarked Audio. In: Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks, Cvejic, N. and T. Seppanen (Eds.). IGI Global, Hershey, Pennsylvania, pp: 260-277.

Arnold, M., M. Schmucker and S.D. Wolthusen, 2002. Techniques and Applications of Digital Watermarking and Content Protection. Artech House, Norwood, Massachusetts, Pages: 281.

Arnold, M., P.G. Baum and W. Voeßing, 2009. A Phase Modulation Audio Watermarking Technique. In: Information Hiding, Katzenbeisser, S. and A.R. Sadeghi (Eds.). Springer, Berlin, Germany, ISBN:978-3-642-04430-4, pp: 102-116.

Chena, X.M., M. Arnolda, P.G. Baum and G. Doerr, 2012. AC-3 bit stream watermarking. Proceedings of the 2012 IEEE International Workshop on Information Forensics and Security (WIFS'12), December 2-5, 2012, IEEE, Tenerife, Spain, ISBN:978-1-4673-2285-0, pp: 181-186.

Griffin, D. and J. Lim, 1984. Signal estimation from modified short-time Fourier transform. IEEE. Trans. Acoust. Speech Signal Process., 32: 236-243.

Haitsma, J., M.V.D. Veen, T. Kalker and F. Brueker, 2000. Audio watermarking for monitoring and copy protection. Proceedings of the 2000 ACM Workshops on Multimedia, October 30-November 3, 2000, ACM, Los Angeles, California, pp: 119-122.

Kabal, P., 2002. An examination and interpretation of ITU-R BS. 1387: Perceptual evaluation of audio quality. J. Interpretation Audio Eng., 1: 1-87.

Lacy, J., S.R. Quackenbush, A.R. Reibman and J.H. Snyder, 1998. Intellectual property protection systems and digital watermarking. Opt. Express, 3: 478-484.

Shahadi, H.I., R. Jidin and W.H. Way, 2014. Lossless audio steganography based on lifting wavelet transform and dynamic stego key. Indian J. Sci. Technol., 7: 323-334.

Wang, X., W. Qi and P. Niu, 2007. A new adaptive digital audio watermarking based on support vector regression. IEEE. Trans. Audio Speech Lang. Process., 15: 2270-2277.