Audio Activity Detection for Watermarking based on Wavelet Denoising

Young-Seok Lee*
Department of Electronics, Chungwoon University, Sukgol-Ro113, Incheon - 22100, Korea; yslee@chungwoon.ac.kr

Abstract

Objectives: The watermark embedding in Non-Activity Region (NAR) of the underlying audio signal results for the degradation of audio quality. This research focused on the development of audio activity detection method for the audio watermarking procedure to prevent from watermark embedding in the NAR. Methods/Statistical Analysis: We propose the audio activity detection method based on the wavelet denoising algorithm to classify the NAR in audio signal. We experiment and apply three- type wavelet denoising algorithms, subsequently the wavelet denoising algorithms showed the same performance. We analyze the audio quality by the Signal to Noise Ratio (SNR) for two cases that the embedding and non-embedding of watermark in the NAR. Findings: We surveyed the effects on the watermarking embedding in the NAR. Decreasing a payload on watermarking, the watermark embedding in the NAR increases the audio quality in SNR and audio test by MUSHRA (Multi Stimulus Test with Hidden Reference and Anchor). In extraction process the extraction rate of embedded watermark is slightly increasing when our method is applied. In comparisons of the other research results such as DFT (Discrete Fourier Transform) based method and MFCC (Mel Frequency Cepstral Coefficients) based method, our method has the better accuracy in AAD. And also our method is in low complexity for the analysis of computational complexity by the effective computational structure of filter bank for implementation of wavelet filter. In additional research results, we find out that the audio signal in the NAR has a random- like pattern by the analysis of power spectrum of it. In statistical experiment a signal in NAR analyzed by statistical parameters such as mean and variance. As a result, our method is effective for computational complexity and produces good audio quality for the watermark embedding procedure. Improvements/Applications: Our method for audio activity detection is very effective computational complexity and produces high audio quality for watermark embedding procedure. It is applicable for the audio watermarking system for high- quality watermarked audio signal.

Keywords: Audio Activity Detection, Audio Watermarking, Non Activity Region, Signal to Noise Ratio, Wavelet Denoising

1. Introduction

Recently the piracy related to digital contents as illegal copying and unauthorized distribution is increasing while development of Internet and digital multimedia technology has advanced much easier and much speedy in digital content distribution such as video, audio and image. The digital watermarking solution is the technical measure against illegal activity for digital contents. The digital watermarking scheme has two sub- schemes as the embedding and extracting schemes for watermark that has the information about the copyright of its contents. The watermark embedding scheme is the embedding the watermark under condition of imperceptible and robust in malicious attacks such as signal processing, A/D, D/A conversion, and the detection scheme is the detecting of
Audio Activity Detection for Watermarking based on Wavelet Denoising

Imperceptible condition of digital audio watermarking scheme is that the human hearing system cannot recognize the underlying audio signal before and after the watermark embedding. The imperceptibility is measured by two evaluations, which are the quantitative evaluation such as SNR (Signal to Noise Ratio) and PSNR (Peak SNR) and the qualitative evaluation such JNE (Just Noticeable Evaluation), aural assessment. International Federation of Phonographic Industry (IFPF) recommends conditions of audio watermark scheme. The first condition is the SNR of the watermarked audio signal is over in 20 dB. The second condition is security that means the watermarked audio signal should leave any evidence in it and also depends on binary information called the secret key in the open state of the watermark embedding scheme. The last condition is robustness of watermarking scheme that means the watermark extraction ability in attacks such as noising, filtering and various malicious attacks. There is a trade-off between three conditions.

With respect to the development of recent audio watermarking technology, the imperceptibility has been further emphasized. The HAS (Human Auditory System) can be analyzed as a temporal-frequency transducer consisting of a set of the critical band pass filter bank in the range of 10 Hz-20 KHz. The bandwidth of critical band pass filters is logarithmic-increasing with each center frequencies of filters. The loudness of sound is defined that the perceived intensity of monotone sound depending on the underlying Sound’s Pressure Level (SPL), frequency and period. But in mixed situation of multi-frequencies sounds the loudness of sound with an interested frequency varies depending on the frequency and/or temporal neighborhood of another sound. This phenomenon called masking effect is composing of two factors, masker and maskee. The masker is a sound that does not mask and the maskee is a sound that is masked.

Masking effect is categorized categories, temporal masking and frequency masking. The temporal masking is the relation between the masker and maskee in temporal neighborhood sound components. The temporal masking is sub-categorized pre-masking and post-masking depending on the position of masker and maskee in temporal domain. Simultaneous masking or frequency masking is also the relation between the masker and maskee in frequency domain. By frequency masking, two sounds that closer in frequency domain heard like tonal or noisy sound. Considering masking effect of HAS, the audio watermarking scheme is implemented by applying of psychoacoustic model that models HAS.

A problem to solve with respect to applying psychoacoustic model for audio watermarking is a watermark embedding in non-audio activity region. The psychoacoustic model for watermarking cannot be applied for non-activity region but activity region, subsequently the watermark embedding in NAR results in disturbing of imperceptibility of watermarking that is necessary condition of watermarking technique as above described. An intuitive candidate for preventing the degradation of imperceptibility is that the watermark is not embedded in NAR so that detects non-auditory signal from underlying signal. We propose NAR detection method to prevent watermark embedding from NAR, which also imply the detection of audio activity region. The paper is organized by following as literature reviews for AAD (Audio Activity Detection) are described in Section 2. Section 3 presents our AAD algorithm including wavelet denoising, watermarking algorithm and overall procedure of method. In Section 4, we show the experimental results in relation to the accuracy of experimental evaluations about detection. In conclusions and further work in Section 5 we summarize our conclusion related to experimental results and additional experiment in future.

In audio signal Audio Activity Detection (AAD) is important problem in separating audio components from noisy components. The AAD is applied to wide fields in audio signal processing such as audio watermarking, audio coding/decoding, audio compression, audio division from noisy environment, speech automatic recognition and audio adaptive filtering and speaker identification. The first condition of AAD scheme is robustness in daily noisy environment that is accustomed to the HAS. In addition to above condition, the simplicity of algorithm is also important to realize real time application. The simplicity depends on its algorithmic procedure while the condition of robustness is related to the inherent condition of audio signal itself. There is trade-off between two conditions. For example, As increasing SNR, for example, the demand for simplicity of AAD scheme payed for the condition of
robustness while payed for simplicity in the condition of robustness.

Various researches for AAD are address in the literatures by its noticeable importance in various audio processing fields. But the procedure of AAD algorithms have two steps, the first step is detection of features that represents characteristics of underlying audio signal and then the second step is identifying the underlying audio signal. Feature related to audio signal is divided in time and frequency domain. Time domain features are that autocorrelation, period are representative and Frequency domain features are that cepstral coefficient and linear predictive coefficients are representative. Apart from above described researches, there are other researches that used to another features such as Fourier coefficients, algebraic features that are eigen value decomposition and singular value decomposition. In recent researches, the Deep learning technology that is a kind of artificial neural network method is experimentally applied to AAD.

2. Proposition of Method for AAD

The overall structure of the proposed method is shown in Figure 1. The first step of algorithm starts from wavelet transform of underlying audio signal that discriminated by frame for 1 bit embedding and then threshold value is estimated using statistical parameters of underlying audio signal. The wavelet denoising procedure is performed by the estimated threshold value. To recover the denoised audio signal, reverse wavelet transform is performed. The last step is the estimation of time interval and classify non-activity and audio activity from there covered audio signal.

The threshold is defined that wavelet coefficients under the threshold value are moved to small values or zeros in wavelet de-noising scheme. The result of wavelet de-noising sustains the original signal and suppresses noisy components that spread all frequency bands. The thresholding scheme is categorized the hard and soft thresholding, the former set wavelet coefficients to zeros under thresholding value, the last set them to small values to shrink under the thresholding value. Two thresholding schemes contribute the removal or weakening of noisy component in the original signal. Some of them are discussed here briefly. In this paper performances of three well-known standard threshold estimation methods are used for underlying audio signals. The first method finds threshold using Minimax principle. Considering mean square error, a constant threshold value is used to calculate Minimax performance that is estimation evaluator in statistics. The Minimax estimator to find optimal thresholding value that is implemented by the minimizing mean square error depending on a given set of function of MMSE (maximum mean square error). The threshold value in (1), the resulting optimized thresholding is shown as

\[ \lambda = \begin{cases} \sigma(0.3936 + 0.1629 \times \log_2 N), & N > 32 \\ 0, & N \leq 32 \end{cases} \]  

(1)

Where \( \sigma = \text{median}\left(\frac{|\omega|}{0.6745}\right) \) and \( \omega \) is the wavelet coefficients at any scale and \( N \) is the length of signal. The threshold value \( \lambda \) of the second threshold is estimated by hard (universal) threshold using square root log evaluation. i.e.

\[ \lambda_j = \sigma_j \sqrt{2 \log(N_j)} \]  

(2)

where, \( N_j \) presents the length at \( j \)-th scale wavelet coefficients and \( \sigma_j \) is Median Absolute Deviation (MAD) at \( j \)-th scale.

\[ \sigma_j = \frac{\text{MAD}_j}{0.6745} = \frac{\text{median}(|\omega|)}{0.6745} \]  

(3)

Where \( \omega \) is the \( j \)-th scale wavelet coefficient. The combination of universal and soft threshold methods is used to selecting of the final threshold. In case
of low SNR, the better threshold estimation is performed by constant threshold value in universal method\(^3\). The combination of two threshold values, the universal and soft threshold \(\lambda_1\) and \(\lambda_2\) generate Heuristic threshold value is shown in (4).

\[
\lambda = \begin{cases} 
\lambda_1, & A > B \\
\min(\lambda_1, \lambda_2), & A \leq B
\end{cases}
\]  

(4)

Where \(A\) and \(B\) is calculated wavelet coefficient and squared wavelet coefficients by \(N\).

In watermark embedding process, we apply quantization-indexed watermarking scheme\(^9\) as shown in Figure 2. The original audio signal is filtering to extract band pass signal, then the characteristic value is calculated from the band pass signal. The watermark embedding is given by,

\[
s'(t) = \begin{cases} 
(s(t) + gs^b(t), & W = 1 \\
(s(t) - gs^b(t), & W = 0)
\end{cases}
\]

Where \(s'(t)\) is an watermarked audio signal, \(g\) is the embedding strength and \(W\) presents 1 bit watermark information. Especially we add the band pass audio components, \(s^b(t)\), to embedding the watermark. Considering bit extraction error in extraction process, we apply the BCH-coded watermark information which can correct the extraction error in 10% bit error rate\(^12\).

### 3. Experimental Results

We apply our AAD algorithm for audio signals. An example of audio signal for experiment shows in Figure 3. The half interval of the underlying audio signal is NAR that characterized by low power and amplitude. The components in NAR are considered background noise and/or very small amplitude of back ground audio components and these two components do not have only meaningful audio information but also affect imperceptibility to human auditory system.

**Figure 2.** Watermarking embedding scheme.

**Figure 3.** Example of audio signal having NAR.

When three wavelet thresholding methods are applied for audio signal, the thresholding results are shown in Figure 4.

**Figure 4.** Results of wavelet denoising by three thresholding methods.

Three thresholding methods perfectly detect NAR in underlying audio signal and we cannot observe the
difference of performances. The average accuracy for 102 audio sample signals is 98.4% by experiment. The experimental results of other researches using FFT- and MFCC-based methods are 96.2%- 97.2%, average accuracy. This result implies that our method has slightly more accuracy comparing with the two conventional AAD algorithms.

Figure 5. Audio quality comparison by difference between original and watermarked audio signal.

The results of watermarked audio signals are shown in Figure 5. The Figure 5 shows the difference of original and watermarked audio signal when the watermarks are embedded not considering the NAR and the below figure shows the difference between them when considering the NAR. The difference values in NAR are zeros because that the watermarks do not embed in there. This result implies that the watermark embedding procedure except for NAR contributes to the increase in audio quality. From audio test called MUSHRA that is abbreviation of ‘multi stimulus test with hidden reference and anchor’, the listeners participated in the test evaluated that the watermarked audio signal by our method was much imperceptible comparing with audio signals by the conventional watermarking method.

In order to analyzing audio signal in the NAR, we also analyses statistical characteristics of audio signal in the NAR. An example of audio signal in NAR is shown in Figure 6. The amplitude of audio signal in the NAR can be observed random-like signal. When the 102 sample audio signals for our experiment are testified for statistical characteristics, the consistency of statistical variables such as mean and variance cannot observe while the frequency characteristics is concentrated on the frequency band, 2.1 kHz- 6 kHz.

Figure 6. Example of audio signal in NAR.

4. Conclusions

The watermark embedding in NAR of the underlying audio signal results for the degradation of audio quality. In this research we are focused on the development of audio activity detection method for the audio watermarking procedure to prevent from watermark embedding in the NAR. We propose the audio activity detection method based on the wavelet denoising algorithm to classify the NAR in audio signal. We experiment and apply three-type wavelet denoising algorithms, subsequently the wavelet denoising algorithms showed the same performance. We analyze the audio quality by the SNR for two cases that the embedding and non-embedding of watermark in the NAR. We surveyed the effects on the watermarking embedding in the NAR. Decreasing a payload on watermarking, the watermark embedding in the NAR increases the audio quality in SNR and audio test by MUSHRA. In extraction process the extraction rate of embedded watermark is slightly increasing when our method is applied. In comparisons of the other research results such as DFT based method and MFCC based method; our method has the better accuracy in AAD. And also our method is in low complexity for the analysis of computational complexity by the effective
computational structure of filter bank for implementation of wavelet filter. In additional research results, we find out that the audio signal in the NAR has a random-like pattern by the analysis of power spectrum of it. In statistical experiment a signal in NAR analyzed by statistical parameters such as mean and variance. As a result, our method is effective for computational complexity and produces good audio quality for the watermark embedding procedure.

5. References

1. Mamatha P, Venkatram N. Watermarking using Lifting Wavelet Transform (LWT) and Artificial Neural Networks (ANN). Indian Journal of Science and Technology. 2016 May; 9(17):1–7.
2. Bilal I, Kumar R. Audio steganography using QR decomposition and fast Fourier transform. Indian Journal of Science and Technology. 2015 Dec; 8(34):1–7.
3. Gokhun S, Ozer H. Voice activity detection in non-stationary noise. IEEE Transaction on Speech and Audio Processing. 2000 Jul; 8(4):478–82.
4. Math Works Document. Available from: http://kr.mathworks.com/help/wavelet/ref/wden.html
5. Xie H, Leland E, Ulaby T. SAR speckle reduction using wavelet denoising and Markov random field modeling. IEEE Transactions on Geoscience and Remote Sensing. 2002; 40(10):2196–212.
6. Sandeep P, Akbar M, Richard G. Improved wavelet denoising via empirical Wiener filtering. Optical Science, Engineering and Instrumentation, International Society for Optics and Photonics. 1997: 389–99.
7. Sylvain S, Paul T, Andrew B. Robust wavelet denoising. IEEE Transactions on Signal Processing. 2001; 49(6):1146–52.
8. Jianbo G. Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: A comparison. IEEE Signal Processing Letter. 2010; 17(3):237–40.
9. Nedeljko C, Djordje T, Tapio S. Increasing robustness of an audio watermark using turbo codes. Proceedings of Multimedia and Expo, ICME’03; 2003. p. 1:1-217–20.
10. Wang Q. Watermark embedding algorithms based on quantizing frequency domain parameters of digital audio signal. Acta Acustica. 2002; 4–15.