AUDIO WATERMARKING OVER THE AIR WITH MODULATED SELF-CORRELATION

Yuan-Yen Tai, Mohamed F. Mansour

Amazon Inc., USA

ABSTRACT

We propose a novel audio watermarking system that is robust to the distortion due to the indoor acoustic propagation channel between the loudspeaker and the receiving microphone. The system utilizes a set of new algorithms that effectively mitigate the impact of room reverberation and interfering sound sources without using dereverberation procedures. The decoder has low-latency and it operates asynchronously, which alleviates the need for explicit synchronization with the encoder. It is also robust to standard audio processing operations in legacy watermarking systems, e.g., compression and volume change. The effectiveness of the system is established with a real-time system under the most general room conditions.

Index Terms— audio watermarking, asynchronous decoder, reverberation, spread-spectrum, second-screen.

1. INTRODUCTION

In most existing audio watermarking scenarios in the literature, the audio signal stays in the digital domain between the encoder and the decoder. This is a typical situation in digital right management of audio distribution, where the watermarking decoder is invoked prior to media playback [1, 2]. Recently, there has been growing interest in audio watermarking that survives indoor acoustic propagation, e.g., for second-screen applications [3]. In this scenario, the watermarked audio is played through a consumer loudspeaker after the encoder, propagates through an indoor acoustic channel, picked by a consumer microphone (usually in another device) before passing to the watermark decoder. This scenario poses a set of new challenges that were not encountered in legacy audio watermarking:

- Room reverberation, which introduces time and frequency smearing of the audio content [4].
- Time/frequency drift between the encoder and decoder due to different system clocks.

The relevant work in the literature has treated these two challenges rather separately, and frequently at the cost of less robustness to standard audio processing operations. For example, few audio watermarking systems have been designed to withstand desynchronization between the encoder and decoder [5, 6, 7, 8, 9, 10]. This robustness can be through using features that are robust to local time-scale variations [5], or through deploying a special synchronization mechanism (through time-warping like procedure) at the decoder [8, 9]. On the other hand, some earlier works have focused on the reverberation impact while assuming perfect synchronization exists [10, 11]. In [10], a special filter bank with a long symbol interval is used, and the watermark is embedded in the specific time-frequency cells that are robust to expected operations. The synchronization has not been explicitly addressed, rather general guidelines from wireless communication systems were described. An end-to-end audio system with practical computation and latency constraints has been largely absent in earlier literature, and this is the focus of this work.

In particular, we develop a novel audio watermarking system that is robust to both reverberation and desynchronization as well as standard audio processing operations. The encoder embeds a spread-spectrum watermark in successive short blocks of the host audio, and the watermark at each block is modulated with a binary ± 1 sequence to improve the detection and suppress host signal correlation. The encoder resembles standard audio watermarking systems, therefore, it inherits their good properties, e.g., imperceptibility of the watermark, and robustness to standard signal processing operation such as audio coding and filtering. The decoder applies a modulated *self-correlation* of successive blocks rather than the standard matched filter that uses cross-correlation with the embedded watermark (which requires perfect synchronization and knowledge of the acoustic channel at the decoder). Although self-correlation is not the optimal detector from detection-theory perspevitve, it effectively and blindly mitigates the impact of both reverberation and desynchronization at a low-cost in both computation and latency, which enables real-time embedded implementation.

2. BACKGROUND

2.1. Spread-Spectrum Watermarking

In the following, we assume the watermark is embedded in selected DCT coefficients of audio blocks. Spread-Spectrum watermarking procedure has the general form [12, 13, 1]. Note, a regular face lower-case letter denotes a vector in the time domain and a bold face lowrer-case letter denotes a vector in the frequency domain.

$$\tilde{\mathbf{x}} = \mathbf{x} + \eta \mathbf{w} \tag{1}$$

where η is the watermark strength (which controls the audibility of the watermark). If $\mathbf{y}$ is the received signal in the DCT domain, then the standard spread-spectrum decoder uses cross correlation of the form

$$\rho = \langle \mathbf{y}, \mathbf{w} \rangle \tag{2}$$

Note in the additive noise case $\mathbf{y} = \mathbf{x} + \eta \mathbf{w} + \mathbf{n}$ (where $\mathbf{n}$ is the noise component), then

$$\langle \mathbf{y}, \mathbf{w} \rangle = \langle \mathbf{x}, \mathbf{w} \rangle + \langle \mathbf{n}, \mathbf{w} \rangle + \eta \|\mathbf{w}\|^2 \tag{3}$$

If the watermark is not correlated with the signal nor the noise, then both $\langle \mathbf{x}, \mathbf{w} \rangle$ and $\langle \mathbf{x}, \mathbf{n} \rangle$ vanish and ρ becomes proportional to the watermark energy. Therefore, at the detector, ρ is compared by a predetermined threshold, γ. If $\rho \geq \gamma$, then the watermark is detected at the decoder; otherwise, it is not detected.

2.2. Acoustic Channel Model

The acoustic propagation channel has few sources of distortions: clock drift between the encoder and the decoder, sampling rate difference, loudspeaker behavior, room reverberation, and analog-to-digital and digital-to-analog distortion. The microphone impact is

usually ignored because of its flat response over frequency of interest. The clock drift is measured in parts-per-million (ppm), and consumer-grade system clocks can have up to few hundreds ppm difference. If the clock drift is 100 ppm, then at 48 kHz sampling frequency, then the effective sampling frequency is 48000 ± 4.8 Hz, which results in a time shift of up to 4.8 samples every second, and also a slight frequency shift. If an explicit synchronization procedure is used, then this clock drift must be estimated and corrected (through PLL-like systems [14]). The operating sampling rate difference could be mitigated by standardizing the sampling frequency at which the watermark is embedded or detected. The other distortions can be broadly modeled as a slowly time-varying channel with additive noise similar to fading channels in broadband wireless communication:

$$y(t) = \sum_{\tau} h^{(t)}(\tau)\tilde{x}(t-\tau) + n(t) \tag{4}$$

where $\{h^{(t)}(\tau)\}$ is the time-varying impulse response, and $n(t)$ is the additive noise. In the frequency-domain, we use the same notation to become

$$y^{(t)}(\omega_k) = h^{(t)}(\omega_k)\tilde{x}^{(t)}(\omega_k) + n^{(t)}(\omega_k) \tag{5}$$

where $y^{(t)}(\omega_k)$ is the frequency response of y at audio frame t, and similarly for $x^{(t)}(\omega_k)$ and $n^{(t)}(\omega_k)$, whereas $\alpha^{(t)}(\omega_k)$ and $\{\alpha^{(t)}(\tau)\}$ are Fourier pairs. In vector-form, each DCT block can be represented as [15]

$$\mathbf{y} = \tilde{\mathbf{x}} \odot \boldsymbol{\alpha} + \mathbf{n} \tag{6}$$

where $\odot$ denotes element-wise vector multiplication, and $\boldsymbol{\alpha}$ is the channel representation in the DCT domain. If α_k changes amplitude and sign with the frequency index k (which is the typical case), then spread-spectrum based audio watermark detection would fail. To see this, consider the cross-correlation factor in this case (assuming perfect synchronization)

$$\begin{aligned} \langle \mathbf{y}, \mathbf{w} \rangle &= \langle \mathbf{x} \odot \boldsymbol{\alpha}, \mathbf{w} \rangle + \langle \mathbf{w} \odot \boldsymbol{\alpha}, \mathbf{w} \rangle \\ &\approx \eta \sum_k \alpha_k \, |w_k|^2 \end{aligned} \tag{7}$$

If the sign of α_k changes with k, then the cross-correlation becomes close to the noise level, and detection fails. In this case, the optimal detector requires knowledge of the channel at the receiver, which is the common approach in wireless communication. The estimation is performed by transmitting a known pilot signal at the start of each frame, which is used for system identification at the receiver. The channel estimation procedure requires perfect synchronization, and it is an expensive procedure in both computation and latency.

3. WATERMARKING SYSTEM

3.1. Watermark Design

If $\mathbf{H}$ is a full-rank symmetric matrix of size κ, then its eigenvectors $\{\mathbf{v}_i\}$ are real and constitute a set of orthonormal basis for $\mathcal{R}^{\kappa}$ [16]. Let the watermark $\mathbf{w}$ be chosen as one of the eigenvectors, e.g., $\mathbf{v}_1$ (without loss of generality). The host signal block, $\mathbf{x}$ in (1), can be expressed as

$$\mathbf{x} = \sum_l a_l \, \mathbf{v}_l \tag{8}$$

where $a_l = \langle \mathbf{x}, \mathbf{v}_l \rangle$. In this case, the cross-correlation between the host signal and the watermark becomes

$$\langle \mathbf{x}, \mathbf{w} \rangle = a_1 \tag{9}$$

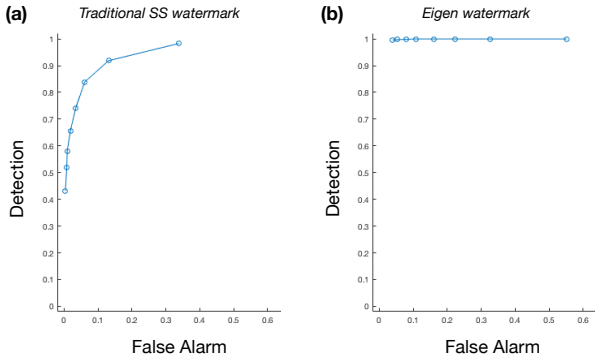

Fig. 1. The ROC metric that compares the '*traditional SS watermark*' and the '*eigen watermark*'. Each data point was collected through out 1000+ audio pieces that embedded with the watermark with simulated reverberation effect added to the watermarked audio source.

and this constitutes the detection noise floor in the noiseless case (i.e., when $\mathbf{n} = 0$ in (3)). To completely remove this noise floor in the noiseless case, the host signal is slightly modified to remove the projection component of the host signal onto the watermark subspace. If we choose $\mathbf{w} = \mathbf{v}_1$, then the watermark embedding equation in (1) is modified to

$$\begin{aligned} \tilde{\mathbf{x}} &= \mathbf{x} - \langle \mathbf{x}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \eta \mathbf{v}_1 \\ &= \bar{\mathbf{x}} + \eta \mathbf{v}_1 \quad (\text{where, } \bar{\mathbf{x}} \triangleq \mathbf{x} - \langle \mathbf{x}, \mathbf{v}_1 \rangle \mathbf{v}_1) \end{aligned} \tag{10}$$

Note that, inorder to surpass the low pass filter (LPF) and high pass filter (HPF) we also restrict the innerproduct of the frequency space operation to be in a specific range,

$$\begin{aligned} \langle \mathbf{a}, \mathbf{b} \rangle &\equiv \langle \mathbf{a}, \mathbf{b} \rangle_{k_L \to k_H} \\ &= \sum_{k=k_L}^{k_H} \mathbf{a}_k \mathbf{b}_k \end{aligned} \tag{11}$$

In the rest of the paper, all of the frequency space operation will be implicitely denoted for a specific frequency range, $k_L \to k_H$. In Fig. 1, we show that Eigen Watermarking significantly improves the ROC of the detector and has a good operation point against the simulated reverberation effect.

3.2. Self-Correlation

The central idea of the proposed system is using self-correlation at the detector, rather than cross-correlation as in standard watermarking detectors. As noted in (7), the cross-correlation with watermark template requires perfect synchronization and perfect knowledge of the acoustic channel, otherwise it will be smeared by the alternating sign of the channel response. This stringent requirement is relaxed if self-correlation is used as described in this section.

Let $\mathbf{y}^a$ and $\mathbf{y}^b$ be two adjacent DCT blocks of the received signal, then self-correlation is defined as

$$\psi \triangleq \langle \mathbf{y}^a, \mathbf{y}^b \rangle \tag{12}$$

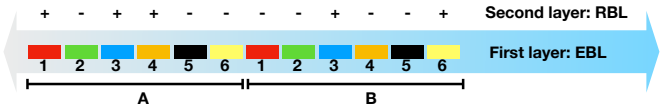

Fig. 2. Illustration of the bi-layered watermark encoding structure with $N_r = 2, N_s = 6$

If each block corresponds to an embedded watermarked block as in (1) after passing through the acoustic channel in (6), then

$$\begin{aligned}\psi &= \langle \tilde{\mathbf{x}}^a \odot \boldsymbol{\alpha} + \mathbf{n}^a, \tilde{\mathbf{x}}^b \odot \boldsymbol{\alpha} + \mathbf{n}^b \rangle \\ &\approx \sum_k \alpha_k^2 x_k^a x_k^b + \sum_k \alpha_k^2 w_k^a w_k^b + \sum_k n_k^a n_k^b \end{aligned} \quad (13)$$

where we assumed that the channel behavior does not change for adjacent blocks, and in the approximation we invoked the assumption of the absence of correlation between the watermark, signal, and additive noise. If the additive noise is zero-mean (which is usually the case), then the last term in (13) vanishes. If adjacent audio blocks are weakly correlated, then the first term in (13) is much weaker than the watermark component (which is the second term in (13)). However, this component might become significant if a music chord is present in the host signal, and that increases the noise floor.

Note that, by employing self-correlation the impact of acoustic channel is neutralized (by making the channel contribution nonnegative) at the cost of higher noise floor due to the host signal self-correlation. The noise-floor is significantly reduced through the sign modulation scheme that is described in the following section.

3.3. Sign-correction encoding/decoding method

The second central component in the proposed watermarking system is the sign-modulation of adjacent blocks in the host signal. A second encoding layer by an ± 1 sequence is applied to modify the binary phase of the watermark in each block. The entire encoded audio sequence is written,

$$\tilde{x} = \bigoplus_{n=1}^{N_r} \bigoplus_{i=1}^{N_s} \left(\bar{x}^{n,i} + \beta \, s_{n,i} \, g_{n,i} \, w^i \right) \quad (14)$$

where $\bigoplus$ denotes block concatenation, N_s denotes the number of segments of basic watermark building blocks, N_r denotes the number of repeats of the set of segments, β is the encoding strength, $g_{i,n}$ is the segment normalization factor, and $s_{i,n}$ is a ± 1 random sequence. An illustration of this encoding process is shown in Fig. 2. Note that, each subblock is modulated by a random sign that will be incorporated at the decoder. Different keys could be used for the generation of the watermark and the sign sequence to allow for increased accuracy or multiple access watermarking.

The decoder modifies the self-correlation procedure in (12) to accommodate multilayered embedding in (14). The multilayered self-correlation has the form

$$\Psi = \sum_{i=1}^{N_s} \sum_{n=1}^{N_r-1} \sum_{m=n+1}^{N_r} \frac{s_{m,i} \, s_{n,i} \, \langle \mathbf{y}^{m,i}, \mathbf{y}^{n,i} \rangle}{g_{m,i} \, g_{n,i}} \quad (15)$$

where

$$g_{m,i} \equiv \sqrt{\langle \mathbf{y}^{m,i}, \mathbf{y}^{m,i} \rangle}, \quad (16)$$

is a normalization factor for the segment vector, $\mathbf{y}^{m,i}$, and $\mathbf{y}^{m,i}$ is the segment of audio from the receiver.

Note that, with this sign modulation arrangement in the encoder and the decoder, the watermark component in (13) is invariant, while the signal component is effectively suppressed. Assuming for now that we have perfect synchronization, we describe how Ψ in (15) behaves under signal and null hypotheses. We have

$$\Psi \equiv \sum_{i=1}^{N_s} \sum_{n=1}^{N_r-1} \sum_{m=n+1}^{N_r} \frac{\psi_{m,n,i}}{g_{m,i} \, g_{n,i}} \quad (17)$$

where

$$\begin{aligned}\psi_{m,n,i} &= s_{m,i} \, s_{n,i} \, \langle \mathbf{y}^{m,i}, \mathbf{y}^{n,i} \rangle \\ &= s_{m,i} \, s_{n,i} \, \langle \bar{\mathbf{x}}^{m,i} \odot \boldsymbol{\alpha} + \mathbf{n}, \bar{\mathbf{x}}^{n,i} \odot \boldsymbol{\alpha} + \mathbf{n} \rangle \\ &+ \beta^2 \, s_{m,i}^2 \, s_{n,i}^2 \langle \mathbf{w}^i \odot \boldsymbol{\alpha}, \mathbf{w}^i \odot \boldsymbol{\alpha} \rangle \\ &+ \beta \, s_{m,i} \, s_{n,i}^2 \, \langle \bar{\mathbf{x}}^{m,i} \odot \boldsymbol{\alpha} + \mathbf{n}, \mathbf{w}^i \odot \boldsymbol{\alpha} \rangle \\ &+ \beta \, s_{m,i}^2 \, s_{n,i} \, \langle \mathbf{w}^i \odot \boldsymbol{\alpha}, \bar{\mathbf{x}}^{n,i} \odot \boldsymbol{\alpha} + \mathbf{n} \rangle,\end{aligned}$$

Note that, fractional delay of the block boundaries is represented as part of the channel. Under the null hypothesis, $\mathcal{H}_0$, i.e., when $\beta = 0$, we get the noise signature

$$\rho_0(t) = \sum_{i=1}^{N_s} \sum_{n=1}^{N_r-1} \sum_{m=n+1}^{N_r} \frac{s_{m,i} \, s_{n,i} \, \langle \bar{\mathbf{x}}^{m,i} \odot \boldsymbol{\alpha} + \mathbf{n}, \bar{\mathbf{x}}^{n,i} \odot \boldsymbol{\alpha} + \mathbf{n} \rangle}{g_{m,i} g_{n,i}} \quad (18)$$

Under the signal hypothesis, and after invoking the assumption of non-correlation between signal/watermark/noise, we get the signal signature

$$\rho_1(t) = \rho_0(t) + \sum_{i=1}^{N_s} \sum_{n=1}^{N_r-1} \sum_{m=n+1}^{N_r} \frac{\beta \, s_{m,i}^2 \, s_{n,i}^2 \langle \mathbf{w}_i \odot \boldsymbol{\alpha}, \mathbf{w}_i \odot \boldsymbol{\alpha} \rangle}{g_{m,i} g_{n,i}} \quad (19)$$

An illustration of the modulated self-correlation behavior under both hypothesis is illustrated in Fig. 3.

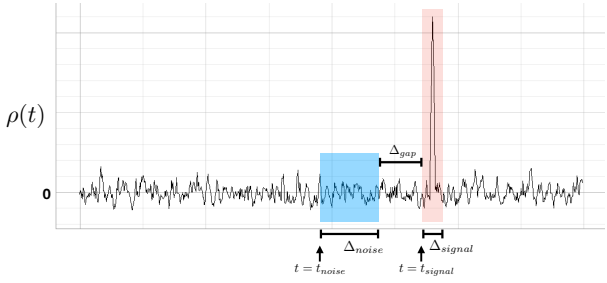

Fig. 3. Illustration of modulated self-correlation under both hypothesis

The difficulty of decoding with self-correlation is that the mean under $\mathcal{H}_1$, and the variance under both hypothesis are dependent on the unknown channel parameters $\boldsymbol{\alpha}$. Nevertheless, there is about 20-30 dB difference in the mean under both hypothesis, which provides flexibility to choose the detection threshold with good overall performance. In our system, the detection threshold is set to be significantly higher, e..g., +10 dB, than the noise floor over a long period

of time. Hence, the detection threshold itself is a function of the acoustic channel.

The above discussion assumed synchronization was achieved prior to self-correlation. In the worst case, synchronization could be achieved at a sample-level by brute-force computation of $\rho(t)$ (where fractional delay is absorbed in the channel response). Nevertheless, it was found that the modulated self-correlation mechanism tolerates imperfect alignment (roughly $\pm 50\%$ of the length of eigenvector length) with acceptable detection rate. For example, if we take the eigenvector to be 10 ms long, a ± 5 ms misalignment can be tolerated. This is due to the blind detection procedure that parameterizes the detection parameters with the channel, and small misalignments can be modeled as part of the channel. The tradeoff between complexity and performance could be further exploited by incorporated only a subset of blocks (out of N_r) and subblocks (out of N_s) in (17).

3.4. System Overview

The overall detection procedure proceeds as follows (where $\rho(t)$ is computed as in the previous section):

1. Calculate the noise-mean throughout the noise region,

$$\bar{\rho}_0 \equiv \frac{1}{\Delta_n} \sum_{t=t_n}^{t_n+\Delta_n} \rho(t) \tag{20}$$

2. Calculate the channel dependent noise variance,

$$\sigma_0^2 \equiv \frac{1}{\Delta_n} \sum_{t=t_n}^{t_n+\Delta_n-1} |\rho(t) - \bar{\rho}_0|^2 \tag{21}$$

3. Set the detection threshold, γ, at the desired point on the ROC curve, e.g., $\gamma = 3\sigma_0$.
4. Calculate the noise-mean corrected signal-mean through out the signal region,

$$\bar{\rho}(t) \equiv \frac{1}{\Delta_s} \sum_{\tau=0}^{\Delta_s-1} \Big(\rho(t+\tau) - \bar{\rho}_0\Big) \tag{22}$$

5. The detector operates as

$$\varepsilon(t) = \begin{cases} 0, \text{ for } \bar{\rho}(t) < \gamma, \\ 1, \text{ for } \bar{\rho}(t) \geq \gamma, \end{cases} \tag{23}$$

Note that, the frequency of computing $\bar{\rho}(t)$ and $\varepsilon(t)$ is determined by the available computation resources.

4. EXPERIMENTAL RESULTS

The first experiment aims at computing the Receiver Operating Characteristic (ROC) curve, which fully captures the detector performance [17]. The ROC curver is computed under various reverberation conditions, and study the impact of watermark length (i.e., the latency). For false accept rate calculation, we scan through a non-watermarked audio ($\sim$41 min) every 5 millisecond to count the watermark detection. For the detection part, the watermark is inserted every 4 seconds in the same host audio (i.e. 600 watermarks). Multiple watermarks with different duration are simultaneously inserted in the host audio. This has minor impact since the watermarks are mutually orthogonal. To study reverberation effect, we apply many reverberations filters for true measured impulse response to the watermarked audio prior to the detector.

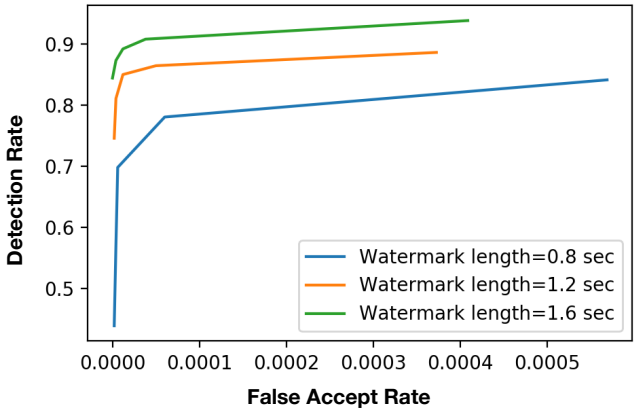

Fig. 4. The ROC metric calculations for the perofrmance of different watermark length.

Fig. 4 shows the ROC for different watermark performance associated versus watermark length. Note that, we zoomed in the horizontal axis of the ROC curve. The system was also evaluated versus standard audio processing operations, e.g., lowpass filtering, highpass filtering, mp4 compression, and mp3 compression, and showed the standard robust performance of spread spectrum systems [13]. The algorithm was also implemented with a python-based real-time demo with a consumer-grade loudspeaker and microphone. The performance was almost perfect for distances more than 20 feet for many watermark durations, and with latency less than 1 second in all cases even in the presence of household noise. The quality of watermarked audio was evaluated by 10 expert listeners and was shown to be indistinguishable from original audio.

5. CONCLUSION

We presented a spread-spectrum based audio watermarking system that is robust to reverberation and desynchronization. The performance is evaluated using both simulation and real-time evaluations. The detector runs asynchronously and deploys sign-modulated self correlation between adjacent audio blocks, which renders it blind to channel reverberation. The complexity of the encoder and decoder is reasonable for embedded implementation, and the latency is less than 1 second. Further, the complexity is scalable according to the available resources with graceful degradation in the performance.

The two bottlenecks of the proposed algorithm are the synchronization search and sensitivity to clock drift when long blocks are used. In a subsequent work, we developed a dynamic programming algorithm to prune the synchronization search by an order of magnitude with negligible impact on the detection performance. The algorithm has shown good performance for clock drift up to 100 ppm, and the robustness can be increased by deploying shorter blocks.

6. REFERENCES

[1] Mitchell D Swanson, Bin Zhu, and Ahmed H Tewfik, "Audio watermarking and data embedding–current state of the art, challenges and future directions," in *Multimedia and Security Workshop at ACM Multimedia*. Citeseer, 1998, vol. 41.

[2] Guang Hua, Jiwu Huang, Yun Q Shi, Jonathan Goh, and Vrizlynn LL Thing, "Twenty years of digital audio watermarking?a comprehensive review," *Signal Processing*, vol. 128, pp. 222–242, 2016.

[3] Pablo Cesar, Dick CA Bulterman, and Jack Jansen, "Leveraging user impact: an architecture for secondary screens usage in interactive television," *Multimedia systems*, vol. 15, no. 3, pp. 127–142, 2009.

[4] Heinrich Kuttruff, *Room acoustics*, Crc Press, 2016.

[5] Mohamed F Mansour and Ahmed H Tewfik, "Time-scale invariant audio data embedding," *EURASIP Journal on Applied Signal Processing*, vol. 2003, pp. 993–1000, 2003.

[6] Chi-Man Pun and Xiao-Chen Yuan, "Robust segments detector for de-synchronization resilient audio watermarking," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 11, pp. 2412–2424, 2013.

[7] Xiang-Yang Wang and Hong Zhao, "A novel synchronization invariant audio watermarking scheme based on dwt and dct," *IEEE Transactions on signal processing*, vol. 54, no. 12, pp. 4835–4840, 2006.

[8] Yong Xiang, Iynkaran Natgunanathan, Song Guo, Wanlei Zhou, and Saeid Nahavandi, "Patchwork-based audio watermarking method robust to de-synchronization attacks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 9, pp. 1413–1423, 2014.

[9] Andrew Nadeau and Gaurav Sharma, "An audio watermark designed for efficient and robust resynchronization after analog playback," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 6, pp. 1393–1405, 2017.

[10] Giovanni Del Galdo, Juliane Borsum, Tobias Bliem, Alexandra Craciun, and Stefan Krägeloh, "Audio watermarking for acoustic propagation in reverberant environments," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2364–2367.

[11] Xia Zhang, Di Chang, Wanyi Yang, Qian Huang, Wei Guo, and Yanbin Zhao, "An audio digital watermarking algorithm transmitted via air channel in double dct domain," in *Multimedia Technology (ICMT), 2011 International Conference on*. IEEE, 2011, pp. 2926–2930.

[12] Ingemar J Cox, Joe Kilian, Tom Leighton, and Talal Shamoon, "Secure spread spectrum watermarking for images, audio and video," in *Image Processing, 1996. Proceedings., International Conference on*. IEEE, 1996, vol. 3, pp. 243–246.

[13] Darko Kirovski and Henrique S Malvar, "Spread-spectrum watermarking of audio signals," *IEEE transactions on signal processing*, vol. 51, no. 4, pp. 1020–1033, 2003.

[14] Roland E Best, *Phase locked loops: design, simulation, and applications*, McGraw-Hill Professional, 2007.

[15] Stephen A Martucci, "Symmetric convolution and the discrete sine and cosine transforms," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1038–1051, 1994.

[16] Roger A Horn, Roger A Horn, and Charles R Johnson, *Matrix analysis*, Cambridge university press, 1990.

[17] Steven M Kay, "Fundamentals of statistical signal processing, vol. ii: Detection theory," *Signal Processing. Upper Saddle River, NJ: Prentice Hall*, 1998.