

Robust Watermarking of Music Signals by Cepstrum Modification

Kaliappan Gopalan

Department of Electrical and Computer Engineering
Purdue University Calumet
Hammond, IN 46323
email: gopalan@calumet.purdue.edu

Abstract— A method of embedding a predetermined watermark in an audio signal is described for audio music copyright protection applications. The proposed technique applies the psychoacoustical masking property of the human auditory system to imperceptibly embed the watermark in the cepstral domain of a host audio signal. Embedded watermark is extracted using an oblivious detection technique without resorting to any correlation procedure. Experimental results show that the watermark is robust to bandpass filtering and additive noise at low power levels.

I. INTRODUCTION

Audio watermarking is concerned with the insertion of a signal of known information or characteristics in an audio signal in an imperceptible way. Detection of the embedded watermark helps in authenticating the audio, identifying illegal copies of the audio, and detecting unauthorized changes made to the audio. While data embedding in an audio signal, or audio steganography, resembles audio watermarking in many ways, the former has applications in covert and/or secure communication of battlefield information, confidential financial transactions, etc., requiring a large payload capacity. Watermarking, on the other hand, is primarily used for copyright protection of digital products that require embedding a small amount of information.

To be effective in the general application of copyright protection, a watermarking technique must satisfy three major criteria, namely, the watermarked audio is perceptually indistinguishable from the original audio, the watermark is robust so that a user is unable to extract the watermark without destroying the audio, and the watermark is unambiguously retrieved by the copyright owner to establish ownership. These requirements, which are also common to data embedding and steganography, cannot all be met simultaneously for any practical application. Because of this challenge, a large body of work has been reported over the past decade with varying degrees of success for steganography and watermarking applications [1-3]. Most of

these techniques directly rely on the psychoacoustic masking property of the human auditory system either in the temporal or spectral domain. Inaudibility of embedding is achieved by altering the amplitudes of the audio signal at spectrally or temporally masked points. Alternatively, phase of the signal may be altered using all-pass filters as hearing is insensitive to absolute phase [4]. Yet other techniques exploit the perceptual masking phenomenon indirectly by modifying speech samples based on their power levels [5, 6]. This paper reports on watermarking a music audio signal by modifying samples in the cepstral domain at perceptually masked spectral points. In the following sections we briefly review the cepstral domain audio processing and embedding, and present a method of modification of cepstra with frequency masking for robust and imperceptible watermarking.

II. CEPSTRAL DOMAIN SPEECH PROCESSING AND EMBEDDING

A. Cepstral Domain Audio Processing

Cepstral domain features have been used extensively in speech and speaker recognition systems, and speech analysis applications. Complex cepstrum $\hat{x}[n]$ of a frame of speech $\{x[n]\}$ is defined as the inverse Fourier transform of the complex logarithm of the spectrum of the frame, as given by

$$\hat{x}[n] = IDFT \left[\ln \{ DFT(x[n]) \} \right] \quad (1)$$

Denoting the Fourier transform of $\{x[n]\}$ by

$$DFT\{x[n]\} = X(e^{j\omega}), \text{ we have,}$$

$$\ln \left[X(e^{j\omega}) \right] = \ln \left[E(e^{j\omega}) \right] + \ln \left[H(e^{j\omega}) \right], \quad (2)$$

where

$$X(e^{j\omega}) = DFT\{x[n]\} = E(e^{j\omega})H(e^{j\omega}), \quad (3)$$

The work on this paper was supported by the Air Force Research Laboratory, Air Force Material Command, USAF, under research grant/contract number F30602-03-1-0070)

and $e(n) \leftrightarrow E(e^{j\omega})$ and $h(n) \leftrightarrow H(e^{j\omega})$ are the discrete Fourier transform pairs from the speech signal production model, $x(n) = e(n)*h(n)$. The cepstrum in (1) effectively separates the excitation source $e(n)$ from the vocal tract system $h(n)$.

B. Embedding in the Cepstral Domain

The additive separation in (2) indicates that modification for data embedding can be carried out in either of the two parts of speech. Imperceptibility of the resulting cepstrum-modified audio from the original audio may depend upon the extent of changes made to the pitch and/or the formants, for instance. Any modification carried out in the cepstral domain in accordance with data alters the speech source, system, or both, depending on the frequencies involved. In the case of music, changing the cepstrum must correspond to a frequency range that retains the perceptual quality. This generally requires lower frequencies where the source is unaffected.

Prior work employing cepstral domain feature modification for embedding includes statistical mean manipulation [7], and adding pseudo random noise sequence for watermarking [8]. More recently, Hsieh and Tsou [9] showed that by modifying the cepstral mean values in the vicinity of rising energy points, frame synchronization and robustness against attacks can be achieved. Alternatively, altering the mean cepstrum in two consecutive ranges of frequencies in a nonreturn-to-zero mode has been shown to increase embedding capacity with low data recovery error rate [10]. Based on the robustness of cepstrum to signal processing operations, the present work proposes watermarking music signals by altering their cepstrum – rather than the mean cepstrum – in regions that are psychoacoustically masked.

III. CEPSTRUM MODIFICATION FOR WATERMARKING MUSIC AT MASKED FREQUENCIES

A. Embedding Watermark

For imperceptibility in hearing, cepstrum of a music signal in the present work is modified in the spectrally masked regions in accordance with a chosen watermark. This modification is carried out in a two-step procedure. In the first step, masked frequency points are determined for each frame of the host music signal. These points are determined from the global masking threshold and normalized sound pressure level (SPL) of each frame [11]. If the SPL at a spectral point is below the corresponding masking threshold at the frequency index by a certain dB, that point belongs in a set of potential points for embedding the watermark. After determining the set of embeddable spectral points in all frames, those points that occurred most commonly in all of the host signal frames are obtained. From this set, a pair of masked frequencies and the frames in which both of these frequencies occurred are selected. These frequencies and the corresponding frames form the

watermark key, and the watermark size is determined by the number of frames in which the selected frequency pair occurs in masked regions.

In the second step, complex cepstrum of a sinusoid at each of the two selected frequencies $f1$ and $f2$ are obtained with the maximum amplitude of the sinusoid set to the full quantization level of the given host signal. For each frame of music that is to be embedded with watermark data, its complex cepstrum is modified as follows.

Initialize: Spectrum at $f1$ and $f2$ = mean of frame spectrum at $f1$ and $f2$

$$\text{To embed a 1: } \text{mod_cep} = \text{cep} + \alpha.c1 - \beta.c2 \quad (4a)$$

$$\text{To embed a 0: } \text{mod_cep} = \text{cep} - \alpha.c1 + \beta.c2, \quad (4b)$$

where

cep = original cepstrum of frame

$c1$ = cepstrum of sinusoid at frequency $f1$, and

$c2$ = cepstrum of sinusoid at frequency $f2$

The parameters α and β are set to low values (one-tenth, empirically, for example), or based on a fraction of frame power. Since the two frequencies are in the masked regions of the selected frames, adding or subtracting low level cepstra at these frequencies ensures that the modification results in minimal perceptibility in hearing. If no bit is to be embedded, as in the case of frames that do not have the frequency pair in their masked regions, only the initialization step is carried out.

Modified frame cepstrum is transformed to time domain and quantized to the same number of bits as the host signal for storage and transmission.

B. Watermark Detection

Embedded watermark information in each frame is retrieved using the spectral ratio at the two frequencies, $f1$ and $f2$. Since an unembedded frame is stored with the same spectral magnitude at $f1$ and $f2$, the ratio is close to unity; hence, no bit is retrieved. At other (key) frames, the recovered bit rb is given by

$$rb = \left\{ \begin{array}{l} 1, \text{ if } \log \left| \frac{X(f1)}{X(f2)} \right| \geq b1 \\ 0, \text{ if } \log \left| \frac{X(f2)}{X(f1)} \right| \geq b0 \\ -1 \text{ (no data), else} \end{array} \right\}, \quad (5)$$

where $b0$ and $b1$ are set close to unity. We note that the indices of the embedded frames are not needed for watermark recovery; however, checking for the log spectral ratio only for the embedded frames speeds up the detection process. In addition, the indices of frames selected for

embedding serve as a second key for added security of watermark from removal attempts.

IV. EXPERIMENTAL RESULTS

Using the above two-step procedure, a CD quality music segment was embedded with a watermark of size up to 259 bits over a duration of approximately 3.2 s. The host music signal was sampled at 44100 samples/s with 16 bits per sample. A frame size of 512 samples were used with 256-sample overlap. To prevent the watermark from being removed by lowpass or highpass filtering, frames with masked frequencies in the range of 1000 Hz to 5000 Hz were determined in the first step. Of the 549 frames in the host segment, 10 frequencies were found to have 50 percent or more of the frames in the masked regions. The pair $f1 = 1894.9$ Hz (frequency index = 22) and $f2 = 3617.6$ Hz (index = 42), both of which occurred simultaneously in 259 frames, were chosen for cepstrum modification. We note that the normalized SPL at these frequencies were lower than their hearing threshold levels by 3 dB or more. Hence, a small change in SPL due cepstrum modification cannot affect perceptibility.

Cepstrum of each frame in which the selected frequencies occurred in the masked region was modified in accordance with (4) after initialization. To test the imperceptibility and watermark recovery, (a) bits of all 0's, and (b) all 1's were first used as watermarks, with $\alpha = \beta = 0.1$ in (4). The cepstrum-modified frames were transformed to time domain and quantized to 16 bits. Informal listening tests of the watermarked music by a group of listeners showed no noticeable difference from the original host music. Employing $b1 = b0 = 1.1$ in (5), all the 259 bits of the watermark in both cases were recovered correctly. Figure 1 shows a segment of the original and cepstrum-modified signals corresponding to a continuous set of 25 frames that were watermarked.

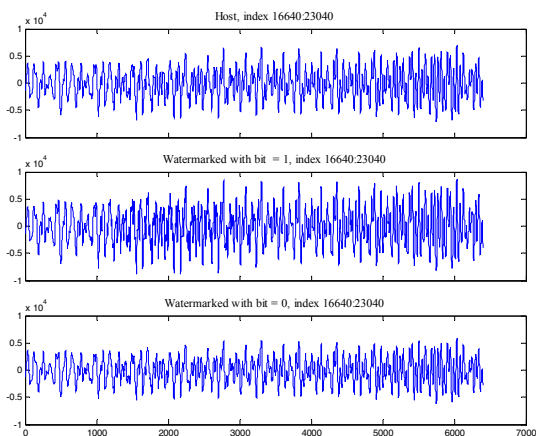


Figure 1. A segment of host signal (top) and its watermarked versions with bits of 1 (middle) and 0 (bottom)

To quantify inaudibility of watermarking, spectrograms of the segments of the three signals corresponding to the 25 frames that were modified are shown in Figure 2. These spectrograms are shown only for the frequency range of 1200 Hz to 4200 Hz in which the watermarking frequencies are present.

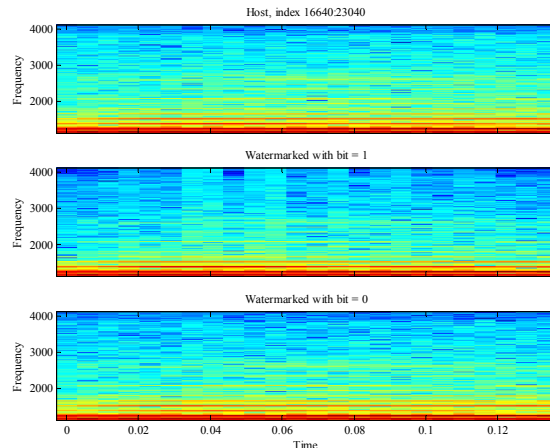


Figure 2. Spectrograms of a segment of host signal (top) and its watermarked versions with bits of 1 (middle) and 0 (bottom)

Watermarking the music signal with a random set of 259 bits yielded similar results of undetectability in hearing perception and accurate and recovery of the embedded watermark. Figure 3 shows the spectrograms of the host and the watermarked signals in the 1200 Hz to 4200 Hz range. Because the modified cepstrum spreads the changes throughout the spectrum, the spectrogram of the watermarked signal reflects the modification slightly over the entire Nyquist range. The low power levels of the change, however, prevent audibility.

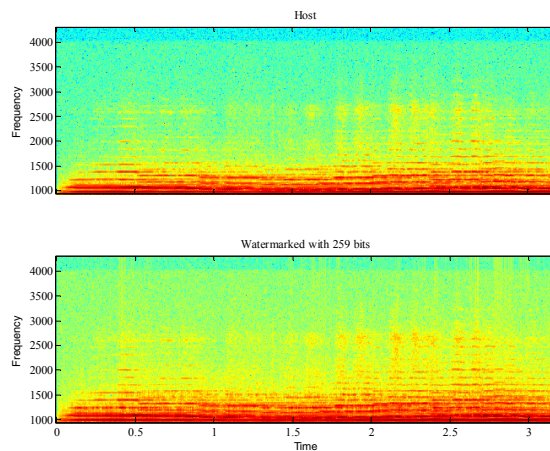


Figure 3. Spectrograms of the host (top) and watermarked signal with 259 bits of random data (bottom)

V. ROBUSTNESS OF EMBEDDED WATERMARK

Unlike in covert communication, a watermarked music signal cannot be attacked for removing the embedded information since any attack is likely to destroy the quality of the resulting music. Still, a CD quality music may be filtered with a bandpass filter to telephone bandwidth, for example, for low quality. In such a case, preserving the original watermark may be necessary for authentication of ownership. Because the watermark is embedded in the midband, it is impervious to bandpass filtering. To prevent tampering of watermarked signal for illegal copying and transmission – by filtering out a band of frequencies, for example – the watermark may be spread over a wide band. With a heavy concentration of energy in the lower frequency band for the signal used (Figure 3, top), however, removing the lower band by filtering creates a poor quality music.

Random noise, unintentionally added to the watermarked signal, affected the retrieved watermark depending on the signal-to-noise ratio (SNR) based on frame power. At an SNR of 40 dB, for example, three bits were in error in the recovered watermark. At lower SNR, music quality of the noisy watermarked signal was severely affected and, consequently, the detected watermark showed many bits in error.

VI. CONCLUSION

A method of watermarking music signals for copyright protection applications has been proposed. By determining perceptually masked spectral points in each frame of a music signal and altering the frame cepstrum at a pair of commonly occurring masked frequencies, it has been shown that an imperceptible watermark can be embedded. Cepstral changes at specified frequencies, unlike direct modification of masked spectral points, spread the change throughout the spectral domain; hence, the embedded watermark is hard to detect in spectrograms, and the low power changes do not cause noticeable difference in hearing perception. The changes in frame cepstrum at two specific frequencies enable watermark detection without requiring correlation with the original watermark data or the host music signal. The oblivious detection is useful in preventing illegal insertion of watermarks on counterfeit signals, for example. The technique retains the watermark under bandpass filtering and at low levels of random noise. Music piracy by transmitting

filtered (low quality) versions can be thwarted by spreading the watermark over several pairs of frequencies spanning a wide range. Further work on this and on increasing the size of watermark on different instrument music signals is underway.

ACKNOWLEDGMENT

The author thanks Dr. Stanley Wennedt and Dr. Andrew Noga of AFRL, Rome, NY for their discussions on embedding detection and imperceptibility of audio.

REFERENCES

- [1] W. Bender, D. Gruhl, N. Morimoto and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, Vol. 35, Nos. 3 & 4, pp. 313-336, 1996.
- [2] M. D. Swanson, M. Kobayashi, and A.H. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proc. IEEE*, Vol. 86, pp. 1064-1087, June 1998.
- [3] D. Kirovski and H. Malvar, "Spread-spectrum watermarking of audio signals," *IEEE Transactions on Signal Processing*, Vol. 51, Issue: 4, pp.1020 – 1033, April 2003.
- [4] T. Ciloglu and S. U. Karaaslan, "An improved all-pass watermarking scheme for speech and audio," *Proc. IEEE International Conference on Multimedia and Expo, 2000 (ICME 2000)*, Vol. 2, pp. 1017 – 1020, July 2000.
- [5] K. Gopalan, S. Wennedt, A.Noga, D Haddad, and S. Adams, "Covert speech communication via cover speech by tone insertion," *Proc. 2003 IEEE Aerospace Conference*, Vol. 4, pp. 4_1647 -- 4_1653, March 2003.
- [6] K. Gopalan, "Audio steganography using bit modification," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '03)*, Vol. 2, pp. 421-424, April 2003.
- [7] X. Li and H.H. Yu, "Transparent and robust audio data hiding in cepstrum domain," *Proc. IEEE International Conference on Multimedia and Expo, (ICME 2000)*, New York, NY, 2000.
- [8] S.K. Lee and Y.S. Ho, "Digital audio watermarking in the cepstrum domain," *IEEE Trans. Consumer Electronics*, Vol. 46, pp. 744-750, August 2000.
- [9] C.-T. Hsieh and P.-Y. Tsou, "Blind cepstrum domain audio watermarking based on time energy features," *Proc. 14th International Conference on Digital Signal Processing*, 2002, vol. 2, pp. 705-708, July 2002.
- [10] K. Gopalan, "Cepstral Domain Modification of Audio Signals for Data Embedding: Preliminary Results," *Proc. of 16th Annual Symposium on Electronic Imaging -- Security, Steganography, and Watermarking of Multimedia Contents VI*, San Jose, CA, January 2004.
- [11] Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, Vol. 88, No. 4, pp. 451-513, Apr. 2000.