Performance Evaluation of Digital Audio Watermarking Algorithms

J. D. Gordy and L. T. Bruton
Department of Electrical and Computer Engineering
University of Calgary
2500 University Drive N.W.
Calgary, Alberta, Canada T2N 1N4

Abstract- **We propose an algorithm-independent framework for rigorously comparing digital watermarking algorithms with respect to bit rate, perceptual quality, computational complexity, and robustness to signal processing. The framework is used to evaluate five audio watermarking algorithms from the literature, revealing that frequency domain techniques perform well under the criteria.**

I. Introduction

In the past few years, a need has arisen for protecting copyright ownership of electronic media. Powerful and low-cost computers allow people to easily create and copy multimedia content, and the Internet has made it possible to distribute this information at very low cost. However, these enabling technologies also make it easy to illegally copy, modify, and redistribute multimedia data without regard for copyright ownership. A recent example of this problem is the controversy regarding piracy of high-quality music across the Internet in MPEG Layer III (MP3) format [1].

Digital watermarking is seen as a partial solution to the problem of protecting digital media, for it allows content creators to embed sideband data into a host signal, such as author or copyright information. Many techniques have been proposed for watermarking audio, image, and video, and comprehensive surveys of these technologies may be found in [2] and [3]. However, the literature lacks an effective means of comparing the different approaches. An evaluation framework was recently described, but is limited to digital image watermarking [4].

The goal of this paper is to present an algorithm-independent set of criteria for quantitatively comparing the performance of digital watermarking algorithms. This framework is then used to evaluate a selection of five audio watermarking algorithms from the literature. The paper is organized as follows. In Section II we present our evaluation criteria, and in Section III we provide experimental data and an analysis of the evaluated algorithms. Finally, in Section IV we summarize the results of this investigation.

II. Performance Evaluation Framework

In this section we provide a description of the performance evaluation framework.

A. Conventions

In order to provide a common basis to describe and compare the algorithms, the following conventions are employed. Let $x(n)$ represent a host signal of length N samples, divided into $B = N/M$ blocks of M samples each. One bit is embedded into each block. A block division was chosen because it conveniently allows for a variable number of bits to be embedded by adjusting the block size. $\tilde{x}(n)$ represents the watermarked signal, and $w(m) \in \{-1,+1\}$ is a bipolar binary sequence of bits to be embedded within the host signal, for $0 \le m \le B-1$. Finally, $\tilde{w}(m)$ represents the set of watermark bits extracted from the watermarked host signal.

B. Evaluation Criteria

Four criteria were carefully selected as part of the evaluation framework. They were chosen to reflect the fact that watermarking is effectively a communications system. In addition, the criteria are simple to test, and may be applied to any type of watermarking system (audio, image, or video). It is important to note that the requirements of a practical watermarking system vary between applications, and so one criterion may be more important in some situations than in others. For example, a low computational cost may be vital to ensure that an algorithm can be implemented in real time on a given DSP system. The criteria are described in the following subsections.

1) Bit Rate

Bit rate refers to the amount of watermark data that may be reliably embedded within a host signal per unit of time or space, such as bits per second or bits per pixel. A higher bit rate may be desirable in some applications in order to embed more copyright information. In this study, reliability was measured as the bit error rate (BER) of extracted watermark data. For embedded and extracted watermark sequences of length B bits, the BER (in percent) is given by the expression:

$$BER = \frac{100}{B}\sum_{n=0}^{B-1}\begin{cases}1, & \tilde{w}(n) \neq w(n)\\ 0, & \tilde{w}(n) = w(n)\end{cases}$$

2) Perceptual Quality

Perceptual quality refers to the imperceptibility of embedded watermark data within the host signal. In most applications, it is important that the watermark is undetectable to a listener or viewer. This ensures that the quality of the host signal is not perceivably distorted, and

does not indicate the presence or location of a watermark. In this study, the signal-to-noise ratio (SNR) of the watermarked signal versus the host signal was used as a quality measure:

$$SNR = 10 \cdot \log_{10} \left\{ \frac{\sum_{n=0}^{N-1} x^2(n)}{\sum_{n=0}^{N-1} [\tilde{x}(n) - x(n)]^2} \right\}$$

3) Computational Complexity

Computational complexity refers to the processing required to embed watermark data into a host signal, and / or to extract the data from the signal. Algorithm complexity is important to know, for it may influence the choice of implementation structure or DSP architecture. Although there are many ways to measure complexity, such as complexity analysis (or "Big-O" analysis), for practical applications more quantitative values are required [5]. In this study, actual CPU timings (in seconds) of algorithm implementations were collected.

4) Robustness to Signal Processing

Watermarked digital signals may undergo common signal processing operations such as linear filtering, sample requantization, D/A and A/D conversion, and lossy compression. Although these operations may not affect the perceived quality of the host signal, they may corrupt the watermark data embedded within the signal. It is important to know, for a given level of host signal distortion, which watermarking algorithm will produce a more reliable embedding. In this study, robustness was measured by the bit error rate (BER) of extracted watermark data as a function of the amount of distortion introduced by a given operation.

III. Experimental Results

In this section, we present some experimental results from applying the evaluation framework to a collection of audio watermarking algorithms.

A. Methodology

Five digital audio watermarking algorithms from the literature were implemented for this study, and are briefly summarized here. The focus of this study was on *public* watermarks, as defined in [6], because they are of more interest in practical applications. Echo coding works by encoding watermark bits as "echo" with an imperceptible delay period, while the phase coding algorithm replaces the short-term phase of an audio signal with a signature [7]. Direct sequence and frequency hopped spread spectrum (DSSS and FHSS) techniques spread watermark data using a bipolar pseudorandom (PN) sequence in the spatial or DCT domain, respectively [8], [9]. Finally, a complex technique using frequency masking properties of the human auditory system (HAS) was considered [10]. The parameters of each algorithm were adjusted so that the embedded watermark is undetectable to a listener. More comprehensive implementation details may be found in [11].

A set of ten wideband audio signals were used as host signals, representing five general classes of music: blues, classical, country, folk, and pop. This delineation was chosen because each class has different spectral properties. Each signal was sampled at 44.1 kHz, represented by 16 bits per sample, and ten seconds in length. The watermarking algorithms were implemented in MATLAB 5.3 under Linux, and in ANSI C using Visual C++ 6.0 under Windows NT 4.0, on an Intel Pentium PC running at 166 MHz. In all of the experiments presented below, a watermarking algorithm was used to embed a random sequence of watermark bits within each of the ten host signals, and then used to extract the bits. This process was repeated for each algorithm – host signal pair, and performed 100 times. The results were averaged for either individual algorithms or for host signals, depending on the experiment. For each run, a different random sequence of bits was generated.

B. Bit Rate

Fig. 1 shows a plot of the bit error rate (BER) as a function of block size M for four of the watermarking algorithms. The phase coding algorithm's decoder is not dependent upon block size, so its BER was zero for all M. It can be seen that, in general, the encoding becomes more reliable as the block size increases. Of the four, the frequency masking algorithm provided the lowest error rate for all block sizes considered. Note that, in general, the error rate approximates $Q(x)$, the complimentary error function given by the expression [12]:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(\frac{-u^2}{2} \right) du$$

where x is a function of the block size.

For a block size of at least $M = 2048$ samples, corresponding to a bit rate of approximately 20 bits per second, the BER was below one percent for each of the algorithms. Although in practice a much lower error rate may be desired, in subsequent experiments a block size of $M = 2048$ samples was used.

C. Perceptual Quality

Table I shows the signal-to-noise ratio for the five watermarking algorithms, obtained from each of the watermarked host signals. It is clear that the echo coding, phase coding, and frequency masking algorithms introduce the most distortion into the audio signals. However, as described in [10], the frequency masking technique "hides" the distortion according to masking properties of the HAS. DSSS and FHSS have a markedly higher SNR, due to the fact that the power of the noise-like watermark signal introduced had to be maintained at a very low level [7], [11].

D. Computational Complexity

Table II shows CPU timings for watermark embedding and extraction obtained from MATLAB and ANSI C implementations of the five watermarking algorithms, averaged over the ten host signals. For three of the algorithms, watermark extraction takes longer than watermark embedding. This is because for public watermarks, additional processing must be performed to compensate for the presence of the host signal [6]. However, for each of the C implementations, extraction takes less than ten seconds, indicating that the algorithms may be run in real time on the given platform. For consumer devices such as MP3 players, extraction speed is more important. Frequency masking embedding was the most expensive operation, because of the need to compute a complex, time-varying perceptual masking analysis on the host signal.

E. Robustness to Signal Processing

The watermarked signals were subjected to distortions representing common signal processing operations, but which do not seriously degrade the quality of the host signals. In particular, varying levels of distortion were introduced under lowpass filtering, additive white Gaussian noise (AWGN), sample requantization, median filtering, and lossy compression using MPEG Layer III, and for each processed signal the random bit sequence was extracted and compared to the original [13]. The results of these experiments are shown in Fig. 2 – 6. Purely time-domain algorithms – echo coding and DSSS – perform poorly under all of the signal distortions applied. This is not surprising, for the DSSS noise power and echo coding magnitude had to be maintained at low levels to keep the watermark undetectable. The best performance came from the frequency-domain phase coding, FHSS, and frequency masking algorithms, because these approaches tend to spread watermark data throughout the time domain of the host signal. Frequency masking localizes and maximizes watermark data within portions of the host signal preserved by the MPEG compression process, indicating why it performed well under the last experiment.

IV. Summary

In this paper, we have presented a straightforward performance evaluation framework for comparing digital watermarking algorithms based on bit rate, perceptual quality, computational complexity, and robustness to signal processing operations. We then implemented five audio watermarking algorithms from the literature, and applied the proposed framework to obtain a quantitative comparison of the techniques. Our results reveal that frequency-domain approaches to watermarking are more costly to implement than purely time-domain techniques, but provide a higher bit rate and more robust watermark extraction under signal processing.

TABLE I
SNR of Watermarked Audio Signals

Host Signal	Echo Coding	Phase Coding	DSSS	FHSS	Freq. Mask.
BLUES1	20.45	26.63	54.38	49.23	20.14
BLUES2	22.86	27.65	54.19	49.25	24.31
COUNTRY1	16.63	21.67	54.48	49.53	19.54
COUNTRY2	21.34	22.05	54.22	49.27	17.94
CLASSIC1	22.54	23.53	54.05	49.10	26.43
CLASSIC2	21.82	25.02	54.07	49.12	28.38
FOLK1	14.52	17.89	54.59	49.64	17.65
FOLK2	16.21	18.22	54.57	49.62	16.94
POP1	15.75	19.08	54.96	50.01	18.53
POP2	15.51	19.09	54.47	49.52	17.84
Mean	18.76	22.78	54.40	49.45	20.77

TABLE II
Timings for Audio Watermarking Algorithms

Watermarking Algorithm	Embedding Time (sec.)		Extraction Time (sec.)	
	MATLAB	ANSI C	MATLAB	ANSI C
Echo Coding	2.95	0.77	6.89	1.93
Phase Coding	8.42	2.27	2.11	0.55
DSSS	1.63	0.39	2.45	0.64
FHSS	16.30	4.56	38.05	9.27
Freq. Masking	351.46	91.38	39.25	9.42

References

[1] J. C. Davis, "Protecting intellectual property in cyberspace," *IEEE Technology and Society Magazine*, vol. 17, no. 2, pp. 12 – 25, Summer 1998.

[2] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1079 – 1107, July 1999.

[3] M. Swanson, M. Kobayashi, and A. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1064 – 1087, June 1998.

[4] M. Kutter and F. Petitcolas, "Fair benchmark for image watermarking systems," *Proceedings of SPIE Security and Watermarking of Multimedia Contents*, vol. 3657, pp 226 – 239, 1999.

[5] N. Wirth, *Algorithms and Data Structures*. Englewood Cliffs, NJ: Prentice Hall, 1986.

[6] S. Craver, N. Memon, B. Yeo, and M. Yeung, "Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 573 – 586, May 1998.

[7] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol 35, no. 3 – 4, pp. 313 – 335, 1996.

[8] R. Pickholtz, D. Schilling, and L. Milstein, "Theory of spread-spectrum communications – a tutorial," *IEEE Transactions on Communications*, vol. 30, no. 5, pp. 855 – 884, May 1982.

[9] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673 – 1687, December 1997.

[10] M. Swanson, B. Zhu, A. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Signal Processing*, vol. 66, no. 3, pp. 337 – 355, May 1998.

[11] J. Gordy, "Performance evaluation of digital watermarking algorithms," Master's thesis, University of Calgary, April 2000.

[12] B. Sklar, *Digital Communications: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice Hall, 1988.

[13] P. Noll, "MPEG digital audio coding," *IEEE Signal Processing Magazine*, vol. 14, no. 5, pp. 59 – 81, September 1997.

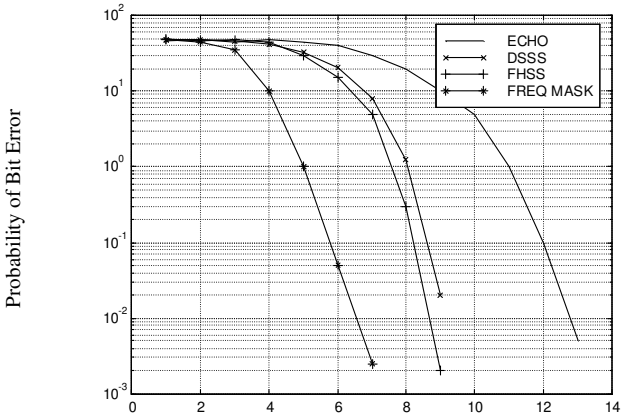

Figure 1: Bit error rate as a function of block size.

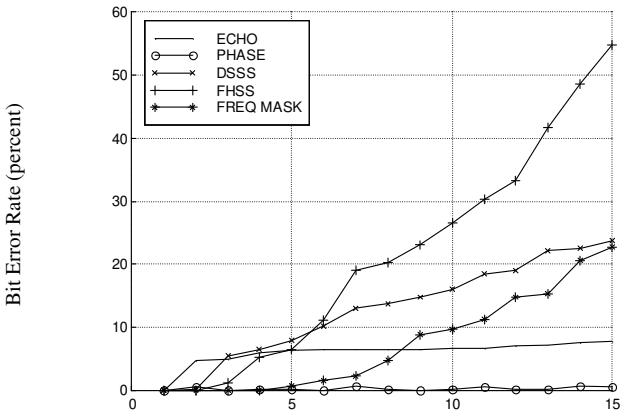

Figure 2: Bit error rate due to lowpass filtering, as a function of filter length.

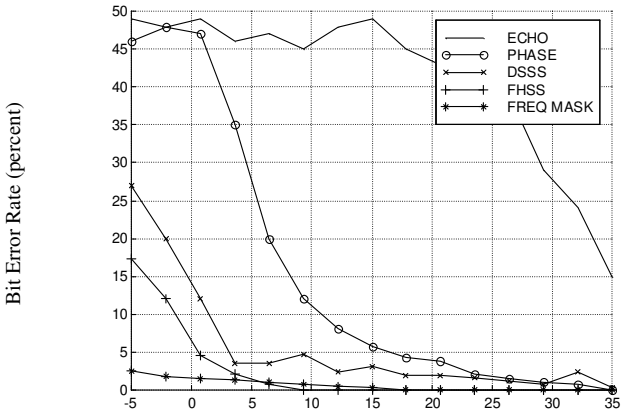

Figure 3: Bit error rate due to additive white Gaussian noise as a function of noise power.

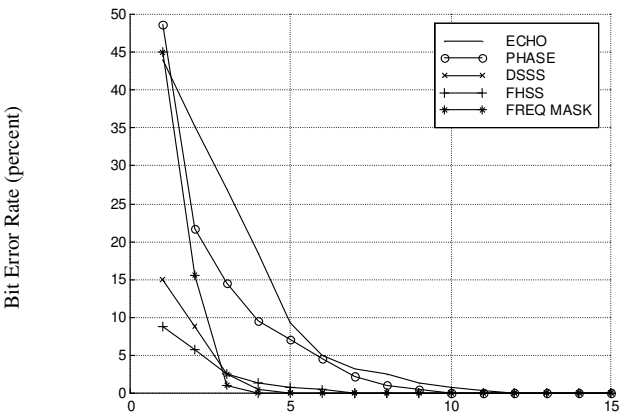

Figure 4: Bit error rate due to sample requantization as a function of bits per sample.

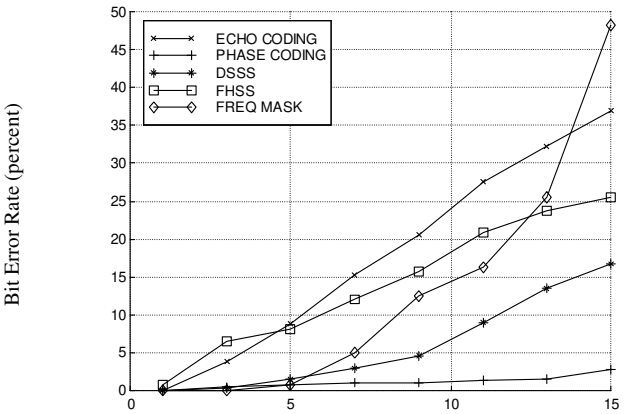

Figure 5: Bit error rate due to median filtering as a function of filter length.

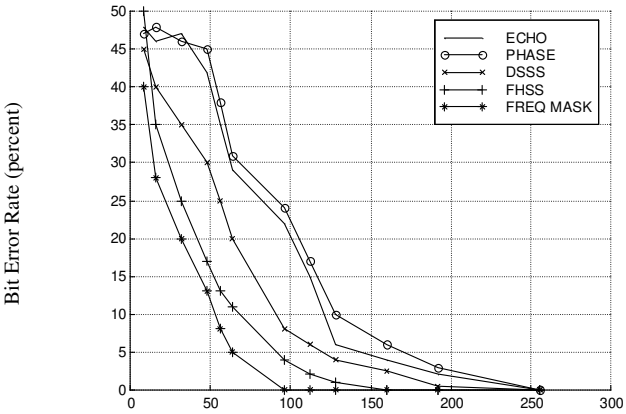

Figure 6: Bit error rate due to MPEG Layer III compression as a function of bit rate.