Audio Watermarking for Monitoring and Copy Protection

Jaap Haitsma, Michiel van der Veen, Ton Kalker and Fons Bruekers
Philips Research Laboratories
Prof. Holstlaan 4
5656 AA Eindhoven, The Netherlands
[jaap.haitsma][michiel.van.der.veen][ton.kalker][fons.bruekers]@ philips.com

ABSTRACT

Based on existing technology used in image and video watermarking, we have developed a robust audio watermarking technique. The embedding algorithm operates in frequency domain, where the magnitudes of the Fourier coefficients are slightly modified. In the temporal domain, an additional scale parameter and gain function are necessary to refine the watermark and achieve perceptual transparency. Watermark detection relies on the Symmetrical Phase Only Matched Filtering (SPOMF) cross-correlation approach. Not only the presence of a watermark, but also its cyclic shift is detected. This shift supports a multi-bit payload for one particular watermark sequence. The watermarking technology proved to be very robust to a large number of signal processing "attacks" such as MP3 (64 kb/s), all-pass filtering, echo addition, time-scale modification, resampling, noise addition, etc. It is expected that this approach may contribute in a wide variety of existing (e.g. monitoring and copy protection) and future applications.

Keywords

audio, broadcast monitoring, copy protection, watermark embedding, watermark detection

1. INTRODUCTION

A digital audio watermark is an information label, which is embedded in an audio signal in an imperceptible manner. During the past few years a number of new audio watermarking techniques have been developed to support applications such as copy control [1] [2] or broadcast monitoring [3]. Most of these operate in time domain and employ methods such as echo-hiding [4] or some kind of noise addition, exploiting temporal and/or spectral masking models of the human auditory system [5] [6].

Based on image and video watermarking techniques [3] [7] we have developed an alternative approach to audio watermarking. Similar to the work of Piva et al.[2], watermark embedding is performed in frequency domain. The principles of spectral masking are exploited in a relatively simple manner by slightly modifying magnitudes of the Fourier coefficients. The embedding algorithm is complemented with a detection procedure adapted from cross-correlation techniques used in image registration [9] and video watermarking [3] [8]. The combination of both algorithms offers several advantages in terms of robustness to some trivial signal processing "attacks" (e.g. all-pass filtering). In this paper, we introduce both embedding and detection algorithms and discuss briefly some key aspects such as payload, perceptual transparency, robustness and detection reliability.

2. EMBEDDING

A sketch of our watermark embedding algorithm is displayed in Figure 1. A random watermark sequence $W(k)$ is drawn from a normal distribution with mean and standard deviation of 0 and 1, respectively. A cyclic shifted version $W_s(k)$ is used to achieve a multi-bit payload for one particular watermark sequence $W(k)$. Every possible shift may be associated with a different information label. Therefore, payload is directly proportional to the watermark size (e.g. 1024-sample watermark corresponds to payload of maximum 10 bit).

The dominant part of the perceptually weighted watermark $w(n)$ is derived in the Fourier domain, where spectral masking is exploited in a relatively simple manner. First, the audio signal $x(n)$ is segmented into frames and transformed to the frequency domain. Here, the magnitude of its Fourier coefficients are slightly modified by utilizing the shifted watermark sequence $W_s(k)$:

$$W_i'(k) = W_s(k)X_i(k), \qquad (1)$$

where i indicates the frame number, $X_i(k)$ the spectral representation of the frame $x_i(n)$, and $W_i'(k)$ the resulting frequency domain watermark. Note that the frame size is a trade-off between perceptual transparency (small frame sizes) and detection reliability (large frame sizes). Several experiments have demonstrated that, in general, frame sizes of 2048-samples provide a good compromise in this trade-off.

Inverse Fourier transforms $\mathcal{F}^{-1}$ are used to reconstruct the time-domain watermark representation $w(n)$. Shaping the watermark in frequency domain (Equation 1) is not sufficient to assure perceptual transparency. Since fixed length Fourier transforms do not provide accurate time-localization, watermarks computed in frequency domain will spread in time over the entire analysis window. This may result in

ACM Multimedia Workshop Marina Del Rey CA USA
Copyright ACM 2000 1-58113-311-1/00/11...$5.00

perceptual distortions such as pre-echos. Therefore, an additional scale parameter α and gain function $g(n)$ are introduced to refine the watermark in the temporal domain:

$$y(n) = x(n) + \alpha g(n) w(n), \tag{2}$$

where α is the global scale parameter, $g(n)$ a data dependent gain function with values between 0 and 1, and $y(n)$ the watermarked audio.

Analog to the frame size, also α is a parameter that influences the trade-off between perceptual transparency and detection reliability: very small/large values of α may result in perceptual transparency/distortions and low/high watermark detection reliabilities. Several informal adaptive up-down listening tests [10] were performed on a variety of watermarked audio excerpts to extract critical values of α. We found perceptual transparency was achieved by selecting α between 0.15 and 0.25, depending on the audio excerpt.

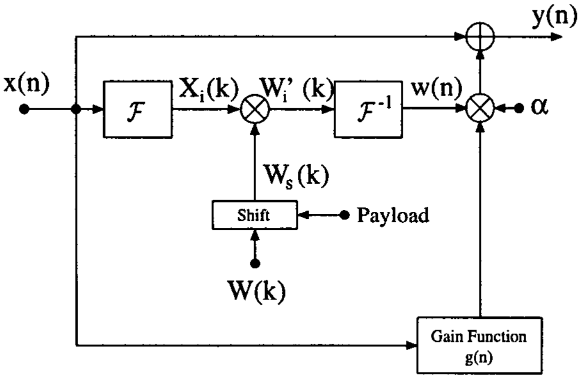

Figure 1: Overview of watermark embedding algorithm for digital audio. $\mathcal{F}$ and $\mathcal{F}^{-1}$ indicate Fourier and inverse Fourier transforms, respectively.

3. DETECTION

Figure 2 gives an overview of the watermark detection algorithm. It relies on a cross-correlation procedure between the watermark sequence $W(k)$ and the audio. Experiments revealed that filtering prior to cross-correlation may improve detection reliabilities significantly. In our detection algorithm, $y(n)$ is filtered with the "equalization" filter $d(n)$ according to:

$$\hat{y}(n) = y(n) * d(n), \tag{3}$$

with filter coefficients $d(n) = [\ -1\ \ 2\ \ -1\ ]$. This signal is segmented into frames and transformed to frequency domain to obtain the magnitude of the Fourier coefficients:

$$\hat{Y}_i(k) = |\ \mathcal{F}(\hat{y}_i(n))\ |, \tag{4}$$

where $\mathcal{F}$ indicates a Fourier transform operation. For each individual frame, the magnitude of Fourier coefficients $\hat{Y}_i(k)$ need to be cross-correlated with every possible shifted version of $W(k)$ to extract the payload. Such a cross-correlation is calculated most efficiently using Fourier transformed signals:

$$\hat{Y}_{i,F} = \mathcal{F}(\ \hat{Y}_i(k)\ ), \text{ and } W_F = \mathcal{F}(\ W(k)\ )^*. \tag{5}$$

The traditional cross-correlation may then be written as:

$$C_i = \mathcal{F}^{-1}\left(\hat{Y}_{i,F} \cdot W_F\right), \tag{6}$$

where C_i is the cross-correlation function. Similar to detection procedures in video watermarking [3], the detection performance may be enhanced by using the Symmetrical Phase Only Matched Filtering approach (SPOMF; [9]). In this cross-correlation procedure, only phase information of the signals $\hat{Y}_{i,F}$ and W_F is used:

$$C_i' = \mathcal{F}^{-1}\left(\mathcal{P}(\hat{Y}_{i,F}) \cdot \mathcal{P}(W_F)\right), \tag{7}$$

where $\mathcal{P}$ is a phase-only operation and $\mathcal{P}(x) = x/|x|$ for $x \neq 0$ and $\mathcal{P}(0) = 1$. To improve detection reliability even further, C_i' is accumulated over a period of time $C'_{sum} = \sum_i C_i'$. Since C'_{sum} is distributed normally its components may be normalized to the standard deviation σ:

$$C_n' = \frac{C'_{sum}}{\sigma(C'_{sum})}, \tag{8}$$

where C_n' is the normalized cross-correlation function. Its peak value, expressed in standard deviation σ, is related directly to the detection reliability, whereas its position corresponds to the cyclic shift (payload).

The detection reliability depends strongly on the number of accumulated frames. In general, cross-correlation functions C_i' need to be added over a period of 2 to 5 sec to exceed a detection threshold of 5σ. This corresponds to a false alarm probability of $2.9 \cdot 10^{-4}$. Figure 3 displays a typical cross-correlation function C_n'. In this example, a peak value of $\sim 13\sigma$ (false alarm probability of $6.3 \cdot 10^{-36}$) is detected at position 512.

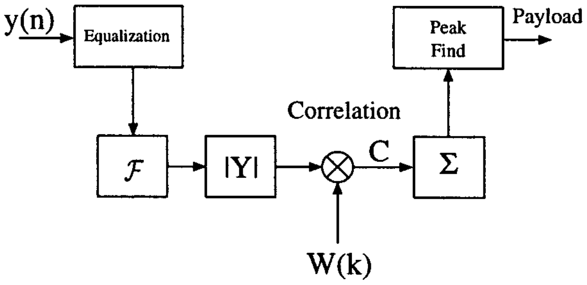

Figure 2: Overview of watermark detection

4. EXPERIMENTAL RESULTS

In a number of experiments we have examined the robustness of our audio watermark to a wide variety of signal "attacks". The following audio excerpts were used: (i) *O Fortuna* from Carl Orff, (ii) *Success has made a failure of our home* from Sinead O'Connor, (iii) *Say what you want* from Texas and (iv) *She works hard for the money* from Donna Summer. The 20 sec. audio fragments were sampled at 44.1 kHz (16 bit, mono). Based on up-down listening tests (section 2) we selected $\alpha = 0.2$ for watermark embedding (Equation 2). All audio excerpts were subjected to the following processing "attacks":

- **MP3 Encoding/Decoding** at 64 kb/s and 32 kb/s.

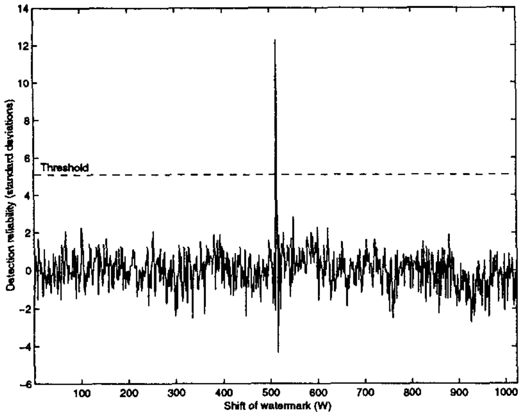

Figure 3: Example of cross-correlation function C'_n accumulated over a period of 5 sec. Dashed line indicates detection threshold of 5σ.

- **All-pass Filtering** using system function: $H(z) = (0.81z^2 - 1.64z + 1)/(z^2 - 1.64z + 0.81)$.
- **Amplitude Compression** with the following amplitude compression ratios: 8.94:1 for $|A| \geq -28.6$ dB; 1.73:1 for $-46.4 < |A| < -28.6$ dB; 1:1.61 for $|A| \leq -46.4$ dB.
- **Equalization** with a 10-band equalizer where signals within each band are suppressed or amplified by 6 dB.
- **Echo Addition** with a delay and decay of 100 ms and 50%, respectively.
- **Band-Pass Filtering** using a second order Butterworth filter with cut-off frequencies 100 Hz and 6000 Hz.
- **Time Scale Modification** of +4% or -4%, where the pitch is unaffected.
- **Resampling** consisting of subsequent down and up sampling to 22.05 kHz and 44.10 kHz, respectively.
- **Noise Addition** with uniform white noise. Maximum magnitude of 150 quantization steps.
- **D/A-A/D Conversions** using a commercial analogue tape recorder.

Processing was performed in MatLab and CoolEdit Pro 1.2. The detection results were calculated by accumulating cross-correlation functions C'_i (Equation 7) over periods of 5 sec and averaging the four detection reliabilities.

The results are displayed in Table 1. Unprocessed watermarked audio excerpts result in typical detection reliabilities between $\sim 13\sigma$ and $\sim 17\sigma$. MP3 compression at very low bit-rates (e.g. 32 kb/s) results in measurements close to the detection threshold of 5σ. The data reveal that detection reliability is affected only marginally by other signal attacks including MP3 compression at 64 kb/s and all-pass filtering. In general, reliabilities are in the range $11\sigma - 17\sigma$, corresponding to a false alarm probability of at least $1.9 \cdot 10^{-25}$.

Table 1: Detection reliabilities expressed in standard deviation σ.

Attack	Orff	Sinead	Texas	Donna
No Processing	13.6	13.4	13.3	17.1
MP3 (64kbit/s)	11.0	11.0	14.3	14.6
MP3 (32kbit/s)	6.0	5.6	6.5	6.7
All-pass Filtering	13.6	13.4	13.3	17.1
Amp. Compr.	13.4	13.1	11.5	17.8
Equalization	13.6	13.4	11.3	18.2
Echo Addition	13.3	12.9	12.9	16.2
Band-Pass Filter	11.7	11.5	13.1	14.3
Time Scale +4%	13.3	13.9	13.3	17.1
Time Scale -4%	14.1	12.5	13.3	16.5
Resampling	10.8	9.1	11.3	12.7
Noise addition	12.7	12.6	12.6	16.4
D/A A/D	11.7	10.5	12.6	11.7

5. CONCLUSIONS

Based on existing technology in image and video watermarking, we have developed new algorithms for embedding and detecting watermarks in digital audio. Important characteristics of this new technique were discussed. Key results of this study are:

1. **Embedding:** The dominant part of the perceptually weighted watermark is derived in frequency domain by slightly modifying the magnitude of Fourier coefficients. An additional scale parameter and time-domain gain function were necessary to refine the watermark. The scale parameter may also be utilized to tune system characteristics such as perceptual transparency and detection reliability.
2. **Detection:** The SPOMF cross-correlation approach offered a robust technology for blind detection of watermarks in digital audio.
3. **Robustness:** Our watermark algorithm proved to be robust to a wide variety of signal processing "attacks" such as MP3 (64 kb/s), all-pass filtering, echo addition, speed change, resampling, noise addition, etc.

With the accomplishments described in paper, and possible future developments, it is expected that our audio watermarking strategy can support a wide variety of existing (monitoring and copy control) and future applications.

6. REFERENCES

[1] E. Koch, and J. Zhao, 1995, "Towards robust and hidden image copyright labeling", in Nonlinear Signal Processing Workshop, Thessaloniki, Greece, pp. 452-455.

[2] A. Piva, M. Barni, and F. Bartolini, 1998, "Copyright protection of digital images by means of frequency domain watermarking", Proceedings of SPIE, vol. 3456, pp. 25-35.

[3] T. Kalker, G. Depovere, J. Haitsma, and M. Maes, 1999, "A video watermarking system for broadcast

monitoring", Proceedings of IS&T/SPIE/EI25, Security and Watermarking of Multimedia Content, vol. 3657, pp. 103-112.

[4] D. Gruhl, W. Bender, and A. Lu, 1996, "Echo-hiding", Information hiding: 1st International Workshop, R.J. Anderson, Ed., vol. 1174 of Lecture Notes in Computer Science, Isaac Newton Institute, England, pp. 295-315.

[5] P. Bassia, and I. Pitas, 1998, "Robust audio watermarking in the time domain", 9th European Signal Processing Conference (EUSIPCO98), Greece, pp. 25-28.

[6] M.D. Swanson, B. Zhu, A.H. Tewfik, and L. Boney, 1998, "Robust audio watermarking using perceptual masking", Signal Processing, vol. 66, 337-355.

[7] I. Cox, J. Kilian, F.T. Leighton, and T. Shamoon, 1996, "A secure, robust watermark for multimedia", In Proc. of the Information Hiding: First Int. Workshop, Lecture Notes in Computer Sciences, vol. 1174, R. Anderson, ed., Springer-Verlag, pp. 183-206.

[8] G.F.G. Depovere, T. Kalker, and J.P.M.G. Linnartz, 1998, "Improved watermark detection reliability using filtering before correlation", Int. Conf. on Image Processing, ICIP, Chicago IL.

[9] L.G. Brown, 1992, "A survey of image registration techniques", ACM Computing Surveys, vol. 24, pp. 325-376.

[10] H. Levit, 1970, "Transformed up-down methods in psychoacoustics", The Journal of the Acoustical Society of America, vol. 49, pp. 467-477.