A Robust Audio Watermarking Scheme Based on MPEG 1 Layer 3 Compression

David Megías
Jordi Herrera-Joancomartí
Julià Minguillón Alfonso

Estudis d'Informàtica i Multimèdia
Universitat Oberta de Catalunya

Seventh IFIP Conference on Communications and Multimedia Security
CMS2003

October 2-3, 2003
Turin, Italy

Outline

1. Introduction
2. Audio watermarking scheme
   a) Mark embedding
   b) Mark reconstruction
3. Performance evaluation
   a) Imperceptibility
   b) Capacity
   c) Robustness
4. Conclusions and future research

Outline

1. **Introduction**
2. Audio watermarking scheme
   a) Mark embedding
   b) Mark reconstruction
3. Performance evaluation
   a) Imperceptibility
   b) Capacity
   c) Robustness
4. Conclusions and future research

Introduction

- Watermarking is a copy detection technique.
- A mark is embedded into a digital medium in order to detect unauthorised copies.
- The most important properties of watermarking schemes are ***imperceptibility***, ***capacity*** and ***robustness***.
- The suggested scheme stems from a previous work in image watermarking which uses JPEG lossy compression.
- Here we use MPEG 1 Layer 3 (mp3) compression to determine the position of the embedded bits.

Outline

1. Introduction
2. **Audio watermarking scheme**
   - a) **Mark embedding**
   - b) **Mark reconstruction**
3. Performance evaluation
   - a) Imperceptibility
   - b) Capacity
   - c) Robustness
4. Conclusions and future research

Watermarking scheme

a) Mark embedding

- S (PCM) is the signal to be marked.
- $\mathrm{FFT}(S) \rightarrow S_F$
- $\mathrm{mp3co}(S) \rightarrow S_{\mathrm{mp3}}$; $\mathrm{mp3dec}(S_{\mathrm{mp3}}) \rightarrow S'$ (PCM)
- Parameter R (bit rate of the mp3 codec)
- $\mathrm{FFT}(S') \rightarrow S'_F$

Watermarking scheme

a) Mark embedding

$$F_{\text{rel}} = \left\{ f \in \left[0, \frac{f_{\max}}{2}\right] : |S_F(f)| \geq \frac{p}{100} |S_F|_{\max} \right\}$$

$$F_{\text{mark}} = \{f_1, f_2, \ldots, f_n\} = \left\{ f \in F_{\text{rel}} : \left| \frac{S_F(f) - S'_F(f)}{S_F(f)} \right| < \varepsilon \right\}$$

- Parameter $p \in$ [0,100] (percentage)
- Parameter $\varepsilon \in$ [0,1] (relative error)

Watermarking scheme

a) Mark embedding

- $\hat{S}_F$ is the marked signal (frequency domain).

$$\hat{S}_F(f_{\text{mark}}) = \begin{cases} S_F(f_{\text{mark}}) \cdot 10^{d/20} & \text{to embed '1',} \\ S_F(f_{\text{mark}}) \cdot 10^{-d/20} & \text{to embed '0'.} \end{cases}$$

$$\hat{S}_F(f) = \begin{cases} S_F(f), & \text{if } f \notin F_{\text{mark}}, \\ S_F(f) \pm d \text{ dB}, & \text{if } f \in F_{\text{mark}}. \end{cases}$$

- Parameter d (magnitude modification)
- $\text{IFFT}(\hat{S}_F) \rightarrow \hat{S}$ (PCM)

Watermarking scheme

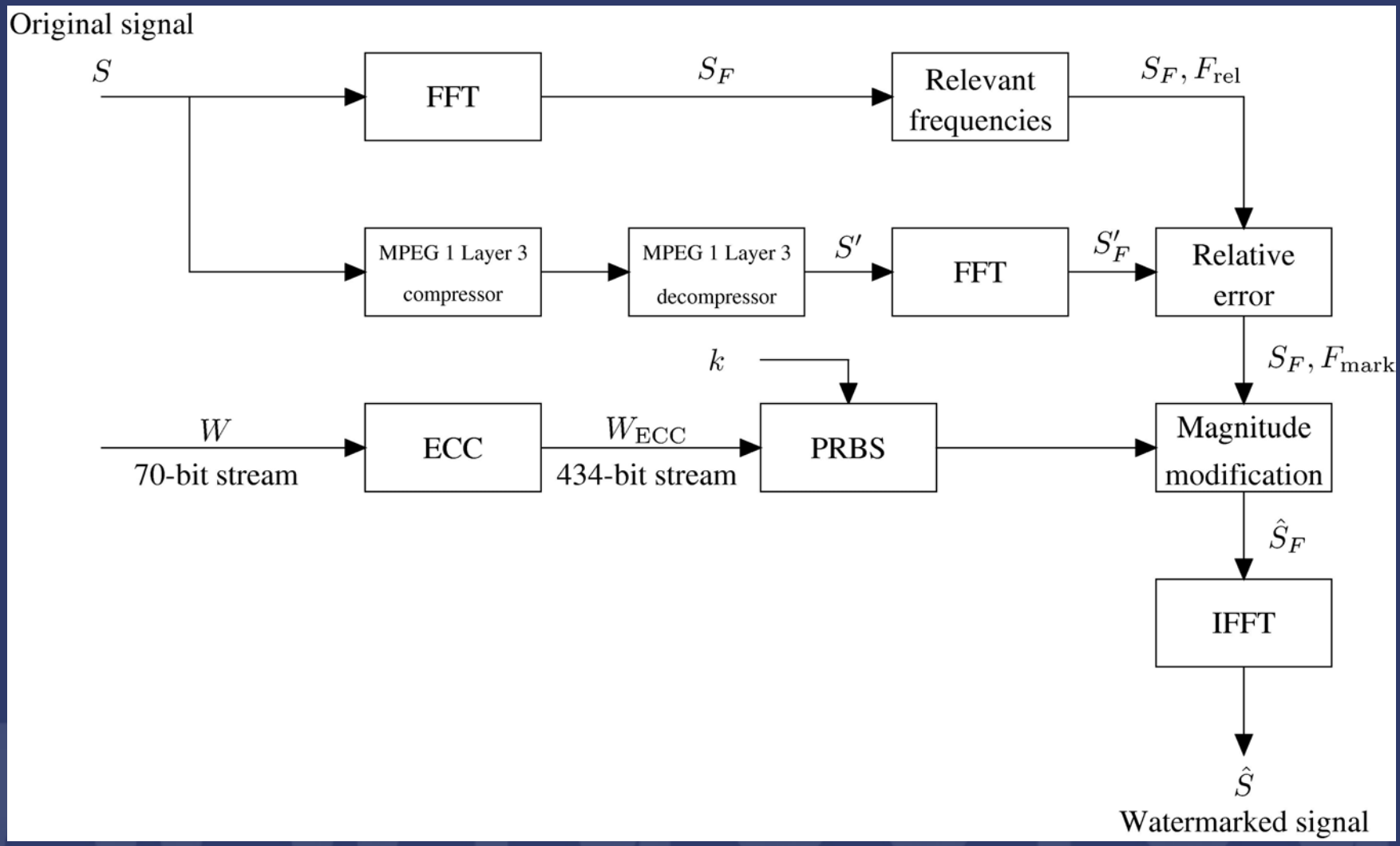
a) Mark embedding

- W (mark) $\rightarrow$ a 70-bit stream
- $\mathrm{ECC}(W) \rightarrow W_{\mathrm{ECC}}$ using a dual Hamming ECC: $\mathrm{DH}(31,5)$
- A Pseudo-Random Binary Signal generated with key k is added to W_{ECC} prior to embedding the mark.

Watermarking scheme

a) Mark embedding

$$\text{Embed}\,(S, W, \text{parameters} = \{R, p, \varepsilon, d, k\}) \rightarrow \left\{\hat{S}, F_{\text{mark}}\right\}$$



Watermarking scheme

b) Mark reconstruction

- T (PCM) is the signal to be tested, a possibly attacked version of $\hat{S}$.
- $\mathrm{FFT}(T) \rightarrow T_F$
- Least squares:

$$\min_{\lambda} \sum_{f \in F_{\mathrm{mark}}} \left(\left| \hat{S}_F(f) \right| - \lambda \left| T_F(f) \right| \right)^2 .$$

Watermarking scheme

b) Mark reconstruction

- In vector form:

$$\boldsymbol{s} = \left[\begin{array}{cccc} |S_F(f_1)| & |S_F(f_2)| & \dots & |S_F(f_n)| \end{array} \right]^{\mathrm{T}},$$

$$\hat{\boldsymbol{s}} = \left[\begin{array}{cccc} \left|\hat{S}_F(f_1)\right| & \left|\hat{S}_F(f_2)\right| & \dots & \left|\hat{S}_F(f_n)\right| \end{array} \right]^{\mathrm{T}},$$

$$\boldsymbol{t} = \left[\begin{array}{cccc} |T_F(f_1)| & |T_F(f_2)| & \dots & |T_F(f_n)| \end{array} \right]^{\mathrm{T}},$$

$$\min_{\lambda} \left(\hat{\boldsymbol{s}} - \lambda \boldsymbol{t}\right)^{\mathrm{T}} \left(\hat{\boldsymbol{s}} - \lambda \boldsymbol{t}\right),$$

$$\lambda = \frac{\hat{\boldsymbol{s}}^{\mathrm{T}} \boldsymbol{t}}{\boldsymbol{t}^{\mathrm{T}} \boldsymbol{t}}.$$

Watermarking scheme

b) Mark reconstruction

- Bit recovery:

$$r_i = \frac{\lambda t_i}{s_i}$$

$$r_i \in \left[10^{\frac{d}{20}}\left(\frac{100-q}{100}\right), 10^{\frac{d}{20}}\left(\frac{100+q}{100}\right)\right] \Rightarrow \hat{b}_i := \text{‘1’},$$

$$\frac{1}{r_i} \in \left[10^{\frac{d}{20}}\left(\frac{100-q}{100}\right), 10^{\frac{d}{20}}\left(\frac{100+q}{100}\right)\right] \Rightarrow \hat{b}_i := \text{‘0’}.$$

- Or $\hat{b}_i := *$ (unidentified) otherwise
- Parameter $q \in$ [0,100] (percentage)

Watermarking scheme

b) Mark reconstruction

- The PRBS is subtracted from $\hat{b}$ giving b.
- Now a voting scheme is applied to identify the mark:

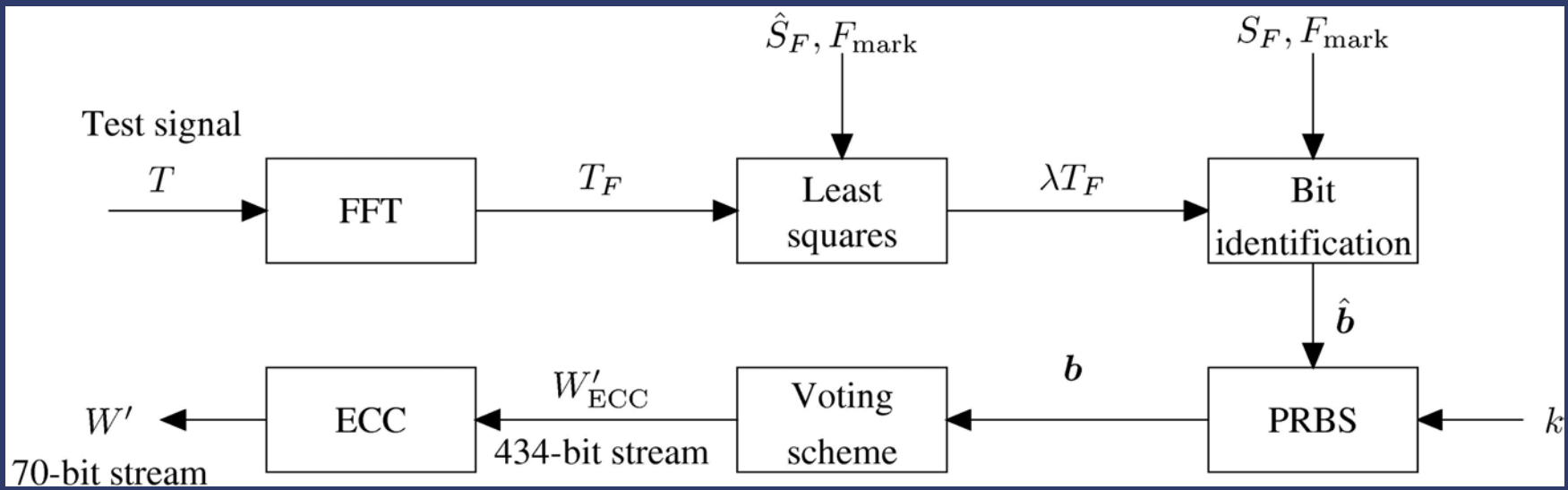
$$\text{bit} := \begin{cases} \text{`*'} & \text{if } n_* > 2\,|n_1 - n_0|\,, \\ \text{`1'} & \text{if } n_* \leq 2\,|n_1 - n_0| \text{ and } n_1 > n_0, \\ \text{`0'} & \text{if } n_* \leq 2\,|n_1 - n_0| \text{ and } n_0 > n_1, \end{cases}$$

- The error correcting code is applied to W'_{ECC} and the identified mark W' is obtained.

Watermarking scheme

b) Mark reconstruction

$$\text{Reconstruct}\left(T, S, \hat{S}, F_{\text{mark}}, \text{parameters} = \{q, d, k\}\right) \rightarrow \{W', \boldsymbol{b}\}$$



Outline

1. Introduction
2. Audio watermarking scheme
   a) Mark embedding
   b) Mark reconstruction
3. **Performance evaluation**
   a) Imperceptibility
   b) Capacity
   c) Robustness
4. Conclusions and future research

Performance evaluation

- ***Imperceptibility***: the extent to which the embedding process leaves undamaged the perceptual quality of the marked object.
- ***Capacity***: the amount of information which can be embedded and recovered.
- ***Robustness***: the resistance to accidental removal of the embedded bits.
- Parameters: R = 128 Kbps, p = 2, ε = 0.05, d = 1 dB, q = 10.
- 10 test files of the SQAM (first 10 seconds).

Performance evaluation

a) Imperceptibility

- Signal to noise ratio (SNR):

$$\mathrm{SNR} = \frac{\sum_{i=1}^{N} S_i^2}{\sum_{i=1}^{N} \left(S_i - \hat{S}_i \right)^2}$$

- Average SNR in blocks of 4 ms.

Performance evaluation

b) Capacity

- 434 marked bits for each 70-bit mark.
- Capacity = 70/434 the number of marked bits.

Performance evaluation

Imperceptibility & Capacity

SQAM file	Marked bits	Capacity (bits)	SNR (dB)	ASNR (dB)
violoncello	4477	722	18.92	20.91
trumpet	3829	617	18.83	19.84
horn	1573	253	18.96	21.10
glockenspiel	1317	212	19.00	20.46
harpsichord	3836	618	19.96	21.48
soprano	5042	813	19.47	21.59
bass	15763	2542	19.02	20.08
quartet	13548	2185	19.22	20.36
female speech	10677	1722	19.57	21.84
male speech	9359	1509	19.44	21.49

Performance evaluation

c) Robustness

- 38 attacks in the Stirmark benchmark for audio:

Name	Number	Name	Number	Name	Number
AddBrumm	1—11	AddDynnnoise	12	Addnoise	13—17
AddSinus	18	Amplify	19	Compressor	20
Echo	21	Exchange	22	FFT_HLPass	23
FFT_Invert	24	FFT_RealReverse	25	FFT_Stat1	26
FFT_Test	27	FlippSample	28	Invert	29
LSBZero	30	Normalise	31	RC-HighPass	32
RC-LowPass	33	Smooth	34	Smooth2	35
Stat1	36	Stat2	37	ZeroCross	38

Performance evaluation

c) Robustness

- Correlation measures the resistance to a given attack:

$$\beta_i = \begin{cases} 1, & \text{if } W_i = W_i' \\ -1, & \text{if } W_i \neq W_i' \end{cases}$$

$$\text{Correlation} = \frac{1}{|W|} \sum_{i=1}^{|W|} \beta_i .$$

- Correlation = 1 means perfect survival of the mark, and Correlation = -1 means total erasure of the mark.
- We consider survival **only when Correlation = 1**.

Performance evaluation

c) Robustness

Number	Survival ratio	Number	Survival ratio	Number	Survival ratio
1	10/10	2	10/10	3	10/10
4	10/10	5	10/10	6	10/10
7	10/10	8	10/10	9	10/10
10	10/10	11	10/10	12	6/10
13	10/10	14	10/10	15	10/10
16	10/10	17	8/10	18	6/10
19	10/10	20	10/10	21	0/10
22	10/10	23	3/10	24	10/10
25	10/10	26	0/10	27	0/10
28	0/10	29	10/10	30	10/10
31	10/10	32	1/10	33	10/10
34	9/10	35	9/10	36	10/10
37	10/10	38	1/10		

Performance evaluation

c) Robustness

- Non-survived attacks: 21 (Exchange), 23 (FFT_HLPass), 26 (FFT_Stat1), 27 (FFT_Test), 28 (Flipp_sample), 32 (RC_Highpass) and 38 (ZeroCross).
- Mp3 compression attacks:

Bit rate (Kbps)	320	256	224	192	160	128	112
Compression ratio	4.41:1	5.51:1	6.30:1	7.35:1	8.82:1	11.03:1	12.60:1
Correlation	1	1	1	1	1	1	1
Bit rate (Kbps)	96	80	64	56	48	40	32
Compression ratio	14.70:1	17.64:1	22.05:1	25.20:1	29.30:1	35.28:1	44.10:1
Correlation	1	0.97	0.97	0.94	0.83	0.80	0.49

Outline

1. Introduction
2. Audio watermarking scheme
   a) Mark embedding
   b) Mark reconstruction
3. Performance evaluation
   a) Imperceptibility
   b) Capacity
   c) Robustness
4. **Conclusions and future research**

Conclusions and further research

Conclusions

- A watermarking scheme for audio based on lossy compression has been suggested.
- The method consists of choosing the frequencies for which the spectrum is not modified after compression and decompression.
- Performance has been evaluated for 10 files of the SQAM.
  - **Imperceptibility**: the noise introduced is roughly 0.01 times the power of the original signal.
  - **Capacity**: for 3-minute files the mark can be embedded several hundreds of times.
  - **Robustness**: the scheme has been successfully tested against the Stirmark benchmark and mp3 compression attacks.

Conclusions and further research

Further research

- Tuning guidelines for the parameters must be obtained.
- Some modification is required to treat the stereophonic case.
- Modification is also needed to treat attacks which change the number of samples significantly.
- The use of models of the HAS, at least to measure imperceptibility.
- The possibility of working with 'slices' of samples instead of the whole file should be addressed.

Thank you very much!

Presentation written using: