A HIGH CAPACITY BLIND WATERMARKING FOR TWO-CHANNEL AUDIO SIGNALS BASED ON CDMA-ICA

Mohammed Khalil, Adil El Bahi and Abdellah Adib

LIM@II-FSTM, B.P. 146 Mohammedia 20650, Morocco
medkhalil87@gmail.com, adil.elbahi@yahoo.com, adib@fstm.ac.ma

ABSTRACT

In this paper, we investigate how data embedding capacity and extraction fidelity can be enhanced keeping imperceptibility of the host audio unaltered. To meet this end, we propose a new blind watermarking system for two-channel audio signals which is based on Code Division Multiple Access (CDMA) and Independent Components Analysis (ICA) techniques. At the emitter, CDMA is used to increase the watermark bit rate by inserting multiple parallel data. At the receiver, ICA algorithm is applied to recover the embedded information. In order to demonstrate the effectiveness of the proposed scheme, simulations under various conditions are conducted. The new system provides an embedding rate up to 60Kbits/s, while ensuring a compromise between inaudibility and extraction reliability.

Index Terms— Audio watermarking, High capacity, two-channel audio signals, CDMA, ICA.

1. INTRODUCTION

Inaudibility, robustness and data embedding capacity are the three main constraints of audio watermarking system. Nevertheless, it is not possible to ensure the three constraints simultaneously [1]. Therefore, if a high robustness is required, the watermark bit rate will be low and vice versa, i.e. high bit rate watermarking systems are usually very fragile in the presence of signal distortions.

In the past decade, several audio watermarking techniques have been proposed to deal with the above-mentioned constraints. Spread spectrum (SS) modulations remain the most commonly used methods due to their inherent anti-jamming and interference rejection property [2]. However, their major shortcoming is the low embedding capacity.

In order to increase data embedding capacity, watermarking-based CDMA (Code Division Multiple Access) schemes have been employed. While much of the papers have used CDMA in image watermarking, only a limited studies were carried out on audio watermarking[1]. In [3], authors propose a multiple audio watermarking algorithm applying CDMA. Two watermarks are embedded in the discrete wavelet transform coefficients of the audio signal and then they are extracted by using FastICA (Fast Independent Components Analysis) algorithm. In that work, however, no experimental results about the amount of data, that can be embedded under the inaudibility constraint, have been discussed. Moreover, FastICA algorithm has been employed without highlighting the observations from which the two watermarks are extracted. In a previous work [4], we proposed a new audio watermarking-based CDMA system that increased the watermark bit rate by considering the multiple embedded watermarks as different users which share the same channel, i.e. the host audio signal. Recovering two signals (watermark and host audio signals) from only one observation (watermarked signal) led us to use an under-determinated BSS (Blind Source Separation) technique to extract the embedded signal [5].

Since several audio signals, e.g. radio, television, music CDs and MP3s, are available in a two-channel formats, we propose, in this work, using ICA as an extraction tool by considering the two channels as two different observations[2]. In this case, we overcome the under-determined problem encountered in [4] [5], by assuming that the left and the right channels are quite similar. Nevertheless, ICA requires that the number of observations must be at least equal or larger than the number of independent sources [6], which means that we can embed only one watermark signal. This constraint limits the data embedding capacity of the watermarking system. Therefore, to sort this out, we exploit CDMA technique to increase the watermark bit rate while keeping the number of sources equal to two (audio signal and watermark one that contains multiple parallel data differentiated by their spreading codes). Thereafter, the two main factors that will limit the capacity of the embedded information are: the inaudibility constraint [4] and the Multiple Access Interference (MAI) [7].

The rest of the paper is briefly outlined as follows: In section 2, we present the proposed two-channel audio watermarking system. The performances of the proposed watermarking system are evaluated by a number of experiments in section 3. Finally, a conclusion is presented in Section 4.

[1]Audio watermarking is well known to be challenging because of the dynamic and characteristics of such signals in comparison with image signals.

[2]Called also mixtures.

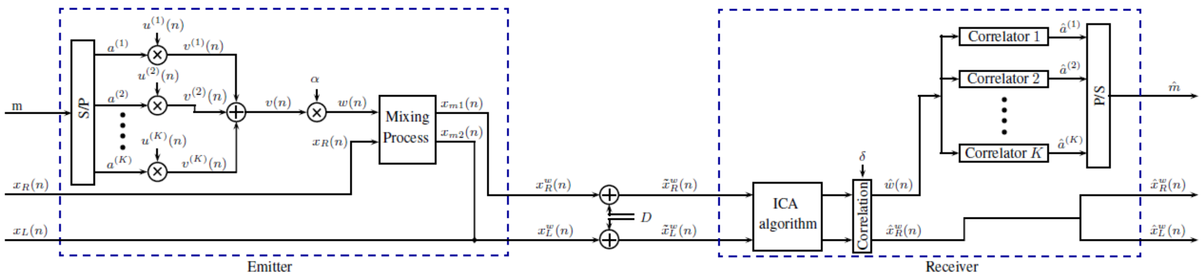

Fig. 1. The proposed CDMA-ICA two-channel audio watermarking scheme.

2. THE PROPOSED TWO-CHANNEL AUDIO WATERMARKING SYSTEM

In this section, we will explain how to increase the capacity of embedded information and how to improve its detection reliability in a two-channel audio watermarking system. The proposed watermarking scheme is illustrated in **Fig** 1. At the emitter, CDMA considers the multiple embedded data as different users that share the same host audio signal [4]. At the receiver, ICA algorithm is applied to estimate the two sources (watermark and audio signals) from the two-channel watermarked signal.

2.1. Emitter

The original binary message m of length N is represented as an independent and identically distributed (iid) binary sequence. The time T_b which used to insert one bit is related to the sampling frequency F_s of the audio signal by:

$$T_b = \frac{N_b}{F_s} \tag{1}$$

where N_b is the discrete bit time. We convert the original binary message m into K parallel streams $a^{(k)} = [a_0^{(k)}, a_1^{(k)}, \ldots, a_{\frac{N}{K}-1}^{(k)}]$. Each symbol $a_i^{(k)}$ of length $N_s = qN_b$ is chosen from a codebook of $M = 2^q$ symbols, where q represents the number of bits per symbol. Therefore, the k^{th} stream $a^{(k)}$ from serial-to-parallel (S/P) conversion is carried out by the spreading sequence $u^{(k)}(n)$ of length L:

$$u^{(k)}(n) = \sum_{l=0}^{L-1} u_l^{(k)} p_{N_c}(n - lN_c) \quad k = \{1, \ldots, K\} \tag{2}$$

where N_c is the chip duration, the rectangular pulse p_{N_c} is equal to 1 for $0 \leq n < N_c$ and zero otherwise and $u_l^{(k)}$ are the chips of the specific spreading sequence $u^{(k)}(n)$. After the spreading step, the signal $v^{(k)}(n)$ corresponding to the k^{th} data stream $a^{(k)}$ will be given by:

$$v^{(k)}(n) = a^{(k)} \sum_{l=0}^{L-1} u_l^{(k)} p_{N_c}(n - lN_c), \quad 0 \leq n < N_s \tag{3}$$

for one data symbol duration $N_s = LN_c$. Hence, the modulated signal results in:

$$v(n) = \sum_{k=1}^{K} v^{(k)}(n) \tag{4}$$

The watermark signal $w(n)$ is, then, constructed by introducing a scale factor α which controls the inaudibility constraint. The resulting watermark bit rate will be then:

$$R = \frac{nKF_s}{N_s} \tag{5}$$

Now, giving the watermark signal $w(n)$, the embedding process is performed by using a mixing matrix $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. Considering one of the two-channel audio signal $x_j(n)$ ($j = R$ or L for right and left channels) and the watermark signal $w(n)$ as two independent signals of the same length $\frac{N}{K}N_s$, the matrix A is used to get two linear observations $x_{m1}(n)$ and $x_{m2}(n)$ as follows:

$$\begin{aligned} \begin{bmatrix} x_{m1}(n) \\ x_{m2}(n) \end{bmatrix} &= \begin{bmatrix} a_{11}x_j(n) + a_{12}w(n) \\ a_{21}x_j(n) + a_{22}w(n) \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_j(n) \\ w(n) \end{bmatrix} \end{aligned} \tag{6}$$

Thereafter, the two-channel watermarked signal $x_R^w(n)$ and $x_L^w(n)$ are obtained by taking:

$$\begin{aligned} x_R^w(n) &= x_{m1}(n) \\ x_L^w(n) &= x_{m2}(n) \end{aligned} \tag{7}$$

Inasmuch as the two-channel audio signal are quite similar[3], the inaudibility constraint will be ensured even if one of the watermarked channels is obtained by simply replacing it with one of the mixtures (produced at the mixing process output). Furthermore, the matrix coefficients a_{ij} are chosen in such a way to create a slight difference between $x_{m1}(n)$ and $x_{m2}(n)$.

[3] Contrary to the stereophonic audio where there is a different amplitude levels and a phase delays between the two channels.

2.2. Receiver

In order to extract the watermark signal $w(n)$, we propose using ICA algorithm. Considered as the most used method in BSS, ICA aims to recover some original signals called *'source signals'* from only their linear instantaneous mixtures[4]. This approach is based on the assumption that source signals are statistically independent [6], which means that the joint pdf[5] $p(y_1, y_2, \ldots, y_n)$ of random variables $y_1, y_2, \ldots, y_n$ can be factorized as follows:

$$p(y_1, y_2, \ldots, y_n) = p_1(y_1)p_2(y_2)\ldots p_n(y_n) \tag{8}$$

where $p_i(y_i)$ is the marginal pdf of y_i. In our case, this assumption is clearly obtained since the knowledge of $x_j(n)$ does not give any information about $w(n)$ and vice versa. Thus, the source recovering (or the watermark extraction in our context) can be achieved by estimating a de-mixing matrix $\mathbf{B}$ that linearly transforms the observations signals [6] and yields to sources that are as independent as possible [6]:

$$\begin{bmatrix} \hat{x}_j(n) \\ \hat{w}(n) \end{bmatrix} = \mathbf{B}. \begin{bmatrix} \tilde{x}^w_R(n) \\ \tilde{x}^w_L(n) \end{bmatrix} \tag{9}$$

where $\hat{x}_j(n)$ for $j = R \, or \, L$ is an estimation of the audio channel used during the mixing process and $\hat{w}(n)$ is an estimation of the watermark signal. Now, to recover the original binary message m, we need to distinguish between $\hat{x}_j(n)$ and $\hat{w}(n)$ because the ICA model can not determine the order of the separated sources [6]. The two estimated source are then correlated with a spreading waveform $u^{(k)}(n)$. The estimated source that will give the highest correlation with $u^{(k)}(n)$ will be the estimated watermark $\hat{w}(n)$. Now, giving $\hat{w}(n)$, the demodulation step will be conventionally carried out by K correlation detectors, and it is assumed that we have a copy of the spreading signals $u^{(k)}(n)$ at the receiver.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

3.1. Experimental protocol

In this section, some experimental results demonstrating the performances of the proposed system will be presented. In these experiments, the well-known SQAM[7](Sound Quality Assessment Material) files are selected as test samples to analyze the performances of the new watermarking system. A total of 29 host audio signals with different styles and sampling at 44.1 kHz are used. All these audio clips are two-channel format belonging to 7 different audio groups as illustrated in **Table** 1. In all simulations, new audio signals are constructed from those belonging to SQAM database in such a way that the left and the right channels are quite similar (i.e. $x_L(n) \simeq x_R(n)$).

Table 1. Test sequences of SQAM database

Host signals	Genres	Mono(M)/Stereo(S)
$X_{01} \rightarrow X_{11}$	Single instruments	S
$X_{12} \rightarrow X_{15}$	Vocal	S
$X_{16} \rightarrow X_{21}$	Speech	M [8]
$X_{22} \rightarrow X_{23}$	Solo instruments	S
$X_{24} \rightarrow X_{25}$	Vocal and Orchestra	S
$X_{26} \rightarrow X_{27}$	Orchestra	S
$X_{28} \rightarrow X_{29}$	Pop Music	S

The inaudibility of the hiding data is the first constraint imposed to a watermarking system. Hence, the Perceptual Of Quality Of Audio Perceived (PEAQ) algorithm will be used to measure the watermark transparency [8]. This algorithm provides an Objective Difference Grade (ODG) value between 0 and -4 revealing the audio quality. Higher ODG values show better quality of the watermarked signal.

Once the watermark transparency is evaluated, the Bit Error Rate (BER) will be used as a tool to measure the extraction fidelity. At the beginning, we assume that the watermarking system is not under any attacks to determine the maximum of information that we can embedded, then we evaluate its robustness against common signal processing attacks described as follows:

- **Additive White Gaussian Noise**: Random noise is added to the watermarked signals, where the ratio of the watermarked signal to noise ration (WSNR) is 10dB and 20dB.
- **Uniform quantization** with a resolution of 13 and 14 bits.
- **MPEG compression**: MPEG 1 Layer III (MP3) compression is performed on the watermarked audio signals at the bit rates 64Kbits/s, 96Kbits/s.

3.2. Simulations

First, we mention that all the results shown in figures below have been computed as a mean over 29 audio signals. In order to investigate the effect of the watermark power on the inaudibility constraint, we fix the number of parallel streams $K = 1$ and then we compute the ODG values according to the scale factor α. As shown in **Fig.** 2, the audio quality is improved when the value of α is decreased. Moreover, we remark that the ODG values of the two channels are not similar due to the faint difference between the mixing matrix coefficients ($a_{11} = 0.9$, $a_{12} = 0.1$ and $a_{21} = 0.95$, $a_{22} = 0.95$)

[4]Which is the case in this work

[5]probability density function.

[6]$\tilde{x}^w_R(n)$ and $\tilde{x}^w_L(n)$ are the two-channel watermarked signal after being subjected to an attack distortion D.

[7]Available at http://soundexpert.org/sound-samples

[8]The left and the right channels are quite similar.

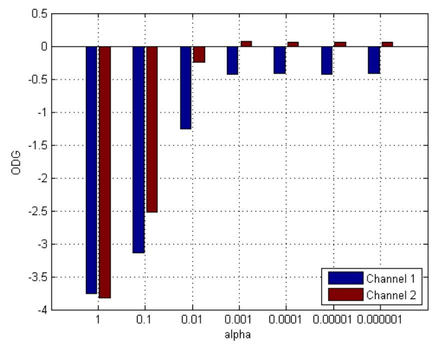

Fig. 2. ODG values according to the scale factor α.

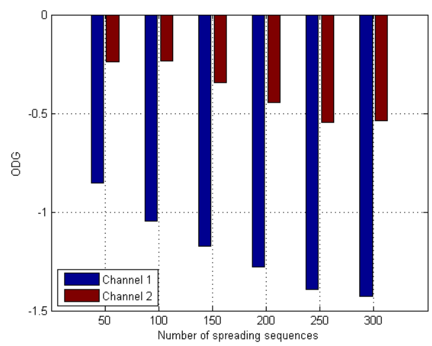

Fig. 3. ODG values according to the number of parallel streams K.

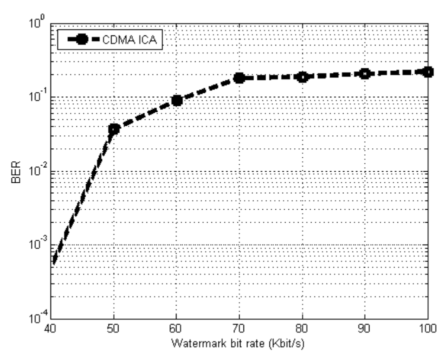

Fig. 4. BER values according to the watermark bit rate.

used to form the two-channel watermarked signal $x_R^w(n)$ and $x_L^w(n)$. In the remaining, we choose $\alpha = 10^{-3}$.

Now, according to equation (5), data embedding capacity can be increased by varying three parameters: the number of bits per symbol n, the number of data streams K and the discrete symbol time N_s. For a given value of n (that we fix in our case at 1), one will try to put K as maximal as possible and N_s as minimal as possible. Nevertheless, these parameters, need to be adjusted to satisfy a trade-off between

Table 2. Additive White Gaussian Noise

WSNR(dB)	Host signals	Average BER	
		10Kbit/s	20Kbit/s
20dB	$X_{01} \rightarrow X_{11}$	0.0029	0.0195
	$X_{12} \rightarrow X_{15}$	0	0.0005
	$X_{16} \rightarrow X_{21}$	0.0099	0.0496
	$X_{22} \rightarrow X_{23}$	0	0.0077
	$X_{24} \rightarrow X_{25}$	0.0156	0.0512
	$X_{26} \rightarrow X_{27}$	0.0001	0.0074
	$X_{28} \rightarrow X_{29}$	0.0001	0.0029
10dB	$X_{01} \rightarrow X_{11}$	0.0981	0.1512
	$X_{12} \rightarrow X_{15}$	0.0044	0.0309
	$X_{16} \rightarrow X_{21}$	0.2087	0.2802
	$X_{22} \rightarrow X_{23}$	0.0256	0.0665
	$X_{24} \rightarrow X_{25}$	0.1356	0.1677
	$X_{26} \rightarrow X_{27}$	0.1016	0.1887
	$X_{28} \rightarrow X_{29}$	0.0478	0.0926

Table 3. Uniform quantization

Resolution	Host signals	Average BER	
		10Kbit/s	20Kbit/s
14	$X_{01} \rightarrow X_{11}$	0.1110	0.1464
	$X_{12} \rightarrow X_{15}$	0.0484	0.1082
	$X_{16} \rightarrow X_{21}$	0.0524	0.1985
	$X_{22} \rightarrow X_{23}$	0.0736	0.1148
	$X_{24} \rightarrow X_{25}$	0.0458	0.0749
	$X_{26} \rightarrow X_{27}$	0.1415	0.1615
	$X_{28} \rightarrow X_{29}$	0.0454	0.2087
13	$X_{01} \rightarrow X_{11}$	0.2142	0.2375
	$X_{12} \rightarrow X_{15}$	0.1955	0.2306
	$X_{16} \rightarrow X_{21}$	0.2265	0.2461
	$X_{22} \rightarrow X_{23}$	0.1863	0.2256
	$X_{24} \rightarrow X_{25}$	0.1960	0.2323
	$X_{26} \rightarrow X_{27}$	0.2375	0.2298
	$X_{28} \rightarrow X_{29}$	0.1873	0.2309

the inaudibility constraint and the extraction fidelity. In **Fig.** 3, we present the evolution of the ODG values according to the parameter K. As it can be seen, by increasing the K value the audio quality of the host signal degrades which is explained by the fact that augmenting K is equivalent to increase the modulation signal power. Furthermore, the parameter K has an impact on the extraction fidelity since it is the source of the MAI between the parallel streams $a^{(k)}$. Hence, we fix it at 100 and then we increase much more the embedding capacity by decreasing N_s.

Even if N_s does not affect the audio quality of the host signal, it has a lower limit that must not exceed to ensure a hight extraction fidelity. By decreasing N_s, the Inter Symbol Interference (ISI) caused between $a_i^{(k)}$ will increase and consequently the watermark detection becomes more difficult. As presented in **Fig.** 4, the BER value increases when

Table 4. MPEG compression

Bit rate (Kbits/s)	Host signals	Average BER	
		10Kbit/s	20Kbit/s
96	$X_{01} \to X_{11}$	0.1979	0.2290
	$X_{12} \to X_{15}$	0.1913	0.2275
	$X_{16} \to X_{21}$	0.2308	0.2458
	$X_{22} \to X_{23}$	0.2090	0.2324
	$X_{24} \to X_{25}$	0.2151	0.2512
	$X_{26} \to X_{27}$	0.2431	0.2590
	$X_{28} \to X_{29}$	0.1826	0.2268
64	$X_{01} \to X_{11}$	0.2550	0.2801
	$X_{12} \to X_{15}$	0.2496	0.2827
	$X_{16} \to X_{21}$	0.2773	0.2851
	$X_{22} \to X_{23}$	0.2530	0.2844
	$X_{24} \to X_{25}$	0.2567	0.2770
	$X_{26} \to X_{27}$	0.2828	0.3115
	$X_{28} \to X_{29}$	0.2454	0.2879

N_s decreases until attaining a value equal to 10^{-1} for a bit rate R$>$ 60Kbits/s.

Note that the spreading codes used in simulations are walsh sequences that consist of a set of orthogonal functions. This choice stems from a previous comparative study already made between various spreading codes, that has shown the superiority of walsh codes in terms of ensuring the best compromise between inaudibility and detection fidelity [9].

It is well known that the host audio signal has an impact on the watermarking system performances. For this purpose, the robustness of the CDMA-ICA is evaluated by calculating the mean values for each genre of different signals as depicted in **Table** 1. Since the data embedding capacity is in complete conflict with the robustness, a compromise should be made. Hence, we put $R = 10$Kbits/s and $R = 20$Kbits/s.

In **Table** 2, we evaluate the proposed system against various amount of additive noise. As expected, we can clearly see that the proposed system gives very low BERs, under noise addition with 20dB, but the performances degrades significantly when the WSNR attains 10dB except for vocal, solo instruments and pop music. Many reasons influence the performances of different signal groups: energy's distribution over time, frequency distribution and bandwidth. In **Table** 3, we remark that the proposed system is generally vulnerable against quantization attacks especially when $R = 20$Kbits/s. However, when the watermark is embedded at $R = 10$Kbits/s, we see that the embedded information can survive in many audio signals like vocal, speech, solo instruments and pop music. Finally, we can see from **Table** 4, that the proposed system is specifically vulnerable to MPEG compression attacks for all audio signal groups.

4. CONCLUSION

In this paper, we proposed a new CDMA-ICA watermarking system for increasing the amount of embedded information and improving its extraction fidelity. The CDMA technique has offered the potential of a high watermark bit rate attaining 60Kbits/s. Extraction of the embedded data has also been enhanced, when the watermarking system is not under any attacks, by introducing the ICA algorithm. The sole limitation of the proposed approach is its fragility against attacks. For this reason, the next stage of our work will be the improvement of the robustness constraint.

5. REFERENCES

[1] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," in *IEEE Trans on Image processing*, 1997, vol. 6, pp. 1673–1687.

[2] Henrique S. Malvar and Darko Kirovski, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Transactions on signal processing*, vol. 51, no. 4, April, 2003.

[3] Y. Chen and J. Chen, "A new multiple audio watermarking algorithm applying DS-CDMA," *International Conference on Machine Learning and Cybernetics*, pp. 2205 – 2210, 12-15 July 2009.

[4] M. Khalil, N. El Hamdouni, and A. Adib, "Increasing information capacity and improving detection reliability in audio watermarking system," in *accepted in ISCCSP'12, Roma, Italy*, 2012.

[5] N. El Hamdouni, A. Adib, S. Larbi, and M. Turki, "A blind digital audio watermarking scheme based on EMD and UISA techniques," *Journal of Multimedia Tools and Applications, Springer*, 2011.

[6] Pierre Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[7] Raymond L. Pickholtz, Donald L. Schilling, Laurence, and B. Milstein, "Spread spectrum for mobile communications," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 2, May 1991.

[8] "Recommendation B.S. 1387 : Method for objective measurements of perceived audio quality," *Int. Telecommunication Union, Geneva*, 2001.

[9] M. Khalil and A. Adib, "Impact of spreading codes on cdma based audio watermarking system," *accepted in International Conference on Telecommunications*, Morocco,2013.