Robust audio watermarking based on multi-carrier modulation

Mohammed Khalil* and Abdellah Adib

LIM@II-FSTM,
B.P. 146 Mohammedia 20650, Morocco
Email: medkhalil87@gmail.com
Email: adib@fstm.ac.ma
*Corresponding author

Abstract: The great challenge, in blind audio watermarking, is to recover the transparent embedded information with minimum errors even in presence of disturbances. Some signal disturbances like MPEG compression and filtering can be modelled as fading-like distortions. Besides the inaudibility constraint, the problem will be viewed as the conception of a particular communication system, where our purpose is to reduce host interference noise (HIN) and inter-symbol interference (ISI) on one hand and to deal with fading perturbations on the other hand. In this paper, we propose adapting multi-carrier (MC) modulation, widely used in wireless communications, to the watermarking context. Our main motivation is to exploit advantages of MC modulations in order to outperform spread spectrum (SS) audio watermarking schemes. The new MC-watermarking system allows us to improve detection reliability and to increase robustness of the embedded watermark. Moreover, the inaudibility constraint will be ensured by adapting the MC signal parameters to the masking characteristics of the human auditory system (HAS).

Keywords: audio watermarking; multi-carrier; fading; OFDM; compression; guard interval; robustness.


Biographical notes: Mohammed Khalil received his PhD in Informatics from Department of Informatics, Faculty of Sciences and Techniques, Mohammedia, Morocco in 2014. He has published four journal papers, 12 communications in international conferences (including EUSIPCO, ICT, ICISP and ISCCSP) and five communications in national conferences. His research interests include digital audio watermarking and digital communication.

Abdellah Adib received his Doctorat de 3rd Cycle and the Doctorat d’Etat-es-Sciences in Statistical Signal Processing from the Mohammed V-Agdal University, Rabat, Morocco, in 1996 and 2004, respectively. From 1986 until 2008, he was with the Institut Scientifique, Mohammed V Agdal University, Rabat, Morocco, as an Associate Professor. Since September 2008, he is a Professor in the Department of Informatics, Hassan II University, Casablanca, Morocco. His research interests are in digital communications and statistical signal processing, with emphasis on (blind) array processing.
1 Introduction

Audio watermarking is the process by which a signature (carrying data stream) called a watermark is embedded into a host audio signal imposing imperceptible loss of audio quality (Boney et al., 1996). Except the inaudibility constraint, blind watermarking systems share the same performance criteria as in digital communication. Their ultimate goal is to successfully transmit the maximum of data while recovering it with minimum errors even in the presence of signal distortions (Cox et al., 1999).

Currently, enhancing the robustness and the watermark bit rate is one of the most desired aspects of each watermarking scheme. However, analysis shows that the watermark bit is in complete conflict with the robustness (Cvejic and Seppänen, 2007). The two most important approaches proposed to deal with this issue are, spread spectrum (SS) methods (Kirovski and Malvar, 2002; Sedghi and Suzuki, 2008) that ensure high robustness and quantisation index modulation (QIM)-based schemes (Chen and Wornell, 1999; Vivekananda et al., 2011) that permit a high transmission rate. In our study, we leave behind us QIM methods and focus on SS-based watermarking.

Several watermarking schemes based on SS methods have been proposed in the last decade (Baras et al., 2006; Cvejic et al., 2003; Khalil and Adib, 2014a; Kirovski and Malvar, 2002; Larbi et al., 2004). This is due to the anti-jamming properties of SS methods that make them robust to narrow-band interference (Haykin, 2001), and also to the low-energy of SS signal that make it transparent (Baras et al., 2006; Khalil and Adib, 2014a). However, SS methods suffer from some drawbacks such as the host interference noise (HIN) that limit the correct detection of hidden information (Sullivan and Moulin, 2003). Furthermore, some effects of watermark distortions including filtering and MPEG compression are not modelled as narrow-band interference, but rather as fading-like distortions (Kundur and Hatzinakos, 2001). In this case, SS methods are not effective and make the watermark fragile against this type of perturbations.

While several SS image watermarking algorithms have been proposed in the literature to deal with fading-like distortions, few works have investigated the robustness of SS audio watermarking systems against this type of perturbation. In Seok et al. (2002), authors present an audio watermarking scheme based on a whitening procedure for linear prediction that improve the extraction of the hidden data. In Cvejic et al. (2003), turbo codes are included in watermarking because they have a large coding gain and good properties in the fading channels. Then, the robustness of the watermark channel is increased even after severe MPEG compression and filtering attacks. In Larbi et al. (2004), authors include a Wiener filter in the audio watermarking in order to improve the detection reliability and increase the robustness against MPEG compression. Another model is presented in Baras et al. (2006), which is based an informed embedding strategy that ensures the local inaudibility constraint of the embedded information and maximises the robustness of the transmission to channel perturbations. In Khalil et al. (2012), the authors proposed an efficient blind audio watermarking detection procedure using under-determined independent subspace analysis. This technique makes the watermark robust to various kind of disturbances.

In this paper, we propose a new multi-carrier (MC) watermarking system that improves detection robustness of standard SS watermarking algorithms. We consider the use of wireless communication tool sets to improve the performance of watermarking system. The MC method is implemented by a specific control of different modulated
signal parameters. The amplitude of the modulated signal is controlled by the masking threshold to ensure the perceptual transparency of the embedded watermark. Then, the hidden data is transmitted in the phase of the modulated signal. Finally, the robustness against fading like distortions will be achieved by choosing adequate carrier frequencies. Note that a fingerprinting system based on MC code-division multi-access technique was proposed in Cha and Kuo (2009). This system is robust against time-varying collusion attacks in protecting continuous media. Nevertheless, an analysis about advantages of MC methods over SS ones has not been significantly studied in the above works. Furthermore, the inaudibility constraint and the robustness of the system against MPEG compression are not included. Another OFDM-based audio watermarking system dedicated to electronic radiotelephone identification was proposed in Shishkin (2010). This system presents an approach different from the proposed one. Indeed, this scheme is based on quantisation step adaptation in the transmitter and receiver.

In summary, the novelty of the proposed approach compared to other ones of state of arts consists of:

1. an analysis about advantages of MC methods over SS ones
2. a new inaudibility control procedure to control the watermark transparency
3. an adequate choice of carrier frequencies that maximises system robustness against fading distortions.

The rest of this paper is briefly outlined as follows: in Section 2, we present the SS watermarking principles and its limitations. In Section 3, we present the MC audio watermarking approach. Performance of the new watermarking system is evaluated by a number of experiments in Section 4. Finally, conclusion and some perspectives end this paper.

2 SS audio watermarking limitations

In blind audio watermarking schemes, the transparent embedded watermark suffers from some effects, which can be characterised as follows: inter-symbol interference (ISI) caused by filters, HIN caused by the host signal and fading caused by some disturbances like MPEG compression. In this section, we first overview the principles of SS-based watermarking system (Baras et al., 2006; Larbi et al., 2004). Then, we discuss the sources of errors (ISI, HIN and fading) in watermarking and we present the limitations of SS approaches.

2.1 SS-based watermarking system

The information contained in the watermark is represented as an independent and identically distributed binary sequence \( b_i = \{0, 1\} \). We group the sequence \( b_i \) to form a symbol sequence \( a_k \) (each symbol \( a_k \) is chosen from an alphabet of \( M = 2^n \) symbols, where \( n \) represents the number of bit per symbol) (Haykin, 2001). The modulation step is characterised by SS waveforms (pseudo-random codes), where each one corresponds to a given symbol \( a_k \) (Larbi et al., 2004). The modulated signal \( v(t) \) will be given then by:
where \( c(t) \) represents the spreading waveforms. To satisfy the inaudibility constraint, the embedded signal power should be controlled by filtering the modulated signal \( v(t) \) by a shaping filter \( H(f) \); so that the resulting watermark signal \( w(t) \) will have the nearest power spectral density (PSD) \( S_w(f) \) to the masking threshold \( S_{mask}(f) \) (Baras et al., 2006; Khalil and Adib, 2014b). Then, the watermark signal \( w(t) \) will be adjusted by a scale factor \( \alpha \) to ensure a compromise between inaudibility and robustness of the embedded information. Finally, the watermarked signal \( y(t) \) is written as:

\[
y(t) = \alpha w(t) + x(t) = \alpha h(t) \ast v(t) + x(t) \quad (2)
\]

where \( h(t) \) is the impulse response of \( H(f) \) and \( \ast \) denotes convolution product.

At the reception, the signal \( y(t) \) is passed through a zero-forcing equaliser \( G(f) = \hat{H}^{-1}(f) \) (Baras et al., 2006; Larbi et al., 2004). The zero-forcing equaliser cancels the ISI by filtering the water marked signal \( y(t) \) with a filter containing a frequency response equal to the inverse of the channel. However, since the audio signal \( x(t) \) used to derive \( H(f) \) is not available at the receiver, an estimate \( \hat{H}(f) \) is computed by psychoacoustic modelling of the watermarked signal \( y(n) \). The estimated signal \( \hat{v}(t) \) is submitted initially to a demodulation step that assesses the similarity between \( \hat{v}(t) \) and the waveforms that are known at the receiver. Finally, a decision step determines the estimated binary sequence \( \hat{b} \).

### 2.2 Robustness against HIN and ISI

HIN is the interference of the watermark \( w(t) \) with the host signal \( x(t) \). This type of interference has a great impact of the detection reliability due to the high power of the audio noise \( x(t) \). In addition to HIN, ISI arises from different filters used at the emitter \( H(f) \) and the receiver \( G(f) \). When the duration of the impulse response of these filters is greater than the symbol time \( T_s \), the symbols \( a_k \) interfere with each other. The detection of the \( i \)th symbol disrupted with the symbols that precede (or follow) it (Haykin, 2001). Indeed, the estimated signal \( \hat{v}(t) \) could be written as:

\[
\hat{v}(t) = \alpha \sum_k a_k p(t-kT_s) + x(t) \ast \hat{h}^{-1}(t) \quad (3)
\]

where \( p(t) = h(t) \ast \hat{h}^{-1}(t) \) and \( \hat{h}^{-1}(t) \) denotes the impulse response of the zero forcing equaliser. The estimated signal \( \hat{v}(t) \) is sampled at time \( t_i = iT_s \) yielding:

\[
\hat{v}(t_i) = \alpha a_i + \alpha \sum_{k=i} a_k p[(i-k)T_s] + x(t_i) \ast \hat{h}^{-1}(t_i) \quad (4)
\]

where the first term represents the \( i \)th transmitted symbol, the second one is the additive noise coming from ISI and the last one still remains the HIN.
Therefore, in the design of watermarking system, the objective is to minimise the
effects of ISI, HIN and thereby recovering data with low error rate. In addition to SS
methods (Baras et al., 2006; Cvejic et al., 2003; Larbi et al., 2004), MC modulations will
constitute a good solution to reduce ISI. Indeed, MC modulation maps the message bits
into a sequence of symbols which will be subsequently converted into many parallel
streams. In this case, we can increase the number of subcarriers; the MC symbol duration
becomes large and the amount of ISI decreases. However, HIN still constitutes a hard
problem, in audio watermarking, to MC modulations.

2.3 Robustness against fading

Previous analytic work in watermarking has considered the effect of distortion as
stationary additive Gaussian noise (Kundur and Hatzinakos, 2001). However, it is clear
that some degradations such as MPEG compression and filtering have the effect of
completely destroying the watermark content in the associated components of the signal.
For example, high-pass filtering will remove the existence of the watermark from
low-frequency coefficients. Therefore, this type of perturbation cannot be modelled as
additive white Gaussian noise (AWGN), due to the unpredictability of signal to noise
ratio (SNR) variation in the watermarked signal. In this work, we assert as in
Kundur and Hatzinakos (2001), that many disturbances including compression and
filtering are more appropriately modelled as fading like distortion. Fading is a term used
in wireless communications to describe the effect of a channel that attenuates the
information-bearing signal amplitude in an unpredictable way (Haykin, 2001). In audio
watermarking, MPEG compression can be described as time and frequency selective
fading. Indeed, distortion due to MPEG-2 Audio Layer III compression varies frame to
frame with time-frequency resolution for each frame varying adaptively. SS approaches
are specifically vulnerable to this type of perturbation; any residual correlation between
the compressed signal and the watermark one can result in unreliable detection. In
addition, they neither take into account spatial non-stationarity of the host signal.
Although SS systems try to exploit spreading to average the fading, they are not designed
to maximise performance. Implementing MC methods in the audio watermarking system
can be a good solution to provide high robustness against fading. An important design
goal for a MC modulation is that the audio channel can be considered as time-invariant
during one MC symbol (Haykin, 2001).

3 MC-based audio watermarking approach

All the arguments given above led us to present MC-based audio watermarking system
illustrated in Figure 1. The watermarking system is the same one presented in 2.1, but
instead of using SS modulation we will employ MC one. The objective is to minimise ISI
and to deal with fading while satisfying the inaudibility constraint.

3.1 MC modulation

The principle of MC modulation is to convert a serial high-rate bits into \( N_c \) low-rate
substreams. Each of \( N_c \) symbols from serial-to-parallel (S/P) conversion is carried out by
different sub-carriers. Since the symbol rate on each sub-carrier is much less than the
Robust audio watermarking based on multi-carrier modulation

initial serial data symbol rate, the ISI effects, significantly decrease. Due to the S/P conversion, the duration of transmission time for $N_c$ symbols will be extended to $N_cT_s$, which forms a single MC symbol with a length of $T_{sym}$ (i.e., $T_{sym} = N_cT_s$). The complex envelope of the modulated MC signal has the form:

$$v_{CE}(t) = \frac{1}{N_c} \sum_{i=0}^{N_c-1} s_i \exp\left(j2\pi f_i t\right) \quad 0 \leq t < T_{sym}$$

(5)

where $s_i$ are the $N_c$ parallel modulated source symbols and the complex exponential signals $\{\exp(j2\pi f_i t)\}_{i=0}^{N_c-1}$ represent different sub-carriers at $f_i$ in the MC signal. The orthogonality of these signals increases the spectral efficiency of the watermarking system, and it will be satisfied if,

$$\frac{1}{T_{sym}} \int_0^{T_{sym}} \exp\left(j2\pi f_i t\right)\exp(-j2\pi f_j t) = 0, \quad k \neq i$$

(6)

Thereby, the $N_c$ sub-carrier frequencies are located at $f_k = \frac{k}{T_{sym}}$ for $k=0,...,N_c-1$ in order to achieve orthogonality between the signals on the $N_c$ sub-carriers. Hence, the resultant signal $v(t)$ is called orthogonal frequency-division multiplexing (OFDM) signal.

Figure 1 The new multi-carrier audio watermarking system

3.2 Guard interval

By extending the symbol duration by $N_c$ times, the ISI effect is greatly reduced. However, fading effect still remains in the MC watermarking scheme. To improve the performance of MC scheme, a guard interval between two consecutive MC symbols should be inserted. Indeed, it is known in telecommunication (Haykin, 2001) that the guard interval eliminates ISI and inter-channel interference (ICI). Furthermore, it allows a great improvement on MC performance over fading and AWGN channels (Haykin, 2001). Therefore, adding guard interval to the MC-watermarking system will increase the performances. Guard interval can be inserted in two different ways. One is the zero padding (ZP) that pads the guard interval with zeros. The other one is the cyclic extension of the MC symbol (for some continuity) with cyclic prefix (CP) or cyclic suffix (CS).CP extends the MC symbol by copying the last samples of the MC symbol into its front. The extended MC symbols have the duration of $T_{sym} = T_{sym} + T_G$ ($T_G$ denotes the length of CP in terms of samples). As the continuity of each subcarrier has been ensured by the CP, its orthogonality with all other sub-carriers is maintained over $T_{sym}$. Note that data rate of the MC symbol is reduced by $T_{sym}/T_{symG} = T_{sym}/(T_{sym} + T_G)$ times due to the guard interval.
3.3 Inaudibility constraint

The most important requirement of the watermarking system is the inaudibility constraint. In Baras et al. (2006) and Larbi et al. (2004), a shaping filter \( H(f) \) is designed by a psychoacoustic auditory model (PAM) in such a way that the modulated signal \( v(t) \) has an unit power,

\[
\sigma_v^2 = 1 \tag{7}
\]

In order to use the shaping filter \( H(f) \) in the proposed scheme, we must adapt the amplitude of the MC signal while satisfying (7). Each sub-carrier component of the MC symbol with the effective duration \( T_{sym} \), can be regarded as a single-tone signal multiplied by a rectangular window of length \( T_{sym} \). The modulated signal \( v(t) \) can be written as the sum of \( N_c \) sinusoidal random phase (Haykin, 2001; Kumarand and Sreenivas, 2007) with constellation points positioned at uniform angular spacing around a circle (Khalil and Adib, 2014c),

\[
v(t, \phi) = \sum_{i=0}^{N_c-1} A \cos(2\pi f t + \phi) \tag{8}
\]

where \( A \) is the amplitude of each single-tone signal and the phase \( \phi \) is a random variable that is uniformly distributed over the interval \([-\pi, \pi]\) with a probability density (Haykin, 2001):

\[
p(\phi) = \frac{1}{2\pi} \text{ with } \phi \in [-\pi, \pi] \tag{9}
\]

According to (8) and (9), we compute the statistical moments of the random variable \( v(t, \phi) \) at a given time \( t_k \),

\[
E\left[v\left(t_k, \phi\right)\right] = E\left[v\left(t_k\right)\right] = \int_{-\infty}^{\infty} p(\phi) \sum_{i=0}^{N_c-1} A \cos(2\pi f t_k + \phi) d\phi = \frac{1}{2\pi} \sum_{i=0}^{N_c-1} A \int_{-\infty}^{\infty} \cos(2\pi f t_k + \phi) d\phi = 0 \tag{10}
\]

and

\[
E\left[v^2\left(t_k, \phi\right)\right] = E\left[v^2\left(t_k\right)\right] = \int_{-\infty}^{\infty} p(\phi) \sum_{i=0}^{N_c-1} A^2 \cos^2(2\pi f t_k + \phi) d\phi = N_c \frac{A^2}{2} \tag{11}
\]

Hence, the modulated signal \( v(t) \) is zero mean with variance \( \sigma_v^2 = N_c \frac{A^2}{2} \). In order to achieve the inaudibility constraint (7), we have to choose \( A = \sqrt{\frac{2}{N_c}} \). Then, the modulated signal \( v(t) \) passed through the shaping filter \( H(f) \) so that the resulting watermark signal \( w(t) \) will have the nearest PSD \( S_w(f) \) to the masking threshold \( S_{mask}(f) \).
Figure 2 shows the PSD of watermark signal $w(t)$ using MC and SS modulations compared to the masking threshold signal in the spectral domain. It illustrates that the MC watermark signal power is considered as the sum of the frequency shifted sinc functions in the frequency domain. The different symbols are transmitted with sub-channels in parallel form.

**Figure 2** Comparison of PSD watermark signal (in the case of MC and SS techniques) with the masking threshold (see online version for colours)

3.4 Extraction process

At the receiving end, we should extract the hiding information $a_k$ from only the watermarked signal $y(t)$ with minimum errors as depicted in Figure 1. The watermarking system is blind because the original signal $x(t)$ is unknown at the detection step. The zero-forcing equaliser $\hat{G}(f) = \hat{H}^{-1}(f)$ is estimated from the masking properties of the watermarked signal $y(t)$. It cancels the ISI by filtering the watermarked signal $y(t)$ with the inverse of the channel (Baras et al., 2006). Now, giving $\hat{v}(t)$, the detection step will be conventionally carried out by $N_c$ correlation detectors. Indeed, the estimated signal $\hat{v}(t)$ is submitted initially to a demodulation step that assesses the similarity between the estimated watermark $\hat{v}(t)$ and elements of the codebook. Finally, a decision step determines the estimated binary message $\hat{a}_k$.

4 Experimental results

4.1 Evaluation criteria

The performance of the proposed watermarking system is assessed according to three criteria: transparency, detection reliability and robustness against disturbances. The
number of bits per symbol \( n = 1 \) and the number of sub-carriers \( N_c = 16 \). The performance of the system depends also on the audio signal type, so a set of 21 audio signals, belonging to SQAM\(^1\) database and sampled at \( F_s = 44.1 \) kHz are used as test signals. They include various styles of audio signals as illustrated in Table 1.

### Table 1 Test sequences of SQAM database

<table>
<thead>
<tr>
<th>Host signals ( X_{i0} \rightarrow X_{i5} )</th>
<th>Genres</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single instruments</td>
<td></td>
<td>Violin, viola, horn, tuba, harp</td>
</tr>
<tr>
<td>Vocal</td>
<td></td>
<td>Soprano, alto, tenor, bass and quartet</td>
</tr>
<tr>
<td>Speech</td>
<td></td>
<td>Female and male English and French speech</td>
</tr>
<tr>
<td>Solo instruments</td>
<td></td>
<td>Trumpet, organ, guitar, violin and piano</td>
</tr>
<tr>
<td>Pop music</td>
<td></td>
<td>ABBA and Eddie Rabbitt</td>
</tr>
</tbody>
</table>

First, the inaudibility of the embedded watermark is evaluated using the perceptual evaluation of audio quality (PEAQ) algorithm which is currently considered among the best measures of the transparency of the watermark (Kondo, 2012). It provides an objective difference grade (ODG) value revealing the quality of the watermarked signal as presented in Table 2. Then, bit error rate (BER) is adopted to compare and evaluate the performance level of detection. The average BER results from 30 Monte-Carlo simulation runs, of detection of random binary messages. Finally, the proposed method is evaluated against the following perturbations are performed:

- Additive white Gaussian noise (AWGN) with SNR \( \epsilon \{ -10, 0, 10 \} \) dB.
- High pass filtering with a cut-off frequency 2 KHz.
- Low pass filtering with a cut-off frequency 8 KHz.
- MPEG/audio layer III compression at bit rates of 64 kbits/s.

### Table 2 Scale perceptual quality of audio perceived

<table>
<thead>
<tr>
<th>ODG</th>
<th>Assessment of transparency of the watermark</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Imperceptible</td>
</tr>
<tr>
<td>−1</td>
<td>Perceptible but not annoying</td>
</tr>
<tr>
<td>−2</td>
<td>Slightly annoying</td>
</tr>
<tr>
<td>−3</td>
<td>Annoying</td>
</tr>
<tr>
<td>−4</td>
<td>Very annoying</td>
</tr>
</tbody>
</table>

### 4.2 Test results

Figure 3 presents the obtained ODGs for SS and MC modulations according to the scale factor \( \alpha \). We have changed the value of \( \alpha \) in order to investigate the effect of watermark power on the inaudibility. By increasing \( \alpha \), lower quality of the watermarked signal is achieved. The obtained ODG values confirm that by using the adaptive regulation amplitude of MC modulation, the inaudibility constraint is ensured almost like SS method. But, it is known that tones are more audible and notable compared to random noise. In the proposed MC approach, the transparency of the watermark is ensured because we satisfy the constraint defined by the equation (7) has been satisfied.
Furthermore, embedding the watermark that high frequencies ensure better inaudibility constraint than low ones. So, 75% of carrier frequencies have been chosen in high frequency band.

**Figure 3**  ODG of the watermarked signal using MC and SS modulations (see online version for colours)

![ODG of the watermarked signal using MC and SS modulations](image)

**Figure 4**  Detection reliability of audio watermarking system using MC and SS modulations (see online version for colours)

![Detection reliability of audio watermarking system using MC and SS modulations](image)
Now, in order to compare SS and MC methods with the same audio watermarked quality and under an acceptable inaudibility constraint ($ODG = -1.3$), we fix the scale factor values of SS and MC watermarking systems to $\alpha_{SS} = 1.5$ and $\alpha_{MC} = 1.75$. Figure 4 shows a small improvement of MC method compared to SS one, in terms of detection reliability, when the channel is free from perturbations. To reduce ISI, and minimise much more the BER, we introduce a sufficient guard band between the transmitted symbols. Figure 5 illustrates a performance comparison between different guard bands: CP, ZP and CS.

First, we remark the detection reliability is improved sacrificing symbol bandwidth, in the MC watermarking system, thanks to adding guard bands. Then, different guard bands used in the new system gives equivalent performances with a little improvement of CS compared to CP and ZP. So, CS will be used in the following test experiments as a guard band of MC modulation.

**Figure 5** MC-based audio watermarking system with three types of guard bands (see online version for colours)

In order to evaluate the robustness of the proposed method, the watermarked signal has been submitted to various disturbances defined above. The choice of these perturbations stems from two main reasons. Firstly, previous analytic works have assumed the effect of various distortions on the overall watermarked signal as a stationary AWGN (Kundur and Hatzinakos, 2001). Secondly, MPEG compression and filtering can be modelled as fading-like distortions (Kundur and Hatzinakos, 2001). In order to make comparison with SS-based watermarking algorithms, we used that one proposed in Khalil et al. (2012), Larbi et al. (2004) and Seok et al. (2002). Figure 6 shows the BER as a function of the SNR of SS watermarking methods and MC watermarking one, in the presence of AWGN. The watermark bit rate is fixed to $R = 600$ b/s. As can be seen, the robustness of MC modulation over SS one is emphasised since the BERs with the MC method are lower than those with SS one even when filters and source separation technique are used. However, MC method did not significantly improve SS-based watermarking system. This
Robust audio watermarking based on multi-carrier modulation

is due to the properties of SS modulations that make the watermark signal robust to a narrow-band interference.

**Figure 6** Robustness of watermarking system against AWGN using SS modulation with and without Wiener filter and MC modulation (see online version for colours)

![Figure 6](image)

**Figure 7** Robustness of watermarking system against MPEG compression using SS modulation with and without Wiener filter and MC modulation with and without adaptive carrier frequencies selection (see online version for colours)

![Figure 7](image)

In order to evaluate the robustness of the proposed method against fading like distortion, we submit the watermarked signal to MPEG compression, low-pass filter and high-pass filter. Figure 7 and Table 3 show the gain provided by MC method, in the presence of
fading distortions, compared to SS method. It is known that MPEG compression quantises spectral components non-uniformly at different frequencies and filters out the highest ones in order to preserve a level of perceptual fidelity (Cvejic et al., 2003). Therefore, the proposed MC method is implemented by choosing only low frequencies of sub-carriers. As expected, the robustness of the new system is once again improved due to the embedding watermark in the low frequency components of the host signal.

Note that, there are many watermarking methods that have been already proposed such as reported in Kalantari et al. (2009), Kirovski and Malvar (2002), Seppanen and Cvejic (2004), Vivekananda et al. (2010), and Lu et al. (2006). These methods are robust but they do not allow a high embedding capacity. On the other side, there are other high capacity watermarking schemes such as reported in Fallahpour and Megias (2011), and Pinel et al. (2010). However, these schemes are not robust against various perturbations such as compression. In Table 4, we compare the performance of several recent audio watermarking schemes against MPEG compression. We notice that the main advantage of the proposed scheme is that it ensures a compromise between robustness and watermark bit rate.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Bit rate (b/s)</th>
<th>Robustness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Kirovski and Malvar (2002)</td>
<td>1</td>
<td>×</td>
</tr>
<tr>
<td>Li et al. (2006)</td>
<td>4.2</td>
<td>×</td>
</tr>
<tr>
<td>Kalantari et al. (2009)</td>
<td>15</td>
<td>×</td>
</tr>
<tr>
<td>Seppanen and Cvejic (2004)</td>
<td>27.1</td>
<td>×</td>
</tr>
<tr>
<td>Vivekananda et al. (2010)</td>
<td>45.9</td>
<td>×</td>
</tr>
<tr>
<td>Proposed</td>
<td>600</td>
<td>×</td>
</tr>
<tr>
<td>Fallahpour and Megias (2011)</td>
<td>11,000</td>
<td>×</td>
</tr>
<tr>
<td>Pinel et al. (2010)</td>
<td>250,000</td>
<td>×</td>
</tr>
</tbody>
</table>
5 Conclusions

In order to improve the performance of the watermarking system, a new strategic technique has been proposed in this paper. The new method is based on MC modulation that has proven a large advantage over standard SS ones. The MC parameters have been adapted to the watermarking context in order to improve the inaudibility of the watermark on one hand, and to increase its robustness on the other hand. The next stage of our work is to evaluate the robustness of the proposed method against various desynchronising and node synchronising perturbations.

References


Notes

1 Available at http://soundexpert.org/sound-samples.