WATERMARKING INFLUENCE ON THE STATIONARITY OF AUDIO SIGNALS

S. Larbi, M. Jaïdane

Laboratoire des Systèmes de Communications
Ecole Nationale d'Ingénieurs de Tunis
BP 37, 1002 Tunis, Tunisia
Sonia.Larbi@enit.rnu.tn– nej.jaidane@planet.tn

ABSTRACT

This paper presents an analysis of the perceptual watermarking impact on the stationarity of audio signals. Indeed, the embedded watermark is piecewise stationary, so it modifies the stationarity of the original audio signal. This study is based on stationarity indices, which represent a measure of variations in spectral characteristics of signals, using time frequency representations. Simulation results with two kinds of signals, test signals and audio signals (speech and music) are presented. Stationarity indices comparison between watermarked and original audio signals show a significant stationarity enhancement of the watermarked signal, especially for transient attacks.

1. INTRODUCTION

The non stationarity of audio signals is a problem encountered in various signal processing fields, such as noise reduction, acoustic echo cancellation, and audio coding. This study is related to previous works on robustness enhancement of an adaptive echo canceller driven by the watermarked speech instead of the original [1, 2]. In this paper, we propose an analysis of the watermarking influence on the stationarity of audio signals.

The considered time domain watermarking scheme is shown on figure 1. The embedded watermark signal t_n is synthesized by spectral shaping of the stationary random sequence v_n, which represents the signature in the watermarking context. This is done through an autoregressive filter H_j, using perceptual properties of the human ear. The watermark t_n is then embedded in the original non stationary audio signal x_n to obtain the watermarked audio y_n.

Since the watermark t_n is piecewise stationary, how does it modify the stationarity (or the non stationarity) of the original audio signal? And precisely, is the watermarked signal, as expected, "more" stationary than the original?

In this paper, we address non stationarities which consist in abrupt changes over short durations in the spectral characteristics of audio signals. Stationarity indices (SI's) based on time frequency representations (TFR's) computed in case of original and watermarked audio are compared.

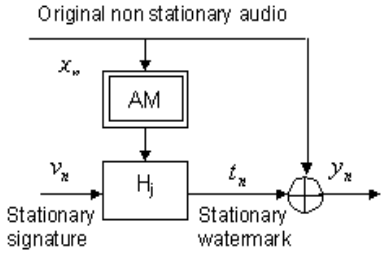


Fig. 1. Perceptual watermarking scheme

The paper is organized as follows: we introduce in section 2 the considered watermarking scheme and the stationarity indices based on Küllback and Bhattacharyya distances. In section 3, we present simulation results obtained with test signals, music and speech, and we make first conclusions on the watermarking influence on the stationarity of audio signals.

2. WATERMARKING AND STATIONARITY

2.1. The perceptual watermarking scheme

We consider here the perceptual watermarking scheme [3] of figure 1. To ensure inaudibility, the embedded watermark signal t_n is obtained by spectral shaping of the stationary and white random sequence v_n through an autoregressive filter H_j of order P. The frequency response module of H_j matches with the masking threshold over the block j of N x_n-samples.

Two kinds of signals are considered here: music and speech. In case of watermarking music signals, a masking threshold $S_m(f)$ is computed by an auditory model[1] (AM) and is updated every 16 ms, corresponding to N-samples blocks. Besides, speech signals can be considered as autoregressive to order P and stationary over durations of 20

[1] In this paper we use the AM N°1 of MPEG.

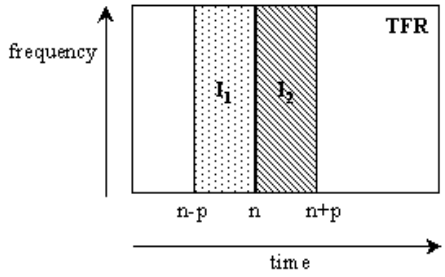


Fig. 2. Subimages I_1 and I_2 of the global TFR

ms. They can be modeled as the output of an allpole filter with transfer function $1/A(z)$. As in perceptual speech coders, we use a perceptual autoregressive filter H_j, given by $1/A(\frac{z}{\gamma})$, where $\gamma \in]0,1[$ is a weighting factor, to shape the white sequence v_n spectrally in order to ensure inaudibility. In both cases (music and speech), the coefficients of H_j are updated over each block j, thus the watermark t_n can be considered as piecewise stationary over each block.

2.2. Stationarity indices [4]

Several non stationarities detection methods based on autoregressive spectral modelling seem to be inefficient in case of short duration transients. The method we choose to detect non stationarities in audio signals, proposed in [4], is based on distance measures between different TFR's of the signal. The TFR's are computed over the audio signal duration. At each analysis instant n, two subimages $I_1(n;\tau,f)$ and $I_2(n;\tau,f)$ with equal duration p are extracted from the global TFR on both sides of the instant n, as illustrated on figure 2, where:

$$\begin{array}{l} I_1(n;\tau,f) = TFR(n-p+\tau,f) \\ I_2(n;\tau,f) = TFR(n+\tau,f) \end{array} \quad (1)$$

In this paper, we use the spectrogram with a Hamming smoothing window of length N_h.

The parameter p is the duration of each subimage and $\tau \in [0,p]$. Both subimages are then normalized as follows:

$$NI_k(n;\tau,f) = \frac{|I_k(n;\tau,f)|}{\int_{\tau=0}^{p}\int_{-\infty}^{+\infty}|I_k(n;\tau,f)|\,df\,d\tau} \quad k=1,2 \quad (2)$$

and compared by computing a distance measure D. The stationarity indices (SI's) used in this paper are based on the following distances:

The Küllback distance:

$$SI_{ku}(n) = \int_{\tau=0}^{p}\int_{-\infty}^{+\infty}(NI_1(n;\tau,f)-NI_2(n;\tau,f)) \cdot \log\left(\frac{NI_1(n;\tau,f)}{NI_2(n;\tau,f)}\right) df\,d\tau \quad (3)$$

The Bhattacharyya distance:

$$SI_{bh}(n) = -Log\left(\int_{\tau=0}^{p}\int_{-\infty}^{+\infty}\sqrt{NI_1(n;\tau,f).NI_2(n;\tau,f)}\right) df\ d\tau \quad (4)$$

If the signal characteristics present no changes at instant n, the distance D is near zero and it peaks otherwise.

Note that the parameter p delimits the considered analysis time at each instant n and it allows the sensitivity control of the SI's: higher p values lead to smoother SI's.

3. WATERMARKING INFLUENCE ON THE STATIONARITY OF AUDIO SIGNALS

In this section, we show that adding a stationary watermark t_n in the time domain modifies the non stationarity characteristics of the original audio signal.

As t_n is stationary over N-samples blocks, we focus our study on the watermarking influence over blocks of samples presenting short duration transients. Indeed, music can be considered as a succession of periods of relative stability, in spite of the presence of transient attacks, such as percussion, inducing high frequency noise [5]. Speech is rather a rapid succession of noise periods, such as unvoiced consonants, periods of relative stability as vowels, and periods of silence.

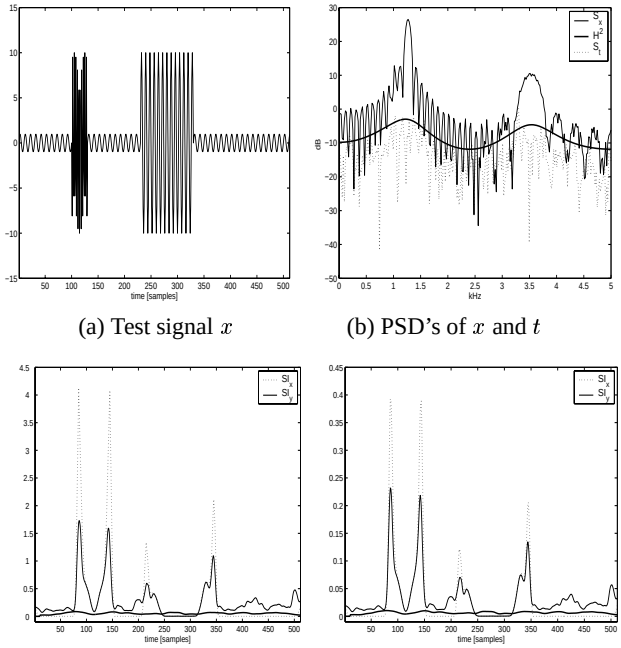
In the following, we begin with presenting simulation results using test signals, which correspond to artificial transients, and their watermarked version. This is done in order to display the behaviour of the SI's in this particular context and to set the right values for the sensitivity parameter p.

3.1. Artificial transients

The test signal has been generated with a sampling frequency $f_e = 10\ kHz$ and is 512 samples long. The corresponding watermark signal t has been synthesized, as depicted on figure 1, using a perceptual spectral shaping filter of order $P = 10$ with a perceptual factor $\gamma = 0.8$.

The test signal x presents two kinds of transients occuring at samples 100 and 230. The first transient segment consists in short duration frequency and amplitude changes and the second one consists in a smaller frequency change but with the same amplitude variation as the first transient. During each segment, the signal x is composed from only one tone. We define in the following a finite duration tone as $x(a,f,[n_1:n_2])$, where a is the amplitude, f the normalized frequency and $[n_1:n_2]$ the time duration. The test signal is then given by (cf. figure 3a):

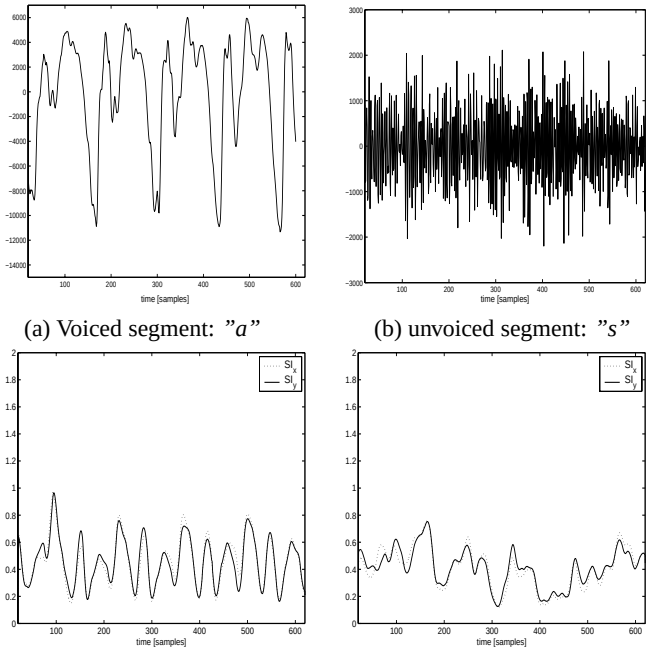
– stationary zones: $x(1,\ 0.1,\ [1:100])$, $x(1,\ 0.1,\ [131:230])$, $x(1,\ 0.1,\ [331:512])$,
– first transient: $x(10,\ 0.35,\ [101:130])$,
– second transient: $x(10,\ 0.125,\ [231:330])$.



(a) Test signal x (b) PSD's of x and t

(c) Küllback distance (d) Bhattacharyya distance

Fig. 3. Simulation results for test signal x.
Simulation settings: $p = 10$, $N_h = 31$, $P = 10$, $\gamma = 0.8$.



(a) Voiced segment: *"a"* (b) unvoiced segment: *"s"*

(c) Küllback distances

Fig. 4. Küllback SI's of the voiced and unvoiced segments.
Simulation settings: $p = 20$, $N_h = 61$, $P = 20$, $\gamma = 0.8$.

We denote respectively SI_x and SI_y the SI's of the original signal x and the one of the watermarked signal $y = x + t$. Figure 3b shows the power spectral densities (PSD's) of the signal x (full line) and of the corresponding watermark t (dotted line) with the squared frequency response module of the perceptual filter (thick line). We compare then the computed Küllback and Bhattacharyya SI's of x and those of its watermarked version y on figures 3c and 3d respectively. As expected, the SI's peak at both transients boundaries and are near zero elsewhere. On both figures, the SI in the watermarked case has been significantly decreased (to about 50%) at the transient boundaries. We notice that the SI's react better (with higher peaks) in case of greater frequency jumps Δf. Indeed, in the first transient $\Delta f = 0.25$ and in the second $\Delta f = 0.025$.

The factor p should not be greater than $\Delta t/2$, where Δt is the duration between successive transient boundaries, otherwise successive SI peaks will overlap and make inaccurate the detection of each transient instant.

The thick line in figures 3c and 3d corresponds to the SI of the watermark t, computed with $p = 10$ and $N_h = 127$. As expected, the stationarity index is near zero.

3.2. Speech signals

In case of speech signals, we consider transitions between unvoiced to voiced zones as transients. We start with analysing 20 ms voiced and unvoiced segments separately (figures 4a and 4b respectively). We notice neither significant enhancement nor degradation of the stationarity. Indeed, both Küllback based SI's, SI_x and SI_y, are approximately the same and they have small values. That is because noise like unvoiced segments and harmonic voiced segments can be considered as stationary. On figure 5a, we consider a segment of the syllable *"sai"* to analyse the unvoiced/voiced transition. As expected, the SI_y value at the transition area (about $n = 220$) has been significantly reduced.

In the following section, we analyse music signals. We will see that the stationarity enhancement by watermarking is more important. Indeed, attack transients are more present in music.

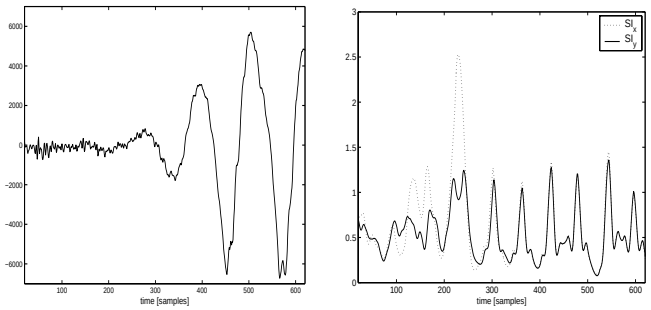
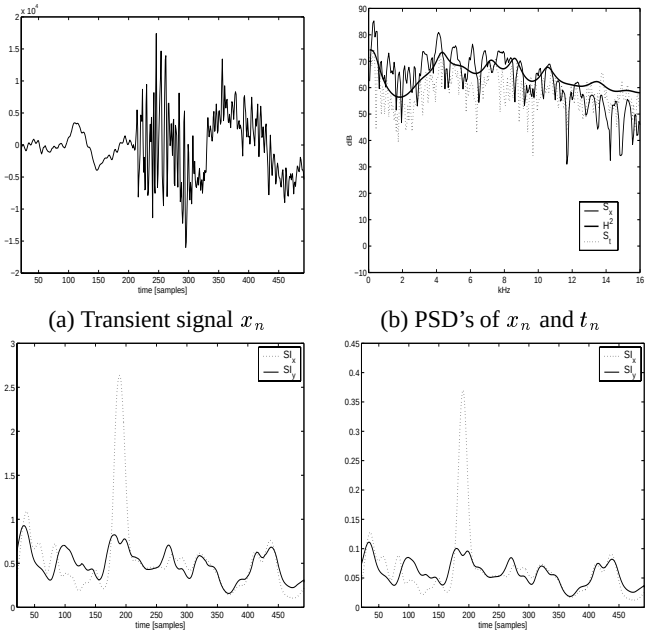


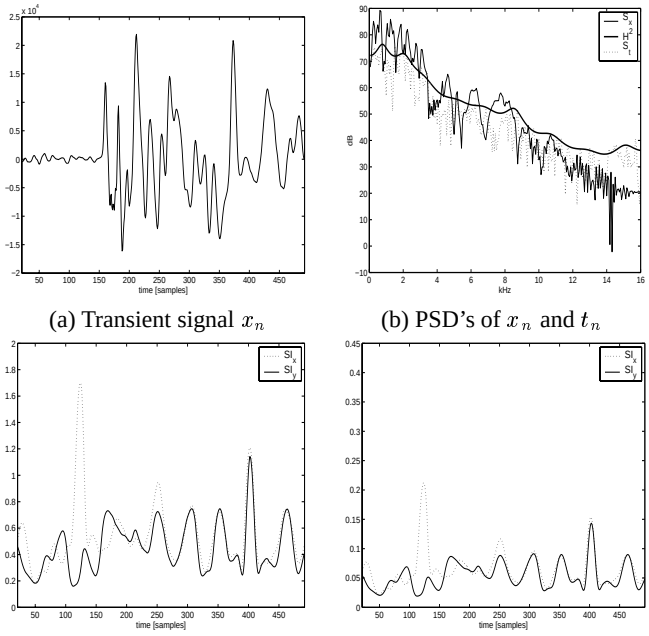
Fig. 5. Unvoiced/voiced transition and Küllback distance (right).
Simulation settings: $p = 20$, $N_h = 61$, $P = 20$, $\gamma = 0.8$.



(a) Transient signal x_n (b) PSD's of x_n and t_n

(c) Küllback distance (d) Bhattacharyya distance

Fig. 6. Percussion attack and consonant *p* (Michel Jonasz). Simulation settings: $p = 20$, $N_h = 61$, $P = 20$.



(a) Transient signal x_n (b) PSD's of x_n and t_n

(c) Küllback distance (d) Bhattacharyya distance

Fig. 7. Indian zyther attack
Simulation settings: $p = 20$, $N_h = 61$, $P = 20$.

3.3. Music signals

Music attacks correspond to the beginning of notes produced by an instrument. They are areas of short duration energy (about a few ms) with rapid variations of the sound signal. Particularly, attacks are accompagned by an abrupt short time energy increase distributed on the whole spectrum and noticeable in the high frequencies, since energy is usually concentrated in the low ones [5].

We analyse in the following two transients over durations of 16 ms, extracted from *Lucille* of Michel Jonasz (figure 6a) and an indian zither excerpt (figure 7b), both sampled at $32\ kHz$. The first transient corresponds to a percussion attack accompanied by the consonant *"p"*. We note on figure 6b the high frequency noise caused by the percussion. Figures 6c and 6d show the important decrease of the SI_y value compared to that of SI_x at the attack instant ($n = 200$) for both the Küllback and the Bhattacharyya distances. Similarly, for the zyther attack of figure 7, SI_y has been significantly reduced to about 50% of the SI_x value. However, as in the speech case, no noticeable stationarity enhancement is achieved for harmonic music segments.

4. CONCLUSION

This paper presented a study of the time domain perceptual watermarking influence on the stationarity of audio signals. Indeed, the embedded watermark is piecewise stationary, so it modifies the stationarity of the original audio signal. The non stationarity measure we used is based on TFR's and stationarity indices. Simulation results with two kinds of signals, test and audio signals, show a significant stationarity enhancement of short segments presenting transient attacks. This enhancement is limited to transient areas and is more important in case of music, since attacks are more present in those signals. Perceptual watermarking can be viewed as a preprocessing step that "*stationnarizes*" audio signals.

Thanks

The authors would like to present their thanks to Professor C. Doncarli and M. Davy (IRCCyN, Nantes) for their helpful advises.

5. REFERENCES

[1] A. Gilloire, V. Turbin, "Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers", in Proc. ICASSP 1998, pp. 3681-3684.

[2] S. Larbi, M. Jaïdane, M. Turki, M. Bonnet, "Acoustic echo canceller robust to speech non stationarities" *in french*, CORESA 2001, Toulouse, France.

[3] L. Boney, A. H. Tewfik, K. N. Hamdy, "Digital watermarks for audio signals", IEEE Int. Conf. Multimedia Computing and Systems, Japan, June 1996.

[4] H. Laurent, C. Doncarli, "Stationarity index for abrupt changes detection in the time-frequency plane", IEEE Signal Processing Letters, vol. 5, no. 2, 1998.

[5] X. Rodet, F. Jaillet, " Detection and modelling of fast attack transients", ICMC 2001, Cuba, Sept. 2001.