ENHANCED AUDIO DATA HIDING SYNCHRONIZATION USING NON LINEAR FILTERS

A. LoboGuerrero, F. Marques, P. Bas, J. Lienard,

Laboratoire d'Images et Signaux-961 rue de la Houille Blanche-38042 St. Martin d'Heres Cedex - FRANCE
Universidad Politecnica de Cataluna-Campus Nord,D5-C/Jordi Girona-08034 Barcelona Spain

ABSTRACT

This paper address the problem of synchronization in the context of audio data hiding. For real time transmissions purposes, the data decoding process has to deal with synchronization issues. This paper proposes an new synchronization scheme that optimizes performances of the systems which are based on spread spectrum synchronization by the use of mathematical morphologic tools that present good performance for peak detection. A brief theoretical presentation of the Top-Hat filter is recalled and the enhanced system is derived from the analysis of the advantages and disadvantages. The final scheme provides a robust synchronization system and is compared with the classical solutions.

1. INTRODUCTION

Audio data hiding [1][2] is a method that allows the insertion of an imperceptible mark in a signal data set. Although this technique is often used to guarantee copyrights, applications in different areas are constantly been developed. This is the case of data hiding for data transmission. Several applications are possible in this particular context. For example, if the priority information which is transmitted is the speech voice, then it may be useful to transmit video analysis information such as lips moving parameters, the speaker's face or any other information identifying the speaking face. Such an application is addressed in the project ARTUS[1] which aims to embed animation parameters into audio and video contents. When implementing audio data hiding schemes the designer has to take into account not only coding but also synchronization constraints. Theses constraints are:

- Minimum acquisition delay: the time devoted for synchronization can not be used for decoding the hidden message;
- Temporal drift management: due to DA/AD conversions, the global system has to cope with resampling and temporal drift;
- Robustness to noise and compression: those two process are part of the transmission channel;
- Minimal audible distortion;
- Maximal transmission rate of the inserted message.

[1]This project was funded by French RNRT Artus Project

Watermarking synchronization for audio is classically achieved using spread spectrum (S.S.) synchronization theory [3] or audio characteristics as beat [5]. In this paper, an enhanced spread spectrum synchronization system is proposed by introducing an improved detection criteria which reduces the sequence length by 90 %. Such an improvement is due to the use of non linear filters (morphology mathematical filters) that are applied to analyze the autocorrelation function. This paper is organized as follows:
At first, a brief introduction of the classical synchronization technique is presented. Its related problems are outlined and the possible improvements are then derived. Then the proposed scheme based on mathematical morphology is introduced. The last section outlines the improvement by comparing the synchronization behaviors of the presented schemes.

2. CLASSICAL SPREAD SPECTRUM SYNCHRONIZATION

2.1. General presentation

The most popular synchronization system in the field of telecommunication is based on the classical spread spectrum theory[3]. This system uses a specific code (also called pseudo random signal) for guaranteeing synchronization. This code is added to the host signal in the transmission stage. The synchronization at the receiver may be implemented by using a sliding correlator matched with the received code. Then, the resulting signal is

$$y(n) = \{s(n) + Ac(n)\} \otimes c^*(n - m) \qquad (1)$$

Where A determines the power of the inserted code. The correlator output is then compared with a threshold to determine the phase of the received synchronization code. The figure 1 shows how the system proceeds.

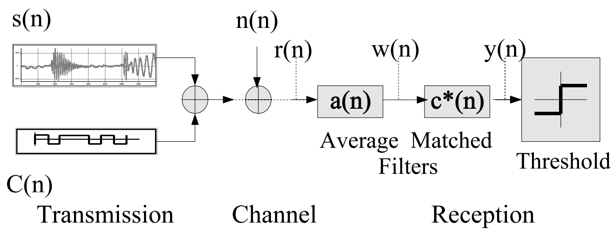


Fig. 1. Synchronization in Spread Spectrum

When the SNR is not large enough (e.g. the power of the code is too weak), the intercorrelation between the code and the noise may prevent the localization of the peak. In this case, before evaluating the threshold, a supplementary filter can be applied to reduce the noise perturbation. One classical solution to solve this drawback is to combine the use of a periodic code and an average filter. This average filter $a(n)$ is expressed by the next equation:

$$w(n) = \frac{1}{K}\sum_{k=0}^{K} r_k(n) \tag{2}$$

In this filter, the received signal $r(n)$ (code plus channel noise) has been divided in blocks noted with the sub index k. The number of total averaged blocks is determined by the constant K. This filter represents an averaging system that aims to reduce the noise effect. Given:

$$y(n) = w(n) \otimes c^*(n-m) \tag{3}$$

Replacing $w(n)$,$R(n)$ and regrouping,

$$y(n) = \frac{A}{K}\left\{\sum_{k=0}^{K} c_k(n) \otimes c^*(n-m)\right\} + \frac{A}{K}\left\{\sum n_k(n) \otimes c^*(n-m)\right\} \tag{4}$$

Taking the limit when K tends toward infinity the first term in equation 4 will give $c_k(n) \otimes c^*(n-m)$, the second term will tend toward zero. It means that the effect of the noise $n(n)$ is eliminated and the codes autocorrelation is preserved guaranteeing synchronization.

2.2. Disadvantages of classical synchronization

This synchronization process leads to several disadvantages when applied to audio data hiding. In this kind of application, the host signal $s(n)$ is often considered as a noise. The constraint of inaudibility imposes a very weak SNR. To deal with such a constraints, three strategies may be used:

1. To increase the power of the inserted codes ;
2. To increase length codes ;
3. To increase K (equivalent to strategy 2)

Theses strategies can be studied in equation 4. The strategy 1 is often discarded because of the high probability of adding audible distortion. The second and third options imply two main disadvantages:

- Important acquisition delays;
- Unsuitable in presence of temporal clock drift.

We propose in this paper to use morphological mathematics to improve the above issues. In the following, the advantages of this technique are introduced and the resulting synchronization scheme is presented.

3. PROPOSED SCHEME

The proposed synchronization system has be designed to reach the following objectives:

- Offer an improved detection criteria;
- Attain an adaptation to SNR conditions of Audio data hiding;
- Reduce the acquisition delay;
- Lower the sensitivity to the clock drift.

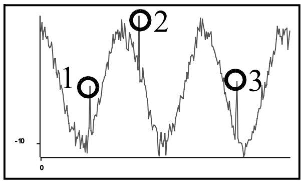
3.1. Dealing with peak detection problem

In the ideal conditions, a threshold is enough to detect the peak of an autocorrelation signal. However, in the case of audio synchronization, the use of a simple threshold is not enough to detect this characteristic: the peaks are often lost as shown in figure 2(a). In such a case, the mathematical morphology[6] yields to a constant threshold. If a signal, as the one showed in figure 2(a), is applied to a Top-Hat morphologic filter, the resulting signal will enhance the presence of the peaks as shown in figure 2(b). From this figure it is easy to observe that the application of a constant threshold will allows de correct detection of the desired peaks. The formalization of the Top-Hat filter is introduced in the following subsection.

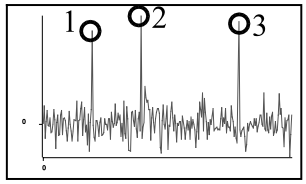
3.2. Mathematical Morphologic tools

Mathematical Morphology is represented by non linear processes and consequently is non widely used for audio processing. Nevertheless, morphological filters are very efficient for ameliorating the quality of a signal (often an image) or for extracting some of its features. Morphological tools enable parameters extraction and signal restoration.

A morphological analysis [7] of the spread spectrum system can improve the behavior of the overall system. To



(a)Noisy Peaks



(b)Top-Hat Effect

Fig. 2. Effect of non linear filtering: if peaks 1,2 and 3 cannot be detected using classical S.S. technique, the use of morphological filters yields successful results

that end, the signal processing is based additionally on operations of max and min filters defined over a N samples window as follows:

$$mm(x(n)) = \max\{x(n)\}\forall n[n_1, n_1 + N] \tag{5}$$

$$nn(x(n)) = \min\{x(n)\}\forall n[n_1, n_1 + N] \tag{6}$$

From this definition, more complex structures can be defined. Erosion and dilatation are the two most important operations. Their equations are:

$$\varphi\{x(n)\} = m\{M(x(n))\} \tag{7}$$

$$\phi\{x(n)\} = M\{m(x(n))\} \tag{8}$$

Conceptually, an opening is an operation that eliminates from the original signal all the structures smaller than the window size N. In the same way, a closing eliminates all the structures bigger than the window size N. In our scheme, we have used a combination of openings and closings operations defined by the Top-Hat filter which is determined by the following equation:

$$th(n) = x(n) - \varphi\{x(n)\} \tag{9}$$

3.3. Top Hat Vs. Classical Filtering

A frequency domain description of morphological filters is not possible. Its non linear behavior gives this description impracticable. Anyway, it could be interesting to show that the top hat filter is not similar with a classic high pass filter. This is easily done by comparison of their impulse responses. To illustrate this, lets take a Top-Hat filter designed to match an unit impulse. Its impulse response will be exactly the same unit impulse. By the other hand, the impulse response will not be an unit impulse but a sync function.

In other words, a high pass filter will eliminate all the low frequencies (including those from the unit impulse) and in consequence the unit impulse will be corrupted. However, if an impulse corrupted by a low frequency signal is applied to the Top Hat filter, this filter will eliminate the low frequencies of the corrupting signal preserving the unit impulse (high and low frequencies) unmodified. Comparative results will be presented at the end of this paper.

3.4. Implementation

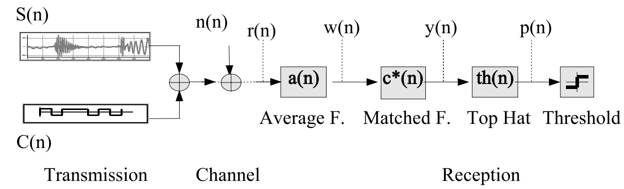


Fig. 3. Mathematical Morphology Synchronization Scheme

The figure 3 shows a scheme composed of three stages. The transmission stage and the channel are the same as the one used in classical spread spectrum systems as presented in subsection 2.1.

The reception stage comports an important amelioration in comparison with the classical system. The filters $a(n)$ and $c*(n-m)$ are the same used in the classical scheme. However, the filter $th(n)$ is the Top Hat filter presented in 3.2. The idea is to define which is the particular characteristic we are looking for.

From the equation 4 we can replace the noise signal $n(n)$ by $s(n) + n(n)$. The signal $y(n)$ will be noted as:

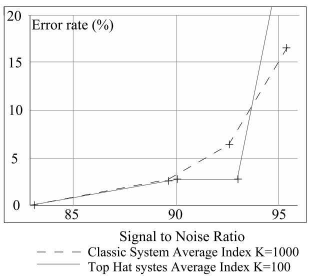
$$y(n) = \frac{A}{K}\left\{\sum_{k=0}^{K} c_k(n) \otimes c^*(n-m)\right\} +$$

$$\frac{A}{K}\left\{\sum \{s_k(n) + n_k(n)\} \otimes c^*(n-m)\right\} \tag{10}$$

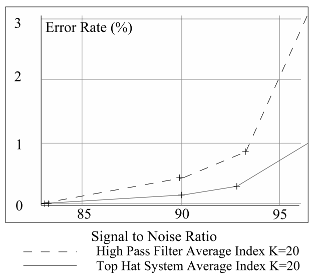
In this equation two main items are noted. The second shows the influence of the host signal and channel noise in the system. The first term represents the averaged auto correlation of the synchronism codes. As explained in section 2.1 the average filter $a(n)$ has not effect on this autocorrelation and by consequence this first term is represented ideally by an impulse $A\delta(n-m)$. It is well known that this function takes values different to zero only in one sample. This special characteristic bring us to the use of the Top-Hat filter. In fact, if the window size N in the Top-Hat filter is defined to match the ideal autocorrelation of the codes(see section 3.2), it will eliminate all parts of the signal which its temporal dynamics are larger than one sample. This allows the detection of a discrete impulse (corresponding to the codes autocorrelation) following the method that was explained in subsection 3.1.

4. RESULTS

In order to test and quantify the behavior of this proposition, tests over a file of 13 minutes long have been realized. The sampling frequency is 44.1 kHz (CD audio Quality), a channel with a SNR=45 dB's have been simulated. We have compared the classical S.S. to the proposed system varying the power of the inserted codes. The codes are 512 samples long. The following figure(a) shows the results in terms of percentage of detected errors. An error represents an instant that have appeared and/or disappeared after transmission. K is fixed equal to 1000 in the classical synchronization system and equal to 100 in the proposed scheme. Then we have compared the proposed system with a high pass (derivative) filter for $K = 20$ and the correspondign results are depicted in the next figure(b). The analyze of these figures shows that the classical spread spectrum synchronism system needs to average the incoming signal until 10 times more in order to obtain the same error probability than the mathematical morphology system. It means that the proposed scheme needs to accumulate ten times less blocks of signals and the acquisition delay and clock drift problems are reduced in consequence.



(a)Classic Vs. Non Linear



(b)High Pass Vs. Non Linear

Fig. 4. Comparative Results

This result can be easy explained: in order to bring out the impulses representing the codes instants, it is always necessary to average the incoming signal. However, the K iterations used in the proposed method will be less than if the top hat filter is not present. In absence of top hat filter the system will detect correctly a unit impulse only if its amplitude A is greater than the threshold value and the rest of the signal is lower than it (cf. figure 2). This implies that K must be enough to cancel almost all the noise influence.

By the other hand, the presence of the top hat filter avoids this situation. In fact, it is simple necessary that the unit impulse be greater than the other peaks that have the same temporal duration as was illustrated in figure 2. In this case, the K value must only attenuate the noise but it does not need to obtain unit impulses greater than the rest of the signal. The immediate effect of this operation is the decrease of the number of iterations K, and by consequence, the acquisition time will be lower. Figure 4(b) confirms the fact that the proposed scheme achieves better synchronization performance than high pass filtering for the same average value K. The comparison between the high pass filter (FIR equals with +1,-1) and the proposed scheme confirms that the top hat filter presents better results for the same average value K.

5. CONCLUDING REMARKS

The mathematical morphology system has been developed in order to improve the behavior of the classical synchronization systems based on the spread spectrum modulation for an application in audio data hiding. This scheme presents until ten times lower time acquisition delay. It gives also better performances than a high pass filter. This system studies the host signal as a white noise. However, the host signal is not white but colored and absolutely deterministic. In the future, the use of this information could be used to optimize the system and to develop mathematical morphology informed approach [8].

6. REFERENCES

[1] L. Boney, A. H. Tewfik and K. N. Hamdy. *Digital Watermarks for Audio*. Eusipco, pages 473-480, Triestre, 1996.

[2] A, LoboGuerrero P. Bas, J. Linard. *Iterative Informed Audio Data Hiding Scheme Using Optimal Filter* ICCT 2003, pages 1408-1412, Pekin, Chine.

[3] S. Haykin, *Digital Communications*. John Wiley and Sons, 1988.

[4] J.G. Proakis, *Digital Communications* McGraw-Hill, New York, 1995.

[5] 'D. Kirovski and H. Attias'. *Audio Watermark Robustness to Desynchronization via Beat Detection*, in book *"Lecture Notes in Computer Science, 5th Information Hiding Workshop"*, 2002,pages 176-160.

[6] J. Serra, *Image Analysis and Mathematical Morphology* Academic Press, 1982.

[7] P. Soille, *Morphological Image Analysis. Principles and applications* Springer-Verlag, 1999.

[8] LoboGuerrero A., Bas P., Lienard J., *An Informed-Synchronization Scheme for Audio Data-Hiding*, SPIE, 2004, to be published.