FACTA UNIVERSITATIS (NIŠ)
SERIES: ELECTRONICS AND ENERGETICS VOL. 12, NO.2 (1999), 103–122

UDC 621.391

DATA HIDING WITHIN AUDIO SIGNALS

This paper is dedicated to Prof. Ilija Stojanović on the occasion of his 75^{th} birthday and the 50^{th} anniversary of his scientific work

Rade Petrović, Joseph M. Winograd, Kanaan Jemili and Eric Metois

Abstract. In this paper we will present general principles of steganography, basic terminology, and an overview of applications and techniques. In particular we will consider data hiding within audio signals, basic requirements and the state of the art techniques. We will propose a novel technique, the short-term autocorrelation modulation, with several variations. The proposed method is characterized by perfect transparency, robustness, high bit rate, low processing load, and, particularly, high security.

1. Introduction

Data hiding is also known as steganography (from the Greek words *stegano* for "covered" and *graphos*, "to write"). In contrast to cryptography, which focuses on rendering messages unintelligible to any unauthorized persons who might intercept them, the heart of steganography lies in devising astute and undetectable methods of concealing messages themselves. An obvious application is a covert communication using innocuous cover signals, like a telephone conversation or an image. Another application, known as (digital) watermarking, refers to embedding an unobtrusive mark into an object, which can be used to identify the object. For example, a digital watermark can be inserted into a piece of music, so that radio and TV broadcasts can be monitored automatically for royalty payment purposes. Many other applications, such as piracy detection and/or prevention, proof

Manuscript received June 15, 1999.

A version of this paper was presented at the fourth Conference Telecommunications in Modern Satellite and Cable Services, TELSIKS'99, October 1999, Niš, Serbia.

Authors are with Aris Technology, Inc.,1972 Massachusetts Ave., Cambrige, MA 02140, e-mail tcrade@aristech.com.

of performance (e.g. monitoring time and duration of advertisement broadcasts), integrity verification (to detect tampering of a cover signal), traitor tracing, (e.g. to identify a source of a leak), transaction identification, automatic inventory, copy control, auxiliary information attachment, etc., have been reported in literature [1 - 19].

General principles of steganography, as well as terminology adopted at the First International Workshop on Information Hiding, Cambridge, U.K. [20] are illustrated in Fig.1. A data message is hidden within a cover signal (object) in the block called embeddor using a stego key, which is a secret set of parameters of a known hiding algorithm. The output of the embeddor is called stego signal (object). After transmission, recording, and other signal processing which may contaminate and distort the stego signal, the embedded message is retrieved using the appropriate stego key in the block called extractor.

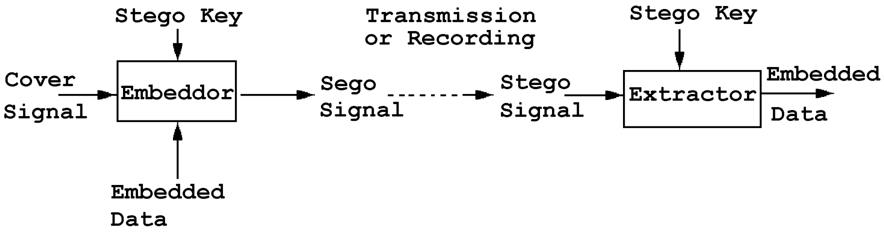

Fig. 1. Block diagram of data hiding and retrieval

A number of different cover objects (signals) can be used to carry hidden messages. Typical cover objects are images, various files (ASCII, .doc, .ps, etc), printed documents, faxes, program files, music, telephone signals, video, radio signals, etc. Data hiding techniques strongly depend on the nature of the cover objects, and vary widely, limited only by designer's ingenuity. However, some general hiding strategies have emerged.

One set of hiding techniques use redundancy of the cover object, where a message is embedded by selecting among valid alternatives in a predefined manner. Those techniques are most common for stego object where no loss of information is permitted, such as computer programs. For example, the order of push and pop operations can hide information.

In the case of analog signals (e.g. audio or video), typical approach is to introduce a small, predefined contamination or distortion, similar to those occurring in normal transmission and/or processing of the cover signal, such as addition of Gaussian noise, low level background signals, fading patterns, echoes, phase shifts, etc. Those modifications should be unobtrusive, but should be detectable by appropriate extractors.

In this paper we will discuss some basic requirements for hiding messages in audio signals, and review the state of the art techniques. Then we will present a novel approach and discuss its performance.

2. Background

Data hiding in audio signals exploits imperfection of human auditory system known as audio masking. In presence of a loud signal (masker), another weaker signal may be inaudible, depending on spectral and temporal characteristics of both masked signal and masker [21]. Masking models are extensively studied for perceptual compression of audio signals such as MPEG, AAC, Dolby AC-3 etc. (e.g. see [5], [22], [32] and [33]). In the case of perceptual compression the quantization noise is hidden below the masking threshold, while in a data hiding application the embedded signal is hidden there.

Besides transparency of the embedding process, it is important to insure robustness of the embedded signal. For example, it is essential that the embedded signal is detectable after a perceptual compression encoding/decoding. In general, the embedded signal should survive any signal processing that produce acceptable quality of the cover signal.

Further, the embedded signal should survive attacks intended to remove and/or forge the message. Many attacks have been designed against different stego systems (e.g. [23], [26], [30], and [31]), and we expect an endless battle between designers and attackers, like in the cryptography arena.

Many other requirements may be imposed on a stego system, like simplicity of embeddor and/or extractor, layering of multiple watermarks, high throughput (long messages), low probability of false detections, etc. A list of requirements put forth by music industry can be found in [11].

Most frequent technique for data hiding within audio signal is based on the direct sequence spread spectrum (DSSS) approach (see [2], [3], [5], [7], [9], [14], [15], [18], and [36]). The spread spectrum signal is shaped in both time and frequency domain to fit under the masking threshold of the audio signal, and then inserted into the cover signal. An extractor uses a sliding correlator that correlates the received signal to the predefined spread spectrum template. In some cases the spread spectrum signal is modulated to fit some sub-band of the audio signal (e.g. [9] and [14]). In other cases signal is "whitened" in the extractor before the correlation (e.g. [15] and

[36]), or processed by a rake receiver (e.g. [15]).

The most serious problem of the DSSS approach is its vulnerability to a time scale modification, which can be an inadvertent effect of a standard signal processing or a deliberate attack. Time scale modifications (speed-up or slow-down) occur in analog tapes (known as wow and flutter) or due to clock mismatches in D/A & A/D conversions. Further, linear speed-ups are common in broadcasts in order to shorten the playtime. Alternatively pitch-invariant time compression or expansion like in [25] is used some times in TV broadcasting. This vulnerability to a time scale modification can be further exploited in attacks designed to erase DSSS watermarks, i.e. prevent their detection (see [23] and [26]).

Further, various techniques that are comprised of inserting modulated or unmodulated tones at pre-selected frequencies are proposed (e.g. [17], [27] and [34]). Those techniques are found to have difficulty meeting the transparency and security requirements. Similar problems were found with so called "notch filtering" techniques (e.g. [16], [27] and [28]), where a narrow-band filter is used to eliminate components in the cover signal spectrum at predefined frequencies.

Techniques that employ echo insertion and cepstral- domain extraction (e.g. [2], [29]) are known to cause perceptible signal distortions and/or show low robustness. Further, they have a relatively high extractor complexity.

Techniques based on replacing one or more of the less significant bits of a digitized audio signal by a hidden data (e.g. [2]), are characterized by a low robustness (and thus a low security as well).

Techniques based on phase modulation [2] exploit the human audio system insensitivity to relative phase of different spectral components. However, many channels do not preserve this phase relationship either, and it is easy to design an attack.

Some proposed techniques are integrated with MPEG AAC compression, using redundancy within the compression process (e.g. [8] and [34]). Apparently, those techniques are appropriate only for compressed audio signals and the watermark is lost as soon as the signal is decompressed.

Overall, none of the proposed techniques meets fully the security requirement, while other requirements can be met at various levels. We will propose here a novel approach, which emphasizes the security issue, but also can achieve very high performance with respect to all other system requirements.

3. Autocorrelation Modulation

Any embedding process can be considered in principle, as adding a difference signal to the cover signal to obtain a stego signal, as illustrated in Fig. 2. The difference signal depends on the cover signal, embedded data, as well as the stego key.

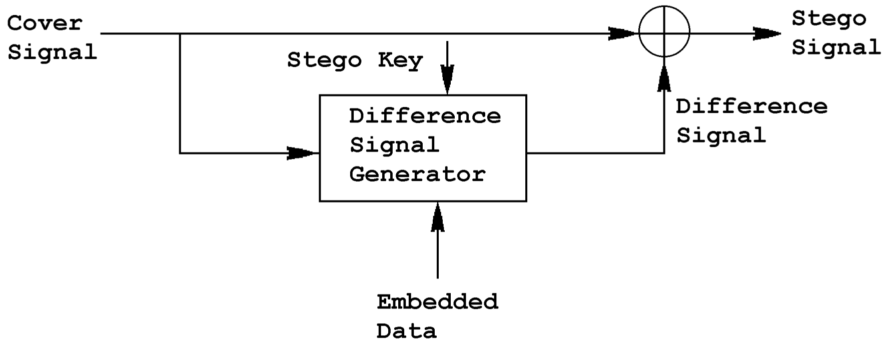

Fig. 2. Block diagram of an embeddor

Fig. 3 shows the block diagram for the difference signal generator for the short-term autocorrelation modulation. Input signal is obtained by applying a filter to the cover signal in order to obtain a portion of the cover signal that will result in minimal disturbance to the audio signal and the best hiding properties. The filtered cover signal is used to generate the difference signal or a component of the difference signal if multiple difference signals are added on top of each other.

The difference signal component generator produces a difference signal component by applying a variable gain or attenuation (of positive or negative value) to a delayed (or advanced) version of the filtered cover signal. The amount of delay or advancement corresponds to the autocorrelation delay at which the signal is being modulated.

The amount of gain that is applied at any time or spatial instant is determined by the gain calculator, and depends on the properties of the filtered cover signal and embedded data. This amount can be determined according to a variety of different methods as in the derivation that follows[1].

[1] These derivations are presented for the case of a one-dimensional signal (such as audio), but can be readily extended to describe the application of the technique to multidimensional signals (such as video).

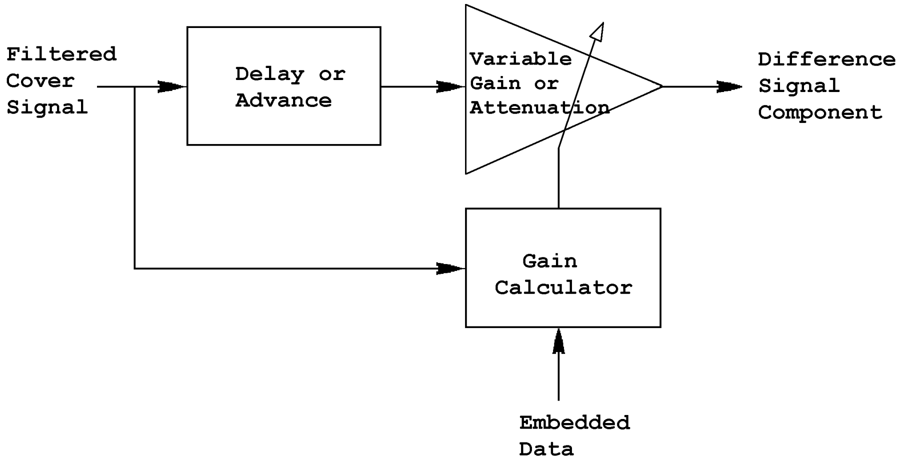

Fig. 3. Difference signal component generator

The short-term autocorrelation of the filtered cover signal can be expressed by the formula:

$$R(t,\tau) = \int_{t-T}^{t} s(x)s(x-\tau)dx \tag{1}$$

where $s(t)$ is the filtered cover signal, $R(t,\tau)$ is its short-term autocorrelation, τ is the delay at which the autocorrelation is evaluated, T is the temporal integration interval, and t is used to denote time.

By adding a difference signal $e(t)$ to the filtered cover signal, the short-term autocorrelation function of the resulting signal is modulated in such a way as to obtain:

$$\begin{aligned} R_m(t,\tau) =& \int_{t-T}^{t} (s(x)+e(x))(s(x-\tau)+e(x-\tau))dx \\ =& R(t,\tau) + \int_{t-T}^{t} (s(x)e(x-\tau)+e(x)s(x-\tau)+e(x)e(x-\tau))dx \end{aligned} \tag{2}$$

With an appropriate choice of the difference signal we can achieve an increase or decrease of the short-term autocorrelation function. While it is apparent that many kinds of difference signals might be used to perform this modulation, (indeed, it would require a special design for difference function to *not* affect the short term autocorrelation function), the simplest application uses delayed or advanced versions of the existing signal multiplied by an amount of gain or attenuation. That is,

$$e(t) = gs(t - \tau) \tag{3a}$$

or

$$e(t) = gs(t + \tau) \tag{3b}$$

Substituting (3a) and (3b) into (2), we find that the short- time autocorrelation of the resulting signal can be written as:

$$R_m(t, \tau) = R(t, \tau) + gR(t, 2\tau) + gR(t - \tau, 0) + g^2 R(t - \tau, \tau) \tag{4a}$$

or

$$R_m(t, \tau) = R(t, \tau) + gR(t, 0) + gR(t + \tau, 2\tau) + g^2 R(t + \tau, \tau) \tag{4b}$$

respectively.

The autocorrelation functions of the existing signal which appear on the right hand side of equations (4a) or (4b) can be measured, and their values used to obtain the solution g that will produce a desired value for $R_m(t, \tau)$.

In a typical implementation it is desirable to have small values for g in order to keep the difference signal transparent with respect to the intended purpose of the existing signal. In this case, the g^2 terms in equations (4a) and (4b) are negligible and the exact gain value can be closely approximated by:

$$g \approx \frac{R_m(t, \tau) - R(t, \tau)}{R(t, 2\tau) + R(t - \tau, 0)} \tag{5a}$$

or

$$g \approx \frac{R_m(t, \tau) - R(t, \tau)}{R(t, 0) + R(t + \tau, 2\tau)} \tag{5b}$$

respectively.

While it is recognized that the technique can be applied to the embedding of analog signals, throughout this discussion we will assume that the embedded signal is digital, i.e. it assumes values taken from a $M-$ary set of symbols $d_i \in \{\pm 1, \pm 3, \ldots, \pm(2M - 1)\}$, for $i = 1, 2, 3, \ldots$ that are transmitted at the time instances $t = iT_s$, where T_s denotes the symbol period.

In one application, each symbol is associated with a corresponding value of the short-term autocorrelation function. Mapping symbols onto the domain of autocorrelation function values can be done in a number of ways. In order to make the difference signal small with respect to the existing signal, we employ the formula:

$$R_m(it_s, \tau) = \varepsilon d_i R_m(it_s, 0) \tag{6}$$

where ε is a small quantity selected so as to balance the requirements of robustness and transparency. By inserting formulas (4a) or (4b) into formula (6) we obtain a quadratic equation with respect to g, whose solution provides the appropriate gain g_i for the symbol transmitted at $t = iT_s$. Alternatively, an approximate value for g_i can be obtained using formulas (5a) or (5b). The gain is held constant at g_i over the symbol interval in order to minimize errors. Further deviation of g_i from its desired value can be employed at the boundaries of the symbol interval in order to avoid abrupt changes in the difference signal which can jeopardize the requirement of difference signal transparency. It has been found that the modulation error incurred by such smoothing does not significantly degrade the performance of the invention.

The integration interval, T, should be shorter than $T_s - \tau$ in order to minimize intersymbol interference. However, certain overlap between adjacent symbols can be tolerated in order to increase the bandwidth of the hidden channel. The embedded signal is retrieved from the stego signal, after transmission, recording, and processing that potentially degrades it, by the extractor (see Fig. 1). In the case where only a single autocorrelation delay is modulated, the extractor consists of a short-term autocorrelation generator followed by a data regenerator, as shown in Fig. 4.

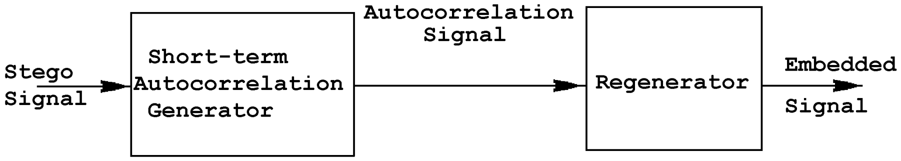

Fig. 4. Block diagram of the extractor

The short-term autocorrelation generator applies first a filter (see Fig. 5), which can be (but is not necessarily) the same as the filter applied in the difference signal component generator of the embeddor, and may be omitted entirely in some circumstances.

The short-time autocorrelation is computed using a delay, τ, corresponding to the amount of delay or advance used in the difference signal component generator of the embeddor. In parallel with the short-term autocorrelation of delay τ, the extractor also calculates the short-term autocorrelation with the delay zero, which is equivalent to calculating the square of the signal and integrating over the interval T. Further, the autocorrelation function generator calculates the output according to the formula:

$$d(t) = \frac{R_m(t,\tau)}{R_m(t,0)} \tag{7}$$

This output is called normalized autocorrelation signal. In a special case where binary data is transmitted as the embedded signal and the embedded information can be recovered as the sign of $R_m(t,\tau)$ at selected sampling instances, then it is unnecessary to calculate $R_m(t,0)$ and perform this normalization.

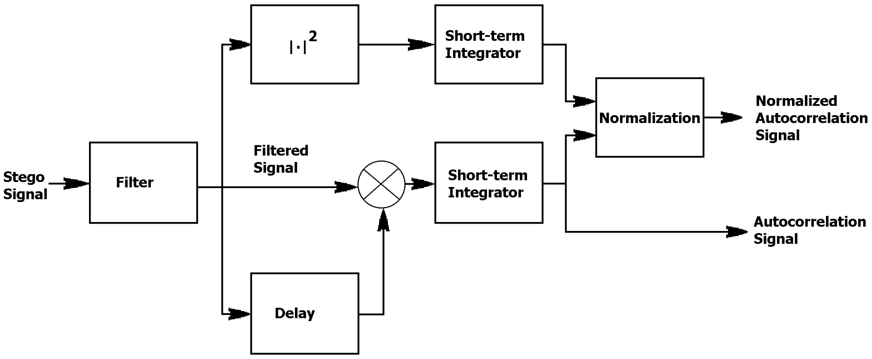

Fig. 5. Block diagram of the normalized short–term autocorrelation generator

The embedded signal is obtained from the (normalized) autocorrelation signal by the regenerator (Fig. 4). Embedded signal extraction may include one or more of filtering, equalization, synchronization, sampling, threshold comparison, and error control coding functions

In the absence of signal distortion, at discrete points in time separated by T_s, $d(t)$ takes on values that are directly proportional to the magnitude of input symbols. This feature of the encoded signal is the basis for the retrieval of the embedded signal.

3.1. Multi-level Symbol Mapping

In another application each embedded data symbol is associated with a set of short-term autocorrelations. The choice of the particular element of the set is done in such a way to minimize the value of g. To illustrate, we describe a method for transmitting a binary-valued embedded signal. In this case, the bit transmitted at time iT_s is associated with the set of short-term autocorrelation values $2j\varepsilon R_m(iT_s, 0)$, for $j = 0, \pm 1, \pm 2, \dots$, if its value is one, or the set $(2j-1)\varepsilon R_m(iT_s, 0)$, $j = 0, \pm 1, \pm 2, \dots$, if its value is zero. The value of j that is selected for each bit so as to minimize the magnitude of g obtained through the solution of equation (4a) or (4b). Alternatively, an approximate calculation can be done using equations (5a) and (5b). It is apparent that the smallest magnitude of g can be obtained from (5a) or (5b) if j is chosen so that $2j\varepsilon R_m(iT_s, 0)$ for one, or $(2j-1)\varepsilon R_m(iT_s, 0)$, for zero, is nearest to $R(t, \tau)$.

In this case, the extractor operates in the same way as for the previous, except that multiple autocorrelation function values are mapped to the same embedded channel symbol using the inverse of the mapping table used in the embeddor.

3.2. Manchester Symbol Encoding

In the third version of the system the embedded channel symbols are encoded as a difference in short-term autocorrelation functions at predefined time instants. For example, the symbol interval is divided in two equal parts and the autocorrelation function is determined for each. Then, the difference between the two autocorrelation functions is changed in such a way to represent embedded channel data. If the data symbol at the instant iT_s is $d_i \in \{\pm 1, \pm 3, \dots , \pm(2M-1)\}$, for $i = 1, 2, 3, \dots$, then the desired difference can be expressed by:

$$R_m(iT_s, \tau) - R_m((i+0.5)T_s, \tau)) = \varepsilon d_i R_m(iT_s, 0) \tag{8}$$

where ε is a small quantity determined in such a way to balance the robustness and transparency requirements stated above. By substituting equation (4a) or (4b) into equation (8), a quadratic equation is obtained with respect to g. By solving this equation we obtain the gain which is applied to the difference signal in the first half of the symbol interval. Gain equal in magnitude, but opposite in sign, is applied in the second half of the symbol interval. In order to minimize intersymbol interference the integration interval should be shorter than $(T_s/2) - \tau$. However, a small amount of intersymbol interference is acceptable and can increase the bit rate with which the embedded signal is encoded.

The extractor of the third version of the system calculates the (normalized) short-term autocorrelation functions in the same manner as in the first version. In the signal processing done in the extractor (Fig. 4), after filtering, equalization, and synchronization, the difference of the short-term autocorrelation functions is used to detect transmitted symbols.

3.3. Multiple Difference Signal Components

In the fourth case, the difference signal is comprised of the sum of a multiplicity of difference signal components. In this case, differing amounts of delay or advancement are employed in each of the difference signal component generators. Either the same embedded signal can be encoded in each of the difference signal components, or else a multiplicity of embedded signals can be encoded with selected difference signal components, under the restriction that for any two component generators which have equal amounts of delay and advancement, and appear in the same or overlapping frequency bands, time intervals and spatial masks, the embedded signals must be the same.

As with the other embodiments, the difference signal components can be of various forms, but in the preferred implementations we use delayed or advanced versions of existing signal itself, as expressed by the formula:

$$e(t) = \sum_{m=1}^{M} g_m s(t - \tau_m) \tag{9}$$

where τ_m, and g_m represent the delay and gain for the $m-$th difference signal component. It is understood that the negative delay represents an advance. The gain can be positive or negative, but should be much less than one in magnitude, in order to maintain the transparency requirement. By substituting (9) into (2) we obtain:

$$\begin{aligned} R_m(t,\tau) = R(t,\tau) + &\sum_{m=1}^{M} g_m (R(t, \tau_m + \tau) + R(t - \tau, \tau_m - \tau)) \\ &+ \sum_{m_1=1}^{M} \sum_{m_2=1}^{M} g_{m_1} g_{m_2} R(t - \tau_{m_1}, \tau + \tau_{m_2} - \tau_{m_1}) \end{aligned} \tag{10}$$

It is a well-known property of autocorrelation functions that for a fixed value of t, the maximum value is attained when the delay, τ, is equal to zero. For a random signal $s(t)$, and a sufficiently large τ, $R(t,\tau)$ is much smaller than $R(t, 0)$. Therefore, the set of delays $\{\tau_m\}$ should be chosen in such a way that $R_m(t, \tau)$ calculated for $\tau = \pm\tau_m$ according to formula (10) has only one term for which the short-term autocorrelation delay is equal to zero. This term will have dominant effect on the modulation of the $R_m(t, \tau_m)$. It is apparent that as different τ_m are chosen, different terms in formula (10) become dominant in the summation, effectively "tuning in" different difference components.

The extractor associated with this design includes a multiplicity of short-term autocorrelation generators, one associated with each advance or delay amount for which a difference signal component was generated in the embeddor. Each short-term autocorrelation generator may have a filter that may be the same or different from the other short-term autocorrelation generators, and which may or may not be the same as the filter that was employed in the corresponding difference signal component generator of the embeddor.

The autocorrelation signals obtained from the short-term autocorrelation generators are together processed by an embedded signal extraction device and either combined in order to obtain the original embedded signal, or independently processed in such a way as to extract a multiplicity of embedded signals. As in the first embodiment, the embedded signal extraction device may include one or more of filtering, masking, equalization, synchronization, sampling, threshold comparison, and error control coding functions.

Different difference components of the embedded channel can carry different embedded signals, to obtain an increase in overall signal throughput, or they can carry the same embedded signal to increase the robustness or security of the embedded signal transmission. In order to enhance the security of the embedded signal against eavesdropping, forgery, and erasure, multiple copies of the embedded signal are difference in such a way that no single one of them, nor small subset of them, is sufficient for reliable embedded signal retrieval. This condition is achieved by using sufficiently small values of the parameter ε.

Reliable recovery of the embedded channel information is achieved by combining information from the short-term autocorrelation values of all difference components. To do this, the extractor must know:

(a) the number of difference signal components,

(b) the autocorrelation delays associated with each component,

(c) the relative time shift between gain changes in the components (e.g. bit boundaries), and

(d) the filtering parameters associated with each component.

By keeping the values of the parameters associated with (a)-(d) secret, the embedded channel information affords a reasonable level of security against even an attacker with complete knowledge of the principles of autocorrelation modulation and with sophisticated expertise and tools for signal analysis.

It is understood in the existing art in this area that the security of the embedded signal can be further enhanced through the use of spread spectrum techniques. Fundamental to this approach is the modulation of the embedded signal with a secret spreading code prior to the encoding process. Such an approach can be used in conjunction with the application of this invention using multiple difference signal components and each component can be assigned a different spreading code. This increases enormously the number of different parameters that must be known for the presence or contents of the embedded signal to be detected, thus providing an even greater level of embedded signal security.

3.4. Delay Hopping

In the fifth variation of the proposal the auxiliary signal components change their autocorrelation delay over time according to a predefined pattern. This process will be called delay hopping. The hopping pattern can be defined as a list of consecutive autocorrelation delays and the duration of each of them. The pattern is kept secret in order to increase security of the embedded signal against the unauthorized eavesdropping, erasure, and forgery. An authorized decoder needs the knowledge of the hopping pattern, as well as the filtering parameters and signaling parameters (symbol duration and other symbol features). Multiple embedded signals can be carried simultaneously in the cover signal if their hopping patterns are distinct, even if other filtering and signaling parameters are the same.

The principles involved in the fifth embodiment of the invention are analogous to the principles of frequency hopping used in radio communication for secure data communications in a hostile environment [24]. Similar hopping rates, hopping sequences, synchronization techniques, and error correction codes can be used, and we will not elaborate them here.

Multiple signal components can be embedded into cover signal using the same hopping pattern, and still be separated at a decoder, providing that beginnings are sufficiently separated in time, and a suitable hopping pattern is used. The extractor will search for a match between the hopping pattern and its template. Once the match is found, the extractor will lock-on to the signal, and track further hops according to the predefined pattern. In order to detect a multitude of embedded signal components the decoder should be programmable to skip first n matches ($n = 0, 1, 2, \dots$), and keep searching on. This way the extractor could lock-on to a signal with a later onset, but the same hopping pattern.

The embeddor can be programmed to start embedding the data with a random delay. This would reduce the possibility that users licensed to operate the embeddor can intentionally damage each others signals using the overwrite process. It is understood that multiple overwrite process will deteriorate reading ability for any of the codes. The effect is analogous to multiple frequency hopping channels interfering with each other and the calculations of the maximum number of channels can be found elsewhere [24]. Specific to our application is the fact that multiple signal embedding is also limited by the transparency requirements. This limit is application dependent, and subject to experimental verification.

Signal parameters, such as number of delays, synchronization sequence length, and error correction code, should be chosen in such a way to withstand the interference of the maximum number of embedded signals determined by the transparency testing. Than the signal will be secure against an overwrite attack which does not cross the transparency threshold.

3.5. Further Enhancements and Modifications

There exist further enhancements and modifications to the proposed technique of which the authors are aware that may be beneficial in certain circumstances. All of them have been tested and used in appropriate applications.

- Improved performance can be obtained by adapting the value of the modulation parameter ε from symbol to symbol in each embedded signal component, in accordance with a measurement of the capacity of the existing signal to transparently withstand the insertion of the difference signal.
- For transmissions that are comprised of a multiplicity of signal channels (such as stereophonic audio), a variety of additional techniques are available for enhancing the performance of this invention. One such technique is that of "common mode rejection" (often referred to in the music undustry as "matrixing"). With this technique, rather than modulating the autocorrelation of the individual existing signals, one or more invertible linear combinations of those signals are obtained and the autocorrelation of the combined signal(s) are modulated. After modulation,

the linear combination is inverted, and the resulting encoded signals are transmitted to the extractor. At the extractor, the same linear combination is obtained, from which the embedded signals are extracted. Another technique available in multi-channel situations is where the same embedded signal is encoded in one or more difference signal components of selected signals. The extractor then extracts the embedded signal from each, combining the information to obtain the original message. With this technique, smaller difference signal components can be used in each channel while obtaining more robust transmission. There exist security benefits from this approach as well.

4. Performance Evaluation

The music industry has set the absolute transparency as a requirement for an embedded signaling system [11]. If any of the subjects in the testing process can distinguish the cover audio from the stego audio on any piece of music the watermarking technology is considered unacceptable. The testing process is very complex and time consuming. It requires utilization of highly talented and trained professionals, so called "golden ears", a large number of subjects in the listening panel (> 20), specially selected music material (which stresses the systems under the test), high quality recordings and reproduction devices, special reference listening rooms and carefully designed test methods.

Most of the tests are conducted using the so-called $A/B/X$ double-blind matching procedure. In this procedure track A represents the original music, track B is the stego signal, while track X is randomly chosen between A or B in each retrial. The embedding process fails the transparency test if a listener can make correct matching with a statistical significance (typically, the probability of achieving the result by a random guessing should be less than $p = 0.05$).

Alternatively, the so-called $A/B/C$ procedure is used, described in Rec. ITU-R BS.1116 as "double-blind triple- stimulus with hidden reference". The cover signal is always available as the stimulus A. The stimuli B and C are randomly assigned to the cover and stego signals depending on the trial. The subjects are asked to assess the impairments on "B" compared to "A", and "C" compared to "A" according to the continuous five-grade impairment scale. If 95 % confidence interval on any of the tests falls completely below the grade 5, the watermark audibility has been established.

Further, it is important to evaluate the watermark's robustness. The proposed technology should be robust to multiple D/A and A/D conversions including conversion between sample rates in the range 12 to 96 kHz and bits per sample 8 to 24, including consumer-grade converters in electrically noisy environment.

The embedded watermark should survive various signal processing techniques such as equalization, dynamic range compression, mono mixdown, stereo expansion, resampling, smoothing/enhancements, reverb, vibrato, noise gates etc. Those testing can be performed using various commercial software packages for audio signal processing, such as Sound Forge [37].

The watermark survival should be tested in various perceptual compression channels, such as MPEG-1 layers 1, 2, and 3, MPEG-2 LBR and AAC, ATRAC (adaptive transform coding), PASC (adaptive transform coding), Dolby AC-2 and AC-3, MPEG-4 AAC, MS audio, Liquid Audio, and Qdesign codec, at various bit rates. Fully transparent watermarking should survive down to 64 kb/s per stereo channel.

The watermark should survive also an additive white Gaussian noise at roughly 36 dB of SNR. Similarly, the watermark should survive voice-overs up to 50/50 split for music vs. voice.

The watermark should also be tested in various time scale modification cases. It should survive wow and flutter of 0.5 % at up to 100 Hz rate. Linear speed-up or slow-down of up to 10 % should be survivable as well. Further, the watermark should be tested against various commercial pitch-invariant time compression/expansion algorithms (e.g. using Sound Forge [37]). Typical pitch-invariant time compression algorithm consists of cutting out pieces of signal in such a way to minimize the disturbance around the cut. Typically the watermark should survive up to ±4 % of such a time scale variation.

A special problem is the evaluation of the watermark security. The objective is to find out if there is a procedure that can systematically eliminate the watermark without degrading the audio signal to an unacceptable level. Typical assumption is that the attacker has all the public information about the watermark algorithm, but not the stego key. The technology described herein offers some distinct features that give it advantage against the most popular alternative - the DSSS watermarks with respect to the attacks published in the literature.

Firstly, if an attacker obtains somehow the original and the watermarked signal, and makes a difference, it is "pure" watermark in the case of DSSS, which can be misused (e.g. after some adjustments can be inserted elsewhere). In our case the difference does not carry any information (that can be interpreted by an extractor); the information is hidden in the relation between the difference and the original.

An example where attacker can obtain the difference signal is a copy control system, where a commercial recorder inserts "no-more-copy" mark, or, more generally, a generational copy control mark. Once the difference signal is extracted, the attacker can invert the difference signal, insert it back into the original, and allow the recorder to cancel it with its own mark; the technique is suggested in [31]. This procedure is not applicable to our method, because the embeddor first calculates the natural autocorrelation, and then adds enough of difference signal to make autocorrelation reach its target level.

Similarly, the DSSS system is more vulnerable to time scale modifications, and related jitter attacks (see e.g. [23], [26]). This comes from the fact that DSSS extractor performs correlation between the incoming signal and a template, while our system performs the correlation between the signal and a delayed version of itself. When a time scale modification occurs, there is a loss of synchronization between the signal and the template for DSSS system, while the relation between signal and the delayed version of itself is only slightly disturbed.

5. Conclusion

This paper presents an overview of techniques used for data hiding within audio signal and presents a novel approach. The proposed approach is called the short-term autocorrelation modulation, and can be classified as a case of modulation of statistical properties of analog signals. The process is very simple, as inserting a delayed and/or advanced version of the signal itself can modify the autocorrelation. In order to optimize the process, we firstly calculate natural autocorrelation, and than determine necessary modification.

In the cases where natural autocorrelation varies in a wide range, it is possible to minimize inserted signal by introducing multiple autocorrelation levels assigned to each embedded data symbol. Alternatively, we can reduce autocorrelation variations by introducing Manchester encoding.

It is possible to introduce autocorrelation modulation on multiple delays in overlapping intervals in order to either increase capacity of the embedded channel, or its security. Introducing the delay hopping technique, where different delays are used for consecutive symbols according to a predefined stego key, further enhances the security.

Music industry has set high requirements for watermarking systems in terms of transparency, robustness, security, data capacity, and complexity. The proposed system offers a unique set of tradeoffs, and is expected to compete favorably with all other systems described in the literature.

REFERENCES

1. D. Kahn: *The History of Steganography.* Proceedings: Information Hiding, First International Workshop, Cambridge, U.K., pp. 1-5, 1996.
2. W. Bender et al.: *Techniques for Data Hiding.* IBM Systems Journal, Vol. 35, Nos. 3&4, pp. 313-336, 1996.
3. M.D. Swanson, M. Kobayashi, and A.H. Tewfik: *Multimedia Data-Embedding and Watermarking Technologies.* Proceedings of the IEEE, Vol. 86, No. 6, pp. 1064-1087, 1998.
4. H. Berghel, and L. O'Gorman: *Protecting Ownership Rights through Digital Watermarking.* IEEE Computer Mag., pp. 101-103, July 1996.
5. L. Boney, A. Tewfik, and K. Hamdy: *Digital Watermarks for Audio Signals.* Proceedings of Multimedia '96, Piscataway, NJ: IEEE Press, pp. 473- 480, 1996.
6. J. Brassil et al.: *Electronic Marking and Identification Techniques to Discourage Document Copying.* IEEE J. Select. Areas Commun., Vol. 13, pp. 1495-1504, Oct. 1995.
7. J. Tiliki and A. Beex: *Encoding a Hidden Digital Signature onto an Audio Signal Using Psychoacoustic Masking.* Proc. 7th Int. Conf. Sig. Proc. Appls. Tech., pp. 476-480, 1996.
8. J. Lacy et al.: *Intellectual Property Protection Systems and Digital Watermarking.* Proceedings: Information Hiding, Second International Workshop, Portland, Or, pp. 158-168, 1998.
9. C. Neubauer, J. Herre, and K. Brandenburg: *Continuous Steganographic Data Transmission Using Uncompressed Audio.* Proceedings: Information Hiding, Second International Workshop, Portland, OR, pp. 208-217, 1998.
10. J. Lacy, D.P. Maher, and J.H. Snyder: *Music on the Internet and the Intelectual Property Protection Problem.* Proc. International Symposium on Industrial Electronics, Guimares, Portugal, July 1997.
11. International Federation of the Phonograph Industry: *Request for Proposals – Embedded signaling systems.* issue 1.0. London, June 1997.
12. M. Cooperman and S. Moskowitz: *Steganographic Method and Device.* U.S. Patent 5,613,004, Mar. 1997 (Available WWW: `http://ww.digitalwatermark.com/patents.htm`)
13. I. Cox, et al.: *Secure Spread Spectrum Watermark for Multimedia.* IEEE Trans. Immage Processing, Vol. 6, no. 12, pp. 1673-1687, 1997.
14. R. Preuss et al.: *Embedded Signaling.* U.S. Patent 5,319,735, June 7, 1994.
15. C. U. Lee, K. Moallemi, and R. L. Warren: *Method and Apparatus for Transporting Auxiliary Data in Audio Signals.* U.S. Patent 5,822,360, Oct. 13, 1998.
16. M. Fardeau, et al.: *Method and Apparatus for Automatically Identifying a Program Including a Sound Signal.* U.S. Patent 5,581,800, Dec. 3 1996.

17. J. M. Jensen et al.: *Apparatus and Methods for Including Codes in Audio Signals and Decoding.* U.S. Patent 5,764,763, Jun. 9, 1998.

18. C. U. Lee, K. Moallemi, and J. Hinderling: *Post- compression Hidden Data Transport.* U.S. Patent 5,687,191, Nov. 11, 1997.

19. B. Chen and G. W. Wornell: *Digital Watermarking and Information Embedding Using Dither Modulation.* Proc. IEEE Workshop Multimedia Signal Processing, Redondo Beach, CA, Dec 1998.

20. B. Pfitzmann: *Information Hiding Terminology.* Proceedings: Information Hiding, First International Workshop, Cambridge, U.K., pp. 347-350, 1996.

21. J. Johnston and K. Brandenburg: *Wideband Coding – Perceptual Consideration for Speech and Music.* Advances in Speech Signal Processing, S. Furoi and M. Sondhi, Eds. New York: Marcel Dekker, 1992.

22. M. Bosi et al.: *IS 13818-7.* MPEG-2 Advanced Audio Coding, AAC

23. R. J. Anderson and F. A. P. Peticolas: *On the Limits of Steganography.* IEEE Journal on Selected Areas in Comm. (J-SAC), May 1998.

24. R.C. Dixon: *Spread Spectrum Systems with Commercial Applications.* 3rd Edition, John Wiley & Sons: New York, 1994.

25. K. N. Hamdy, A. H. Tewfik, T. Chen, and S. Takagi: *Time-scale modification of audio signals with combined harmonic and wavelet representations.* Proc. International Conference on Acoustics, Speech and Signal Processing–ICASSP '97, volume 1, pages 439- 442, Munich, Germany, April 1997.

26. F.A.P. Petitcolas, R.J. Anderson, and M.G. Kuhn: *Attacks on Copyright Marking Systems.* Proceedings: Information Hiding, Second International Workshop, Portland, Or, pp. 218-238, 1998.

27. S.J. Best, N. Johnson, and A.M. Sandford: *Signal identification system.* U.S. Patent No. 5,113,437. 1992.

28. S.J. Best, R.A. Willard, and E. Thorn: *Signal identification.* U.S. Patent No. 4,876,617. 1989.

29. D. Gruhl, A. Lu, and W. Bender: *Echo Hiding.* Proceedings: Information Hiding, First International Workshop, Cambridge, U.K., pp. 295-316, 1996.

30. S. Sowers and A. Yousef: *Testing Digital Watermark Resistance to Destruction.* Proceedings: Information Hiding, Second International Workshop, Portland, Or, pp. 239-257, 1998.

31. I.J. Cox and J.P.M.G Linnartz: *Public Watermarking and Resistance to Tampering.* Proc. IEEE Internat. Conf. On Immage Processing (ICIP97), Santa Barbara, CA, Vol. 3, Oct. 1997.

32. J.D. Johnston: *Transform Coding of Audio Signals Using Perceptual Noise Criteria.* IEEE J. Select. Areas Commun., Vol. 6, pp. 314-323, Feb. 1988.

33. K. Brandenburg and M. Bosi: *Overview of MPEG- audio: Current and Future Standards for Low-bit Rate Audio Coding.* J. Audio Eng. Soc., Vol. 45, No. 1/2, pp. 4-21, Jan./Feb. 1997.

34. J. Lacy et al.: *On Combining Watermarking with Perceptual Encoding.* Proceedings IEEE ICASSP, 1998.

35. A. Nagata et al.: *Method and Apparatus for Protection of Signal Copy.* US Patent 5,073,925, Dec. 1991.

36. C.U. Lee, K. Moallemi, and R.L. Waren: *Method and Apparatus for Transporting Auxiliary Data in Audio Signals.* US Patent 5,822,360, Oct. 1998.

37. Sound Forge Version 4.5a: *Copyright Sonic Foundry, Inc.*. www.sonicfoundry.com.