www.ijiarec.com

ISSN:2348-2079

Volume-5 Issue-2

International Journal of Intellectual Advancements and Research in Engineering Computations

Design of audio watermarking based on energy comparison technique implementation using internet of things

[1]SathishKumar.U.K, Leo.F.P, Dinesh Kumar.R,DurgaDevi.B,
Department of Electronics and Communication Engineering, Nandha College of Technology, Erode, India
Email: Leojohn110@gmail.com, dinuraju1601@gmail.com, durgabhojan04@gmail.com

***Abstract*—This paper introduces a new audio watermarking technique based on a perceptual kernel representation of audio signals (spikegram). Spikegram is a recent method to represent audio signals. It is combined with a dictionary of gammatones to construct a robust representation of sounds. In traditional phase embedding methods, the phase of coefficients of a given signal in a specific domain (such as Fourier domain) is modified. In the encoder of the proposed method (two-dictionary approach), signs and phases of gammatones in the spikegram are chosen adaptively to maximize the strength of the decoder. Moreover, the watermark is embedded only into kernels with high amplitudes where all masked gammatones have been already removed. The efficiency of the proposed spikegram watermarking is shown via several experimental results. First, robustness of the proposed method is shown against 32 kbps MP3 with an embedding rate of 56.5 bps. Second, we showed that the proposed method is robust against unified speech and audio codec (24 kbps USAC, linear predictive and Fourier domain modes) with an average payload of 5-15 bps. Third, it is robust against simulated small real room attacks with a payload of roughly 1 bps. Lastly, it is shown that the proposed method is robust against a variety of signal processing transforms while preserving quality.**

***Index Terms*—Copyright protection, Watermarking, Spikegram, Gammatone filter bank, Sparse representation,Multimedia security**

I. INTRODUCTION

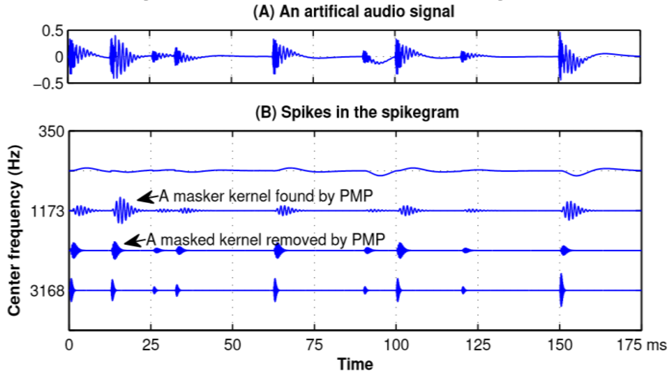
Every year global music piracy is making 12.5 billion of economic losses, 71060 U.S. jobs lost, a loss of 2.7 billion in workers' earnings and a loss of 422 million in tax revenues, 291 million in personal income tax and 131 million in lost corporate income and production taxes. Most of the music piracy is because of rapid growth and easiness of current technologies for copying, sharing, manipulating and distributing musical data [2].

As one promising solution, audio watermarking has been proposed for post-delivery protection of audio data. Digital watermarking works by embedding a hidden, inaudible watermark stream into the host audio signal. Generally, when the embedded data is easily removed by manipulation, the watermarking is said to be fragile which is suitable for authentication applications, whereas for copyright applications, the watermark needs to be robust against manipulations [3].

Watermarking has also many other applications such as copycontrol, broadcast monitoring and data annotation [3], [4], [5]. For audio watermarking, several approaches have been recently proposed in the literature. These approaches include audio watermarking using phase embedding techniques [6], cochlear delay [7], spatial masking and ambisonics [8], echo hiding [9], [10], [11], patchwork algorithm [12], wavelet transform [13], singular value decomposition [14] and FFT amplitude modification [15]. State of the art methods introduce phase changes in the signal representation (i.e., from the phase of the Fourier representation) [6], [16], while we adopt a more original strategy by using two dictionary of kernels and by shifting the sinusoidal term of the gammatones [17], [18]. In this paper, the watermarking is of multi-bit type [19] and could be used for data annotation.

Multiple dictionaries for sparse representation has already drawn the attention of researchers in signal processing [20], [21], [22], [23]. For example, in [20], a two-dictionary method is proposed for image inpainting where one decomposed image serves as the cartoon and the other as the texture image. Also, a watermark detection algorithm was proposed by Son et al. [21] for image watermarking where two dictionaries are learned for horizontally and vertically clustered dots in the half tone cells of images. In [23], authors propose an audio denoising algorithm using a sparse audio signal regression with a union of two dictionaries of modified discrete cosine transform (MDCT) bases. They use long window MDCT bases to model the tonal parts and short window MDCT bases to model the transient parts of the audio signals.

Two random dictionaries are used to improve the cryptographic security of spread spectrum (SS) image watermarking. In all mentioned methods, the goal is to



have an efficient representation of the signal. However for audio watermarking, one goal is to manipulate the signal representation in a way to find adaptively the spectro-temporal content of the signal for efficient transmission of watermark bits.

In this paper, we propose an embedding and decoding method for audio watermarking which jointly uses two type of gammatone dictionaries (including gammasinesandgammacosines) and a spikegram of the audio signal. It is shown in [24] that in comparison to block based representations, spikegram is time-shift invariant, where the signal is decomposed over a dictionary of gammatones. To generate the spikegram, we use the Perceptual Matching Pursuit (PMP) [25]. PMP is a bio-inspired approach that generates a sparse representation and takes into account the auditory masking at the output of a gammatone filter bank (the gammatone dictionary is obtained by duplicating the gammatone filter bank at different time samples).

Robustness against lossy perceptual codecs is a major requirement for a robust audio watermarking, thus we decided to evaluate the robustness of the method against 32 kps MP3 (although not used that often anymore, it is still a powerful attack which can be used as an evaluation tool).The proposed method is robust against 32 kbps MP3 compression with the average payload of 56.5 bps while the state of the art robust payload against this attack is lower than 50.3 bps [26]. In this paper, for the first time, we evaluate the robustness of the proposed method against USAC (Unified Speech and Audio Coding) [27], [28], [29]. USAC is a strong contemporary codec (high quality, low bit rate), with dual options both for audio and speech. USAC applies technologies such as spectral band replication, CELP codec and LPC.

Figure 1. A 2D plane of gammatone kernels of a spikegram generated from PMP [25] coefficients. The 2D plane is generated by repeating N_c= 4 gammatones at different channels (center frequencies) and at each time samples. A gammatone with non-zero coefficient is called a spike.

Experiments show that the proposed method is robust against USAC for the two modes of linear predictive domain (executed only for speech signals) and frequency domain (executed only for audio signals), with an average payload of 5-15 bps. The proposed method is also robust against simulated small real room attacks for the payload of roughly 1 bps. Lastly, the robustness against signal processing transforms such as resampling, re-quantization, low-pass filtering is evaluated and we observed that the quality of signals can be preserved.In this paper, the sampled version of any time domain signal is considered as a column vector with a bold face notation.

II. SPIKEGRAM KERNEL BASED REPRESENTATION

A. Definitions

With a sparse representation, a signal $x[n], n = 1:N$ (or x in vector format) is decomposed over a dictionary $\Phi = \{g_i[n]; n = 1:N, i = 1:M\}$ to render a sparse vector $\alpha = \{\alpha_i; i = 1:M\}$ which includes only a few non-zero coefficients, having the smallest reconstruction error for the host signal x [24], [25]. Hence,

$$x[n] \approx \sum_{i=1}^{M} \alpha_i g_i[n], \qquad n = 1,2,..,N \qquad (1)$$

where α_i is a sparse coefficient.

A 2D time-channel plane is generated by duplicating a bank of N_c gammatone filters (having respectively different center frequencies) on each time sample of the signal. Also, all the gammatone kernels in the mentioned 2D plane form the columns of the dictionary Φ (Hence, $M = N_c \times N$). Thus $g_i[n]$ is one base of the dictionary which is located at a point corresponding to channel $c_i \in \{1,..,N_c\}$, and time sample $\tau_i \in \{1,2,..,N\}$ inside the 2D time-channel plane (Fig.1). Thespikegram is the 2Dplot of the coefficients at different instants and channels (center frequencies). The number of non-zero coefficients in α_i per signal's length N is defined as the density of the representation (note that sparsity = 1-density).

To compute the sparse representation , many solutions have been presented in the literature including Iterative Thresholding Orthogonal Matching Pursuit (OMP) ,Alternating Direction Method (ADM) , Perceptual Matching Pursuit (PMP) [25]. Here, we use PMP for three different reasons: PMP is not computationally expensive, it is a high resolution representation for audio signals, and it generates auditory masking thresholds and removes the inaudible content under the masks [25].

PMP is a recent approach which solves the problem in (1) for audio and speech using a gammatone dictionary [25] PMP is a greedy method and an improvement over Matching Pursuit. PMP finds only audible kernels for which the sensation level is above an iteratively updated masking threshold and neglects the rest. A kernel is considered as a masked kernel if it is under the masking of (or close enough in time or channel to) another masker kernel with larger amplitude. The efficiency of PMP for signal representation is confirmed in [25] . The gammatone filter bank (used to generate the gammatone dictionary) is adapted to the natural sounds [24] and is shown to be efficient for sparse representation [25]. A gammatone kernel equation [17] has a gamma part and a tone part as below $g[n] = an^{m-1}e^{-2\pi ln}cos[2\pi(f_c/f_s)n + \theta], n = 1,..,\infty$ (2) in which, n is the time index, m and l are used for tuning the gamma part of the equation. $f_s$is the sampling frequency, θ is the phase, f_c is the center frequency of the gammatone. The term a is the normalization factor to set the energy of each gamatone to one. Also, the effective length of a gammatone is defined as the duration where the envelope is greater than one percent of the maximum value of the gammatone. In this paper, a 25-channel gammatone filter bank is used (Table I). Their bandwidths and center frequencies are fixed and chosen to correspond to 25 critical bands of hearing. They are implemented at the encoder and the decoder using (2). Also, a gammatone is called a gammacosine when $\theta = 0$ or a gammasine when $\theta = \pi/2$. In Table I, center frequencies and effective lengths for some gammatones, versus their channel numbers are given. In Fig.2, channel 8 gammasine and gammacosine are plotted.

Figure 2. A sample gammacosine (blue) and gammasine(red) (for channel-8) with a center frequency of 840 Hz and an effective length of 13.9 msec. Gammasines and gammacosines are chosen in the watermark embedding proceess based on their correlation with the host signal and the input watermark bit. The sampling frequency is 44.1 kHz.

B. Good characteristics of spikegram for audio watermarking

1) Time shift invariance: In most traditional watermarking techniques, the signal representation is block-based, where the signal is divided into overlapping blocks and watermark is inserted into each block. The conventional methods have two drawbacks. First, they might misrepresent the transients and periodicities in the signal. Moreover, in the block-based representation of nonstationary signals, small time shifts in the time domain signal might produce large changes in the representation, depending on the position of a particular acoustic event in each block [24]. The spikegram representation in (1) is time-shift invariant and is suitable for robust watermarking against time shifting de-synchronization attack.

2) Low host interference when using spikegram: In (1), many gammatones have either zero coefficients or are masked, thanks to PMP. Therefore, compared to traditional transforms such as STFT and Wavelet transforms, spikegram is expected to yield less host interference at the decoder.

3) Efficient embedding into robust coefficients: The watermark bits are inserted only into large amplitude coefficients obtained by PMP, where all inaudible gammatones have been a priori removed from the representation.

III. TWO-DICTIONARY APPROACH

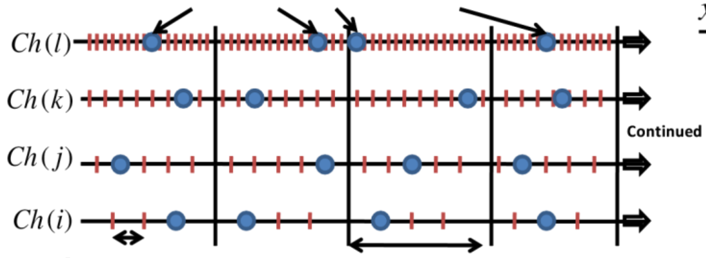
The watermark bit stream is symbolized by b which is an $M_2{\times}1$ vector ($M_2 < M$). The goal is to embed the watermark bit stream into the host signal. K, a $P \times 1$ vector ($P < M_2$), is the key which is shared between the encoder and the decoder of the watermarking system. Also, the sparse representation of the host signal x on the gammacosine dictionary (i.e., α_i) is assumed to be known.

The proposed method relies on the fact that the change in signal quality should not be perceived when changing the phase of specific gammatone kernels. Moreover, it is called a two dictionary approach, as a candidate kernel for watermark insertion, is adaptively selected from a gammacosine or gammasine dictionary.

For inserting multiple bits, the host signal $x[n]$ (x in vector format) is first represented using (1). Then, M_2 gammatones$g_k[n]$ from the representation in (1) are selected (the selection of watermark kernels is detailed in section III-D). These gammatones form the watermark dictionary D_1 and carry the watermark bit stream $b_k, k= 1,2,..,M_2$. Other $M_1 = M - M_2$ kernels form the signal dictionary D_2. The signal and watermark dictionaries are disjoint subsets of the gammatone dictionary used for sparse representation in (1), thus $D_1 \cap D_2 = \emptyset$. Each watermark bit $b_k$serves as the sign of a watermark kernel. Hence (1) becomes

$$y[n] = \sum^{M1} \alpha_i g_i[n] + \sum^{M2} b_k|\alpha_k| g_k[n] \qquad (3)$$

i=1 k=1 where $y[n]$ is the watermarked signal. In (3), if the watermark and signal dictionaries use the same gammatone kernels, the watermarking becomes



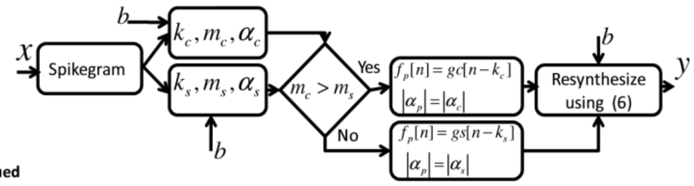
a one dictionary

Figure 3. Watermark insertion using the two-dictionary method. First, the spikegram of the host signal is found using PMP with a dictionary of 25channel gammacosines, located at each time sample along the time axis. Then for each processing window and each channel and based on the embedding bit b, the gammacosine, or gammasine (located at a blue circle) with maximum strength factor (m_c or m_s) is chosen for the watermark insertion. In this work, gammatone channels Ch^0s are selected in the range of 1-4 and 919 (odd channels only) for the watermark insertion. Also, to get the same embedding strength for different embedding channels, processing windows of different channels have the same length.

In one dictionary method, the watermark bits are inserted as the sign of gammatone kernels. In two dictionary method, in addition to the manipulation of the sign of gammatone kernels, their phase also might be shifted as much as $\pi/2$, based on the strength of the decoder. Hence, for the two-dictionary approach, each watermark kernel is chosen adaptively from a union of two dictionaries, one dictionary of gammacosines and one dictionary of gammasines.

The k^{th}watermark kernel in the watermark dictionary is found adaptively and symbolized with $f_k$which is either a gammasine or a gammacosine. Thus for the two dictionary method, the embedding equation in (3) becomesTo decode of the p^{th}watermark bit, we compute the projections of the watermarked signal on the p^{th}watermark kernel.

The number of samples used to compute the projection in (5) is equal to the gammatone effective length. The goal is to decode the watermark bit as the sign of the projection $<y,f_p>$. We later show how to find the best watermarkkernels so that the first two terms in the right side of (5) have the same signs as the watermark bit b_p. There are two sources of interference in (5). First, the right term in the right side of (5) is the interference that the decoder receives from other watermark bit insertions. To remove this interference term, the watermark insertion is performed into limited number of channels so that the watermark gammatones are uncorrelated. In fact, to design the watermark dictionary, we choose a subset



of the full overcomplete dictionary in such a way that the watermark kernels are spectro-temporally far enough such that they are uncorrelated. Thus the watermark bits will be decoded independently. Hence, in Fig. 3, for each channel and time sample, two neighbor watermark kernels should be separated with at least one effective length and at least one channel. With this assumption, the correlation between watermark gammatones will be less than 0.02. The second source of interference is the left term in the right side of (5) which originates from the correlations between watermark and signal gammatones, that is shown in (7). We reduce this interference in the encoder of the system in the next section, by adaptively searching for and embedding into the strongest watermark gammatones in the spikegram.

As embedding of multiple watermark bits are performed independently, thus in the next section, only the single bit watermarking using the two dictionary method is explained.

A. The proposed informed embedder

Equation (1) is used to resynthesize the host signal x from sparse coefficients and gammacosines.

Now, we want to embed one bit $b \in \{-1,1\}$ from the watermark bit stream b by changing the sign and/or the phase of a gammacosine kernel g_p(the p^{th}kernel found by PMP, still to be determined later in this section) with amplitude α_p (to be determined) located at a given channel and processing window (each processing window is a time frame including several effective lengths of a gammatone, Fig.3).

To find an efficient watermark kernel $f_p$which bears the

greatest decoding performance for the watermark b, we write the 1-bit embedding equation as follows:

$$y[n] = \sum_{i=1, i\neq p}^{M} \alpha_i g_i[n] + b|\alpha_p| f_p[n] \quad (6)$$

where the watermarked kernel $f_p$for a given channel number can be a gammacosine (gc) or a gammasine (gs) which are zero and $\pi/2$ phase-shifted versions of the original gammatone kernel g_p, respectively. The correlation between the watermarked signal y and the watermarked kernel f_p, is found as below

Hence, to design a simple correlation-based decoder, the sign of the correlation in the left side of (7) is considered as decoded the watermark bit. In this case, for correct detection of the watermark bit b, the interference term should not change the desired sign at the right hand side of (7). Moreover, the gammatone dictionary is not orthogonal, hence the left term in the right side of (7) may cause erroneous detection of b. For a strong decoder, two terms on the right side of (7), should have the same sign with large values. We later show that by finding an appropriate gammacosine or gammasine in the spikegram, the right side of (7) can have the same sign as the watermark bit b. In this case, the module of correlation in (7) is called watermark strength factor m_p for the bit b and a greater strength factor means a stronger watermark bit against attacks. In this case, (7) becomes

$$<y,f_p>= bm_p \tag{8}$$

For a large value strength factor (and with the same sign of the watermark bit), we search the peak value of the projections using (7) when a gammatone candidate is a gammacosine or gammasine. Thus, for a given channel, a processing window and watermark bit b, the signal interference is minimized at the decoder using the informed encoder in (7). We do the following procedure to find the phase, position and the amplitude of the watermarked kernel f_p (Fig. 4).

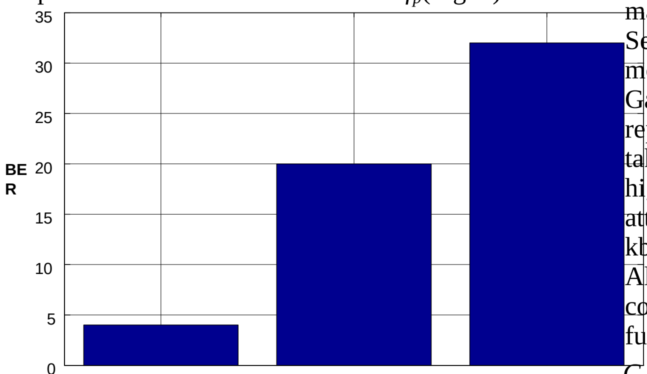


Figure 4. The proposed embedder for a given channel and processing window. The gammasine or gammacosine with maximum strength factor is chosen as the watermark kernel and its amplitude is set to its associated sparse coefficient in the spikegram. Finally (6) is used to resynthesize the watermarked signal y (in vector format). m_s and m_c are respectively the strength factors for gammasine candidate and gammacosine candidate.

In the given channel, we consider the watermark gammatone candidate f_p (the p^{th} gammatone kernel in the signal representation of (1)) to be a gammacosine gc or a gammasine gs. Then, do the following steps:

- Shift the watermark gammatone candidate f_p alongside all processing windows, at time shifts equal to multiples of the gammatones' effective length. For each shift compute the correlation of the watermarked signal with the sliding watermark candidate kernel. Then, find the absolute maximum of the correlation (watermark strength factor) $|<\boldsymbol{y},\boldsymbol{f}_p>|$ using (7) (Fig.3). The result is a strength factor, symbolized as m_c for gammacosine, located at time sample k_c with amplitude α_c and also another strength factor, symbolized as m_s for a gammasine kernel located at k_s with the amplitude α_s. Thus $m_c = |<y,gc[n-k_c]>|$, $m_s= |<y,gs[n-k_s]>|$.
- Afterwards, the gammacosine or gammasine with greater strength factor is chosen as the final watermark gammatone f_p and its time shift (sample), amplitude and phase are registered. Gammatone or gammasine with greater strength factor is chosen as the final watermark gammatone f_p with the final watermark strength factor being $m_t= max(m_c,m_s)$. The respective k_c or k_s, amplitude α_c or α_s and phases are kept. Therefore, the algorithm finds the optimal watermark gamatone from two dictionaries including one dictionary of gammacosines and one dictionary of gammasines.

It is called two-dictionary approach. The encoder and the decoder search in a correlation space to find the maximum projection (minimum signal interference). Second, the proposed approach is a phase embedding method on gammatone kernels with uses of masking. Gammatone kernels are the building blocks to represent the audio signal. Third, the proposed method takes care of efficient embedding into non-masked, high value coefficients which make it robust against attacks such as universal speech and audio codec (24 kbps USAC) [29] and 32 kbps MP3 compression. Also, thanks to the use of PMP, by removing many coefficients under the masks, the signal interference is further reduced at the decoder.

G. Robustness against analogue hole experiments

Here, the robustness of the proposed method against analogue hole is evaluated in a preliminary experiment. The BER of the proposed method against a simulated real room are given using the image source method for modeling the room impulse response (RIR) . We embed one bit of watermark in each second of the host signal (1 bps payload). We use an open source MATLAB code to simulate the room impulse responses. A cascade of RIR of a $4m \times 4m \times 4m$ room with a 20 dB additive white Gaussian noise is considered as the simulated room impulse response. Also, only one microphone and loud speaker are modeled. The experiments are done for three distances d between the loudspeaker and the microphone including d = 1,2and 3

The d meters (d denotes the distance between the microphone and the speaker). For watermark embedding, all the bits in each 1-second frames are generated using a pseudo random number generator. A spread spectrum (SS) correlation decoder is used. Hence, the 1-second sliding window is shifted sample by sample until the correlation of the SS decoder is above 0.75. Then, the watermark bit is decoded as the sign of the SS correlation.

VI. CONCLUSION

A new technique based on a spikegram representation of the acoustical signal and on the use of two dictionaries was proposed. Gammatone kernels along with perceptual matching pursuit are used for spikegram representation. To achieve the highest robustness, the encoder selects the best kernels that will provide the maximum strength factors at the decoder and embeds the watermark bits into the phase of the found kernels. Results show better performance of the proposed method against 32 kbps MP3 compression with a robust payload of56.5 bps compared to several recent techniques. Furthermore, for the first time, we report robustness result against USAC (unified speech and audio coding) which uses a new standard for speech and audio coding. It is observed that the BER is still smaller than 5% for a payload comprised between 5 and 15 bps. The approach is versatile for a large range of applications thanks to the adaptive nature of the algorithm (adaptive perceptive masking and adaptive selection of the kernels) and to the combination with well established algorithms coming from the watermarking community. It has fair performance when compared with the state of the art. The research in this area is still in its infancy (spikegrams for watermarking) and there is plenty of room for improvements in future works. Moreover, we showed that the approach can be used for realtime watermark decoding thanks to the use of a projectioncorrelation based decoder. In addition, two-dictionary method could be investigated for image watermarking.

REFERENCES

[1] YousofErfani, RaminPichevar, Jean Rouat, “Audio watermarking using spikegram and a two dictionary approach, Vol 2

[2] I. Cox, M. Miller, J. Bloom, J. Fridrich and T. Kalker, “Digital Watermarking and Steganography”, San Francisco, USA: Morgan Kaufmann Publishers Inc., 2nd ed., 2007.

[3] M. Steinebach and J. Dittmann, “Watermarking-based digital audio data authentication”, *Eurasip J. Appl. Signal Process.*, pp.1001-1015, 2003.

[4] A. Boho, G. Van Wallendael, A. Dooms, J. De Cock, et al., “End-ToEnd Security for Video Distribution”, IEEE Signal Processing Magazine, vol.30, no.2, pp.97-107, 2013.

[5] S. Majumder, K.J. Devi, S.K. Sarkar, “Singular value decomposition and wavelet-based iris biometric watermarking”, IET Biometrics, vol.2, no.1, pp.21-27, 2013.

[6] M. Arnold, X. Chen, P. Baum, U. Gries, and G. Dorr, “A phase-based audio watermarking system robust to acoustic path propagation”, IEEE Trans. on IFS, vol.9, no.3, pp.411-425, 2014.

[7] M. Unoki, R. Miyauchi, “Robust, blindly-detectable, and semi-reversible technique of audio watermarking based on cochlear delay”, IEICE Trans. on Inf. Syst. vol.E98-D, no.1, pp.38-48, 2015.

[8] R. Nishimura, “Audio watermarking using spatial masking and ambisonics”, IEEE Trans. on ASLP, vol.20, no.9, pp.2461-2469, 2012.

[9] G. Hua, J. Goh, and V. L. L. Thing, “Time-spread echo-based audio watermarking with optimized imperceptibility and robustness”, IEEE Trans. ASLP, vol.23, no.2, pp.227-239, 2015.

[10] G. Hua, J. Goh, and V. L. L. Thing, “Cepstral analysis for the application of echo-based audio watermark detection”, IEEE Trans. on IFS, vol.10, no.9, pp.1850-1861, 2015.

[11] Y. Xiang, I. Natgunanathan, D. Peng, W. Zhou, S. Yu, “A dual-channel time-spread echo method for audio watermarking”, IEEE Trans. IFS, vol.7, no.2, pp. 383-392, 2012.

[12] Y. Xiang, I. Natgunanathan, S. Guo, W. Zhou, and S. Nahavandi, “Patchwork-based audio watermarking method robust to desynchronization attacks”, IEEE Trans. ASLP, vol.22, no.9, pp.1413-1423, 2014.

[13] C. M. Pun and X. C. Yuan, “Robust segments detector for desynchronization resilient audio watermarking”, IEEE Trans. ASLP., vol.21, no.11, pp. 2412-2424, 2013.

[14] B. Lei, I. Y. Soon, and E. L. Tan, “Robust SVD-based audio watermarking scheme with differential evolution optimization”, IEEE Trans. ASLP, vol.21, no.11, pp.2368-2377, 2013.

[15] D. Megas, J. Serra-Ruiz, M. Fallahpour, “Efficient self-synchronised blind audio watermarking system based on time domain and FFT amplitude modification”, Signal Processing, vol.90, no.12, pp.3078-3092, 2010.

[16] N. M. Ngo, M. Unoki, “Robust and reliable audio watermarking based on phase coding”, IEEE ICASSP, pp.345-349, 2015.

[17] R.D. Patterson, B.C.J. Moore, “Auditory filters and excitation patterns as representations of frequency resolution”, Academic Press Ltd., Frequency Selectivity in Hearing, London, pp.123-177, 1987.

[18] M. Slaney, “An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank”, Apple Computer Technical Report 35, 1993.

[19] N. Nikolaidis, I. Pitas,“ Benchmarking of Watermarking Algorithms”, in Book: Intelligent Watermarking Techniques, World Scientific Press, pp. 315-347, 2004.

[20] S.M. Valiollahzadeh, M. Nazari, M. Babaie-Zadeh, C. Jutten, “A new approach in decomposition over multiple-overcomplete dictionaries with application to image inpainting”, Machine Learning for Signal Processing, IEEE MLSP2009, pp.1-6, 2009.

[21] Ch. H. Son, H. Choo, “Watermark detection from clustered halftone dots via learned dictionary”, Signal Processing, vol.102, pp.77-84, 2014. [22] A. Adler., V. Emiya, M.G. Jafari, M. Elad, R. Gribonval, M.D. Plumbley, “Audio Inpainting”, IEEE Trans. ASLP, vol.20, no.3, pp.922-932, 2012.

[23] C. Fevotte, L. Daudet, S.J. Godsill, B. Torresani, “Sparse Regression with Structured Priors: Application to Audio Denoising”, IEEE ICASSP, pp.57-60, 2006.

[24] E. Smith, M. S. Lewicki, “Efficient Coding of Time-Relative Structure Using Spikes”, Neural Computation, vol.17 , no.1 pp.19-45, 2005.

[25] R. Pichevar, H. Najaf-Zadeh, L. Thibault, H. Lahdili, “Auditory-inspired sparse representation of audio signals”, Speech Communication, vol.53, no.5, pp.643-657, 2011.

[26] K. Khaldi, A.O. Boudraa, “Audio Watermarking Via EMD”, IEEE Trans. ASLP, vol.21, no.3, pp.675-680, 2013.

[27] S. Quackenbush, "MPEG Unified Speech and Audio Coding", IEEE MultiMedia, vol.20, no.2, pp. 72-78, 2013.

[28] Y. Yamamoto, T. Chinen and M. Nishiguchi, "A new bandwidth extension technology for MPEG Unified Speech and Audio Coding", 2013 IEEE ICASSP, pp.523-527, 2013.

[29] M. Neuendorf, P. Gournay, M. Multrus, J. Lecomte, B. Bessette, R. Geige, S. Bayer, G. Fuchs, J. Hilpert, N. Rettelbach, R. salami, G. Schuller, R. Lefebvre, B. Grill, "Unified speech and audio coding scheme for high quality at low bit rates", IEEE ICASSP, pp.1-4, 2009.