Paper

Robust Audio Watermarks in Frequency Domain

Przemysław Dymarski and Robert Markiewicz

Faculty of Electronics and Information Technology, Warsaw University of Technology, Warsaw, Poland

Abstract—In this paper an audio watermarking technique is presented, using log-spectrum, dirty paper codes and LDPC for watermark embedding. This technique may be used as a digital communication channel, transmitting data at about 40 b/s. It may be also applied for hiding a digital signature, e.g., for copyright protection purposes. Robustness of the watermarks against audio signal compression, resampling and transmitting through an acoustic channel is tested.

Keywords—annotation watermarking, audio watermarking, digital signature, dirty paper codes, LDPC.

1. Introduction

Audio watermarking techniques are still being developed, due to many applications, e.g., copyright protection, voice messages authentication, annotation of audio files, etc. [1]. In most applications blind watermarking is used, i.e., watermark reception is effectuated without knowing the original audio signal (so called cover signal). For annotation watermarking, e.g., the song lyrics or the singers names encoding, less than 100 b/s bit stream is needed [2]–[4]. Sometimes real time decoding is required. For copyright protection, the recording owner's logo may be encoded in a several bits per second bit stream [5]. It may be also hidden in an audio file as a several hundreds of bits frame [6], [7]. In this case it is difficult to attribute any bit rate to the watermark. Such watermarking system is rather characterized by its capacity, i.e., different digital signatures number (logos) being recognized without mistakes. The same concerns watermarks used for identification of broadcast stations, certain kinds of emissions (e.g. commercials), and watermarks used for authentication of spoken messages. In such systems an increase of capacity is a main issue [8], [9].

In this paper an audio watermarking method in frequency domain is described. Two variants are presented. The first one consists in transmitting a bit stream with bitrate about 40 b/s, which is sufficient for most annotation watermarking applications. The second method lies in a digital signature in an audio signal embedding. This signature is difficult to detect and remove by an unauthorized person. The authors have presented some watermarking algorithms in conference proceedings [10]–[14], but the complete systems are described in this paper, together with a thorough characterization.

In Section 2 requirements for audio watermarks and digital signatures are discussed. In Section 3 the first variant of proposed method is described, namely the digital transmission in log-spectrum domain using dirty paper codes and LDPC. In Section 4 the second variant is presented, i.e., digital signatures embedding and detection. Section 5 is devoted to testing the robustness against audio compression, resampling and transmitting in acoustic channel. Short summary follows in Section 6.

2. Requirements for Audio Watermarks and Digital Signatures

First of all, watermark should not affect the quality of audio signal. Particularly it concerns music, where the watermark should be inaudible. According to the recommendations of International Federation of Phonographic Industry (IFPI), signal to watermark ratio should be greater than 20 dB. Some researchers follow these recommendations [7], but according to authors, they are not enough restrictive and yield audible distortions. The 6 musical recordings, i.e., songs, piano with orchestra, violin and trumpet, sampled at 44100 Hz was tested, with embedding the hidden bit stream (Fig. 1a) and the digital signature (Fig. 1b). The watermark spectrum was fit to the masking curve calculated using the MPEG-Audio algorithm [15], but its attenuation was adjusted using the offset parameter. Two objective audio quality criteria were evaluated: the segmental signal-to-watermark ratio (SNRseg) and the Objective Difference Grade (ODG) calculated with the PEAQ algorithm [16]. ODG values around zero indicate that the original and watermarked files are perceptually equivalent. ODG value around –1 indicates audible, but not annoying distortions.

Results shown in Fig. 1 indicate a substantial drop of quality of some musical recordings even at SNRseg = 30 dB. Informal listening tests confirm these results. The target is to keep the watermark inaudible, so ODG $>$ –0.2 was chosen as a condition of inaudibility. In order to fulfill this condition for all tested signals, watermark attenuation (offset) should be set at 20 dB. This is a very restrictive constraint, because for many audio signals watermark may be much stronger. Thus it is possible to adapt the offset value, adjusting it to the audio signal. Note that in watermarking bit stream transmission task the synchronization signal is also embedded, so the watermark itself should be weaker than in the digital signature transmission task.

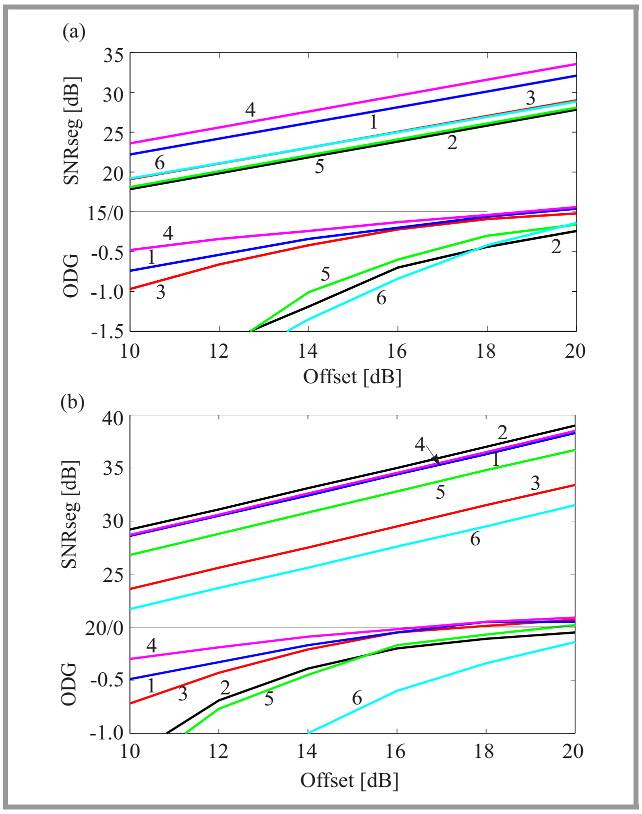

Fig. 1. Signal to watermark ratio (SNRseg) and ODG as functions of the watermark attenuation for 6 audio files: 1, 4 – songs; 2, 5 – piano with orchestra; 3 – violin; 6 – trumpet. Picture (a) concerns watermarking bit stream transmission, (b) with hiding of digital signature.

In watermarking of speech signal (e.g. for authentication of spoken messages), the quality issue is not as important as for musical signals. Slight distortions are allowed but they should not be annoying.

In any case the watermark must be robust against compression of audio signal, i.e., MPEG-Audio in case of musical signals, speech coders such as GSM-EFR in case of telephonic speech. The compression itself should not affect much the signal quality. Take for example the compression

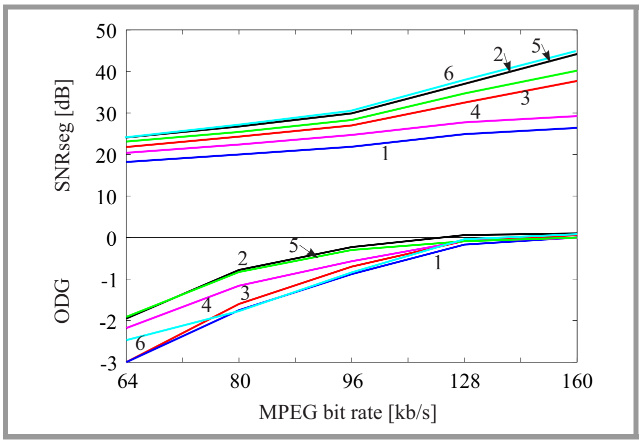

Fig. 2. SNRseg and ODG as functions of bit rate of the MPEG1-Audio codec for the same 6 audio files.

of musical signals using the MPEG-Audio coder. The objective quality measures (SNRseg and ODG) drop as the bit rate of the MPEG-Audio coder decreases (Fig. 2). Below 128 kb/s distortions become audible, and at 96 kb/s they may be annoying. Copyright protection of such distorted recordings makes no sense, because they have no commercial value. So the authors demand that the watermark be robust against the MPEG-Audio compression at 128 kb/s for watermarking bit rate transmission task and at 96 kb/s for digital signature embedding task.

Watermark should be also robust against resampling of the audio signal [2]. In commercial broadcast resampling is used for make a song shorter or longer in order to fill in the time slot of a predefined duration. If the watermarked audio is to be transmitted in analog form, i.e., FM analog broadcast or acoustic channel [5], [17], then resampling is due to slightly different sampling frequencies in DAC and ADC converters. Difference of both frequencies may attain few dozen of Hz.

In some applications, i.e., copyright protection, authentication of spoken messages, watermark should be robust against malicious attacks of persons aiming at its removal or modification. First of all it should be hard to detect for an unauthorized person using conventional signal processing tools. In this context watermarking algorithms based on embedding of tones should be discussed. These algorithms are quite robust against compression, filtering and digital-to-analog and analog-to-digital conversion [8], [9]. Tones, hidden below the masking threshold, are not audible, but they are easily detected using Discrete Fourier Transform (DFT). Moreover, they may be removed using a stop-band filter. They may be also maliciously inserted to the audio signal, in order to compromise the message authentication system.

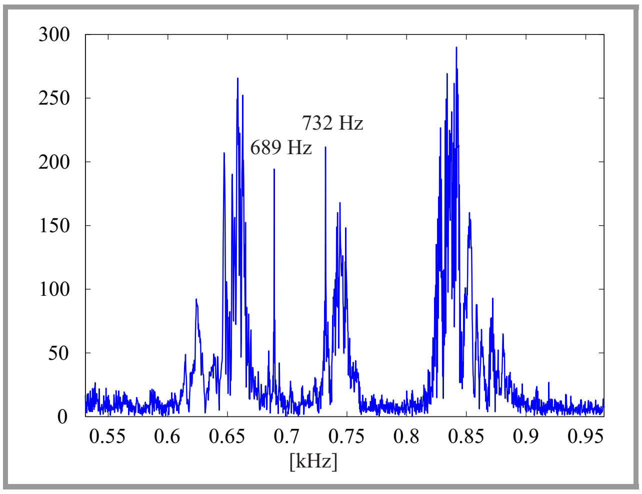

Fig. 3. Pilot tones (689 and 732 Hz) detected using DFT, window duration 2 s.

Take for example the symbol synchronization algorithm described in [11]. Two pilot tones (689 and 732 Hz), embedded below the masking threshold, are used for symbol localization in time domain and for sampling frequency offset detection. These tones may be easily detected using DFT and removed (Fig. 3).

This issue is less important in annotation watermarking, where some surplus information (e.g. lyrics of a song) is embedded in the audio signal. Nobody is interested in removing such information. Therefore, these tones were used for symbol synchronization if a bit stream is transmitted, but embedding them in digital signatures was avoided.

3. Watermark as a Bit Stream

3.1. Overview of the System

Embedding watermark in frequency domain yields better robustness against imperfect synchronization and resampling as compared with time domain watermark embedding [10]. Therefore, this research is focused on frequency domain watermarking. Firstly, the low bit rate (about 2 b/s) watermarking scheme, firstly described in [5], was adapted to annotation watermarking purposes [12]. In [11] the modified embedding and detecting algorithms were presented to improve robustness against time shifts of transmitted watermark symbols. In connection with [11] a synchronization algorithm based on transmission of two masked tones is proposed. In the annotation watermarking, the audio signal is known at the transmitter but not known at the receiver – this is a channel coding problem in presence of side information. It is solved with the informed (dirty paper) coding [18], consisting in using many symbols for transmission of the same message. The other approach is the informed watermark embedding, which tends to adapt the watermark to the audio signal by simulating the watermark reception at the transmitter's side [19]. Both approaches were adapted to proposed annotation watermarking system [13], transmitting the inaudible watermarks at the bit rate of 43 b/s with about 1% of errors caused by MPEG-Audio compression at 128 kb/s.

Recently the authors have presented an improved audio watermarking system transmitting inaudible data at 43 b/s. It is robust against MPEG compression at 96 kb/s (BER $< 10^{-4}$). This is achieved by using M-ary orthogonal codes, proposed in [20] and applied in [3] in time domain and the low-density parity-check code (LDPC) [21]. In order to obtain the orthogonal patterns in log-spectrum domain, Walsh functions are used [22].

3.2. Watermark Embedding and Detection

In order to keep a comparable watermark strength over the frequency axis watermark embedding and detection in log-spectrum domain is performed. The log-spectrum is increased or decreased in frequency subbands according to sign of a kernel, in presented case one of the Walsh functions w_i. The watermark is embedded in the band 1030–6550 Hz (Fig. 4), lower frequencies are reserved for transmission of the synchronizing tones [11]. Higher frequencies are not used because they may be suppressed in lossy compression of the watermarked audio (e.g. MPEG-Audio).

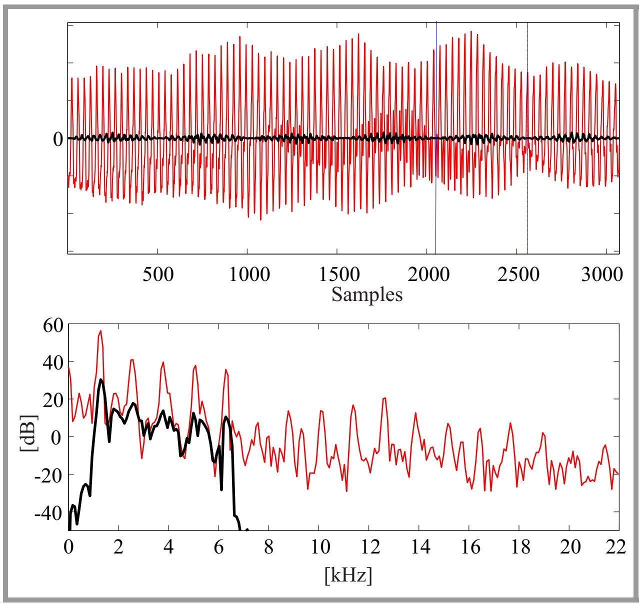

Fig. 4. The audio signal and the watermark (thick line) in time and frequency domain.

In order to reduce the strong spectral peaks influence of the audio signal on watermark detection, a difference of log-spectra in the neighbor windows is used as the cover signal [5], [12]. This approach is referred to as differential coding, mirrored kernels or Manchester signaling. Thus the time slot used for one watermark symbol transmission (i.e. 1024 samples at the sampling frequency 44.1 kHz) is split into two subintervals (subframes, vertical lines in Fig. 4). In each subframe a kernel of opposite sign is used: $-w_i$, $+w_i$.

The watermark embedding algorithm is shown in Fig. 5. Firstly the log-spectra of the audio signal are calculated in two subframes and subtracted, thus forming the cover signal (Fig. 5a):

$$\Delta \log |\overline{X}| = \Delta \log |\overline{X}^2| - \Delta \log |\overline{X}^1| \,. \tag{1}$$

Then the masking threshold is calculated in each subframe and signal to mask ratio is obtained in linear amplitude scale, as a function of frequency k: smr_k^1, smr_k^2. It means that the k-th component of the amplitude spectrum $|\overline{X}^2|$, i.e. $|\overline{X}_k^2|$, may be increased or decreased by $|X_k^2|/smr_k^2$ without violating the masking threshold (the same concerns $|X_k^1|$). In the other words, $|X_k^2|$ may be multiplied by any number greater than $1-1/smr_k^2$ and less than $1+1/smr_k^2$. These maximal modifications of the amplitude spectrum are expressed in decibels and denoted $\Delta^- \log|X_k^2|$ and $\Delta^+ \log|X_k^2|$. They are also calculated for masked spectrum components ($smr_k < 1$). In this case modifications must not exceed the masking threshold and they are limited to 10 dB. Note that according to the Eq. (1), the watermark is embedded in the difference of two log-spectra. Its strength is thus bounded

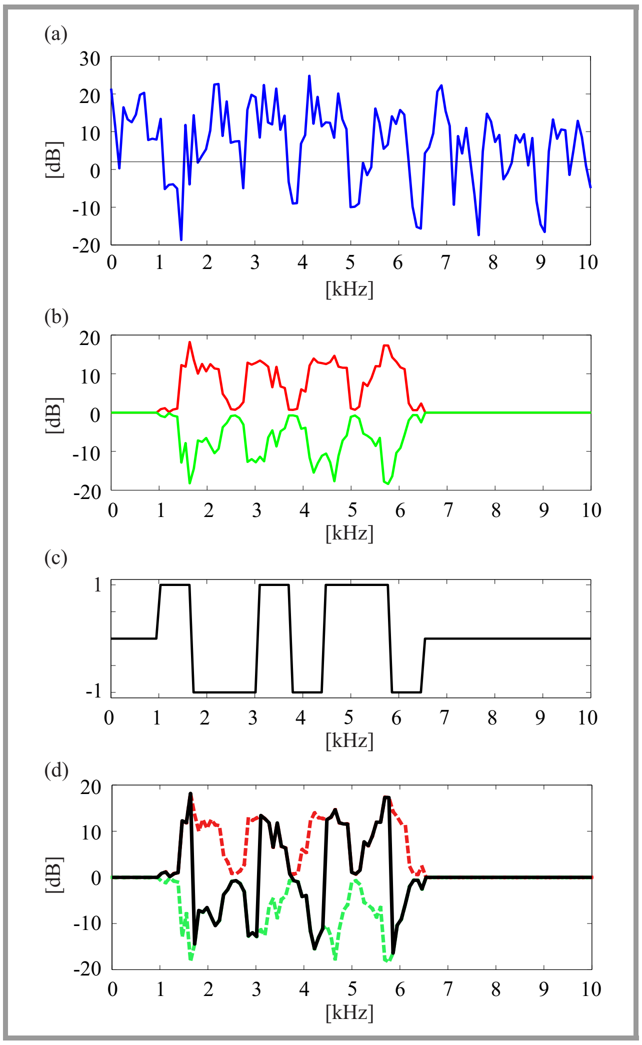

Fig. 5. The scheme of watermark embedding: (a) the cover signal $\Delta\log|\overline{X}|$, (b) maximal changes within the masking threshold $\Delta^{+}\log|\overline{X}|$ and $\Delta^{-}\log|\overline{X}|$, (c) the 6th Walsh function, (d) absolute value of the watermark in log-spectrum domain.

with the following values (frequency index k is skipped for simplicity) – see Fig. 5b:

$$\begin{aligned} \Delta^{+}\log|\overline{X}| &= \Delta^{+}\log|\overline{X}^{2}| - \Delta^{-}\log|\overline{X}^{1}| \quad \text{and} \\ \Delta^{-}\log|\overline{X}| &= \Delta^{-}\log|\overline{X}^{2}| - \Delta^{+}\log|\overline{X}^{1}|\,. \end{aligned} \tag{2}$$

Indeed, in order to increase $\Delta\log|\overline{X}|$ to maximum, $\Delta\log|\overline{X}^{2}|$ must be increased and $\Delta\log|\overline{X}^{1}|$ must be decreased. Now the log spectra of two subframes may be modified, up to maximum allowable values, using the kernel w_i. Thus the watermark log-spectrum is obtained (Fig. 5d).

The watermark receiver is shown in Fig. 6. The received symbol $\overline{y}$ is transformed to its delta-log-spectrum $\Delta\log|\overline{Y}|$ and then it is compared with Walsh functions used for transmission. Decision is made according to maximum of the correlation:

$$\arg\max_{i} \left\langle \Delta\log|\overline{Y}|, w_i \right\rangle, \tag{3}$$

where $\langle a, b\rangle$ is the scalar product.

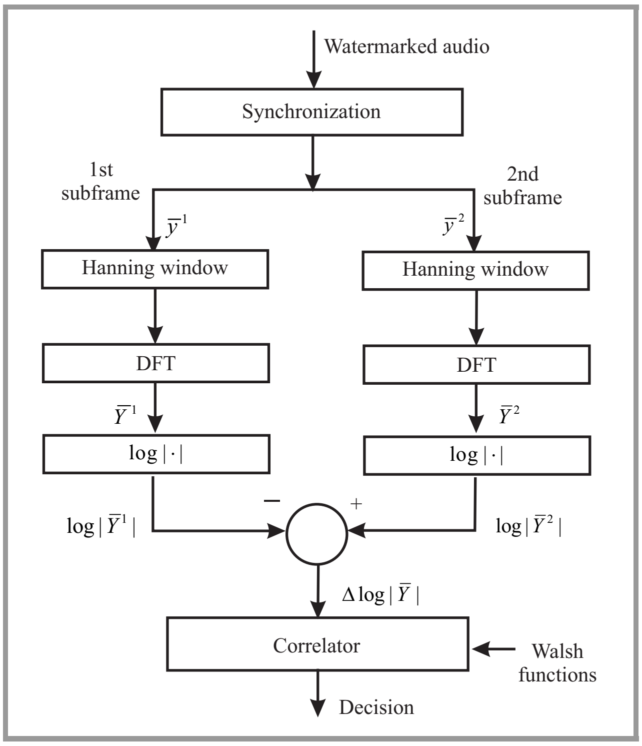

Fig. 6. Reception of a single symbol $\overline{y}$.

3.3. Dirty Paper Codes

Dirty paper codes use many symbols for transmission of the same information [18]. Of course this symbol (in this case the Walsh function) is used, which, in presence of known distortion (here the cover signal) yields maximum of $\langle\Delta\log|\overline{X}|, w_i\rangle$. In [14] the plain codes (Fig. 7a,b) have been tested and the dirty paper codes using up to 16 symbols per bit (two of these codes are shown in Fig. 7c,d).

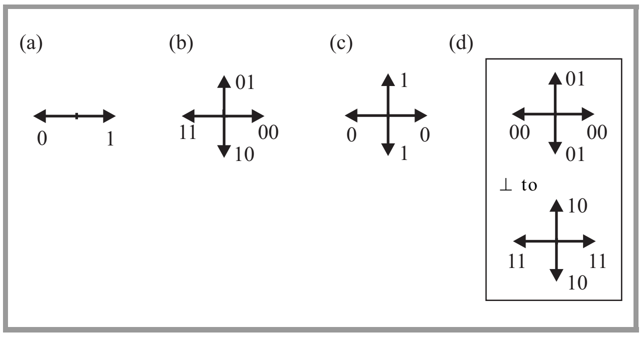

Fig. 7. Coding schemes: (a) antipodal, (b) quaternary bi-orthogonal, (c) bi-orthogonal dirty paper with 2 symbols per bit, (d) quaternary bi-orthogonal dirty paper with 2 symbols per duo-bit.

The plain bipolar code (Fig. 7a) and dirty paper bi-orthogonal code (Fig. 7c) transmit about 43 b/s (precisely there are 44100/1024 symbols per second). If the cover signal known at the watermark embedding is a dominant distortion, then the dirty paper code outperforms the plain binary code – compare BER values obtained without

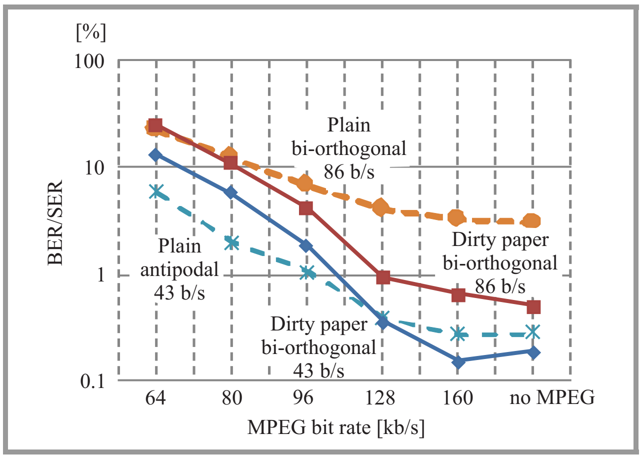

Fig. 8. Comparison of plain (dashed line) and dirty codes (solid line) concerning cases presented in Fig. 7.

MPEG coding in Fig. 8. If the quantization noise of the MPEG codec is the dominant distortion, then the plain code performs better than the dirty paper code. The advantage of dirty paper code is much more evident in transmission of duo-bits at 86 b/s. Dirty paper code (Fig. 7d), using two antipodal symbols for transmitting of the same duo-bit, outperforms evidently the plain quaternary code (Fig. 7b).

3.4. Combining the Dirty Paper and LDPC Codes

Dirty paper codes perform well if distortions known at the transmitter (i.e. the cover signal) are greater that the unknown distortions, i.e. the quantization noise of the MPEG-Audio coder. In order to obtain a watermarking scheme robust against both kinds of distortions a dirty paper code should be combined with an error correcting code (ECC). As a dirty paper code the quaternary bi-orthogonal code shown in Fig. 7d have been selected. The transmission rate is 86 b/s, but using an ECC of a code rate $\frac{1}{2}$ it is reduced to 43 b/s. As an ECC the low-density parity-check

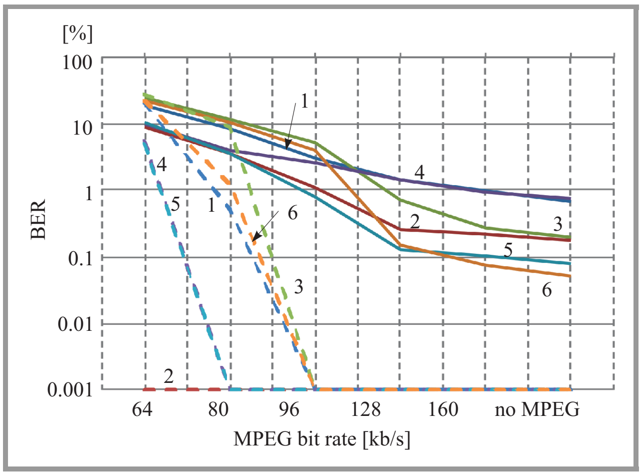

Fig. 9. BER for 6 audio files before (solid line) and after (dashed line) LDPC decoding.

code LDPC(400,800) is selected [21], [23]. Soft decision decoder, based on the log-likelihood ratio [23], is used.

The combination of the quaternary dirty paper code and the LDPC of a code rate $\frac{1}{2}$ outperforms the other coders tested in this paper as presented in Fig. 9. At the MPEG-Audio transmission rate 96 kb/s no single error was observed. Because only 24000 watermark bits were sent, it seems that BER is lower than 0.01% at the confidence level 0.95. Without MPEG coding 144000 bits were transmitted, so BER is below 0.002% at the confidence level 0.98.

3.5. Symbol and Block Synchronization

For symbol synchronization two pilot signals are used, of frequencies $f_k \cong 689$ and $f_{k+1} \cong 732$ Hz (the 16th and 17th base function of the DFT, $N = 1024$). The synchronization signal, $\sin(2\pi n f_k / f_s) - \sin(2\pi n f_{k+1} / f_s)$ is shown in Fig. 10. Note, that its envelope attains zero at the frame edges, which enables a smooth adjustment of its amplitude. It should not exceed the masking threshold. The angle between the 16th and the 17th DFT coefficients is proportional to the symbol shift $m: \psi = 2\pi m/N$. The DFT coefficients (only two are calculated) are cumulated for at least 3 seconds, which assures the symbol alignment accuracy of 20 samples (for details see [11]).

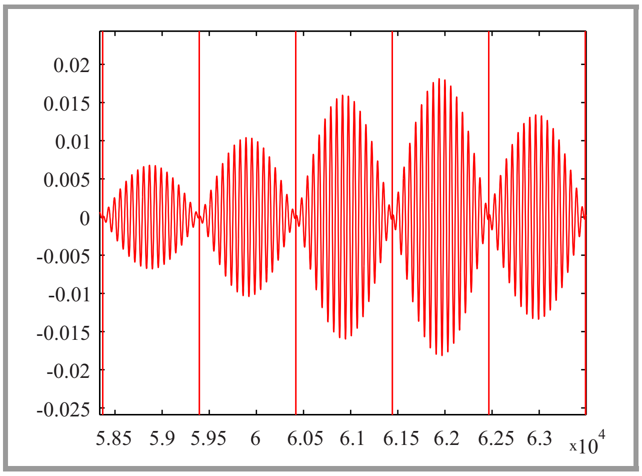

Fig. 10. The symbol synchronization signal.

The same pilot signals are used for sampling frequency offset measurement and correction in the case of resampling or different sampling frequencies in ADC and DAC converters). The algorithm, based on techniques used in OFDM [24], has been used for watermarking purposes in [8], [9] and tested in presented system [11].

Block synchronization is based on LDPC decoding with a shift of one symbol. The iterative decoding process is then observed. In case of block alignment, if the received watermarked audio signal is not severely distorted, the process converges quickly and stops after several iterations. Without block alignment there is no convergence, even in reception of undistorted watermarked audio.

4. Embedding and Detection of Digital Signatures

For embedding of the digital signature (logo) the same algorithm is used as for data transmission symbols embedding. Main steps of this algorithm are shown in Fig. 5. In this case, however, the pilot tones used for symbol synchronization are not embedded. Such tones are easy to detect and remove, which could compromise the copyright protection or spoken command authentication systems.

As digital signatures the Walsh functions are used. Number of Walsh functions being used defines system capacity. The same symbol is transmitted in each frame of 1024 samples, i.e., the same Walsh function is used, yielding the same frequency subbands (Fig. 5c). Due to differential watermark coding, spectrum increases and decreases in an alternate way. Changes occur every 512 samples (i.e. duration of a subframe). This regularity, however, does not cause any observable pattern neither in time, nor in frequency domain. In Fig. 11 spectrum of a digital signature is shown (i.e. spectrum of a difference between the watermarked and original signal). The same conditions are maintained as in the experiment shown in Fig. 3 (DFT window duration 2 s), but no particular spectral lines are observed. An unauthorized person could have difficulties in detecting and removing such watermark.

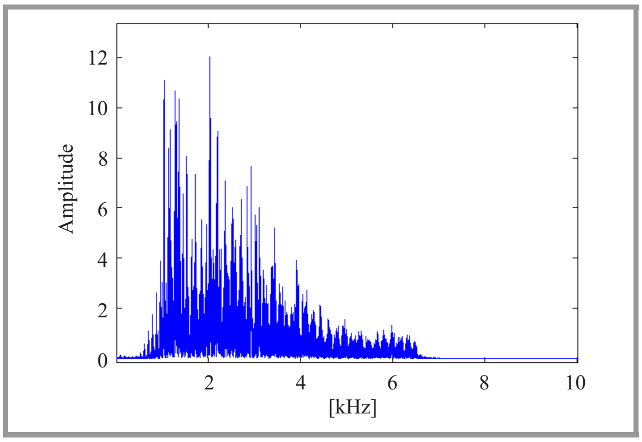

Fig. 11. Spectrum of a digital signature.

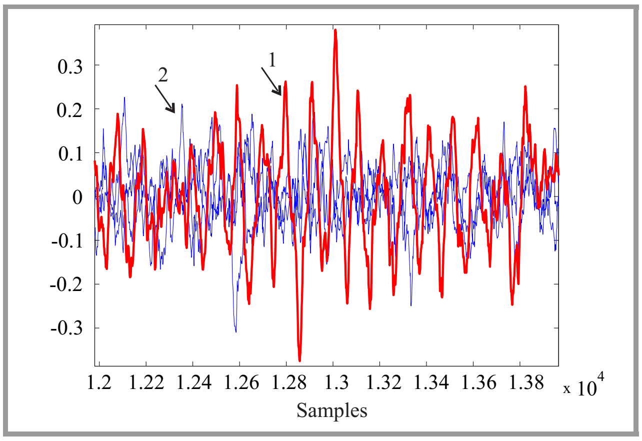

Fig. 12. Correlations $\Delta C_i = \langle \Delta \log |\overline{Y}|, w_i \rangle$ with: (1) a proper Walsh function and (2) improper ones.

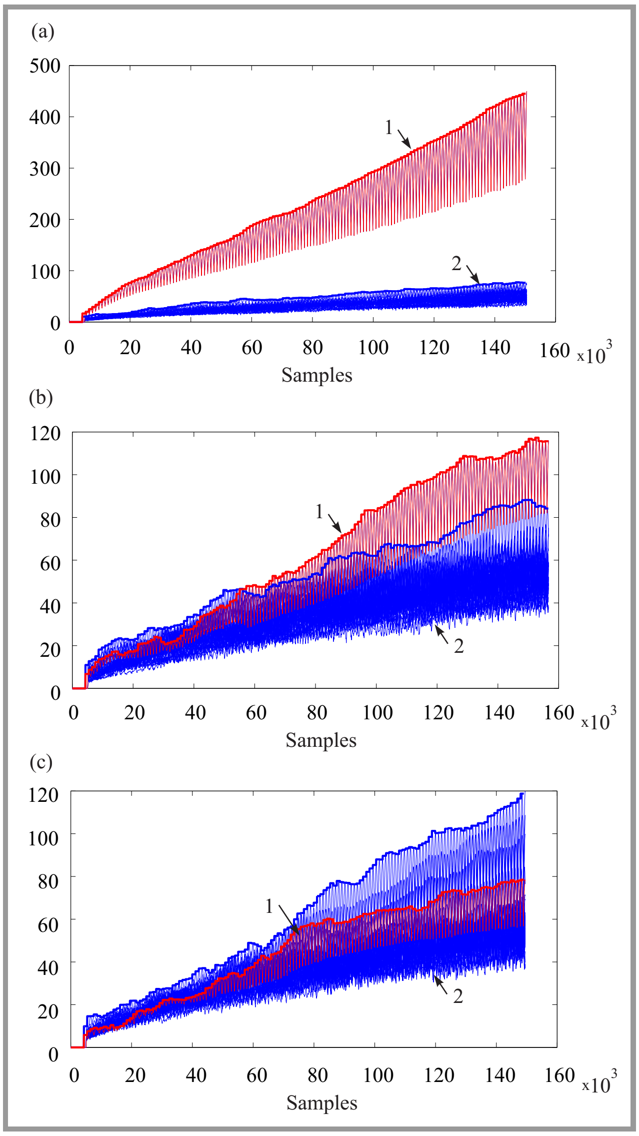

Fig. 13. The amplitude spectrum of the correlation series ΔC_i maximum calculated with: 1 – a proper Walsh function and 2 – improper ones.

Watermark detection algorithm does not require any synchronization. It uses the same scheme shown in Fig. 6, but the input signal $\overline{y}$ comes from the sliding window containing 1024 samples, shifted by 10–20 samples. The shift value may be increased because our detection algorithm is robust against symbol shift up to 50 samples. In every run of this algorithm correlations of the delta-log-spectrum $\Delta \log |\overline{Y}|$ with Walsh functions used in the system are calculated and stored. The correlation with the proper Walsh function should exhibit a maximum if the transmitting and receiving frames are aligned (i.e. every 1024 samples). Between the neighbor maxima a minimum should be expected. Correlations of the delta-log-spectrum with improper Walsh functions should not exhibit this quasi-periodic behavior – see Fig. 12 for evidence.

This quasi-periodic behavior of a correlations series may be detected by calculating the amplitude spectrum. It should exhibit a peak, exceeding the spectra of the improper series of correlations. These amplitude spectra are calculated in a growing window. Detection algorithm starts if this window is long enough to contain several periods of correlations series.

In Fig. 13 the maximum values of amplitude spectra are shown for a proper correlation series and 31 improper ones. Note that each maximum value varies in time, so it is better to use its envelope rather than its instantaneous value. Detection is made according to the margin, which is defined as a difference of the greatest envelope and the next one. In ideal conditions, i.e. small distortions of a watermarked signal – Fig. 13a, margin grows quickly in time and there is no problem in detecting the proper signature (logo). In more difficult conditions, i.e. greater distortions – Fig. 13b, margin starts to grow after some synchronization period. If no watermark is present or distortions are too great (Fig. 13c), margin is small and sometimes the "winning" Walsh function changes in time.

5. Testing

Both systems, i.e., the watermarking bit stream transmission system and the digital signature embedding system, were tested using 6 musical recordings. In some tests wideband speech signal was also used. These files contained monophonic signals, sampled at 44100 Hz. The following distortions were considered:

- quantization noise caused by the MPEG1-Audio codec,
- resampling caused e.g. by different sampling frequencies in DAC and ADC converters,
- symbol shift caused by improper synchronization or lack of synchronization,
- acoustic channel (DAC conversion, propagation of acoustic waveform in an office room, ADC conversion).

In annotation watermarking systems usually ASCII characters are transmitted. Any bit error yields an improper ASCII character, therefore transmission should be practically errorless. That is why error correcting codes (ECC) are implemented, that yield very small BER if Symbol Error Rate (SER) before ECC is sufficiently small. According to tests using LDPC, the acceptable SER values are less than 5% (Fig. 9). In order to reduce SER, dirty paper codes are used and informed watermark embedding [13]. Thus presented annotation watermarking algorithm is obtained, enabling transmission at 43 b/s, robust to quantization noise of the MPEG-Audio codec operating at 128 kb/s (BER $< 0.01\%$).

The watermark detecting algorithm (Fig. 6) is robust to symbol shift up to 50 samples (no BER increase is observed for shift < 50 samples, BER = 0.6% for shift = 100 samples [11]). Hence a symbol synchronization algorithm is required. Developed synchronization algorithm based on tones embedding (see Section 3.5) localizes a symbol with an error less than 20 samples in presence of distortions caused by the MPEG-Audio codec [11]. It is fully sufficient to assure robust operation of annotation watermarking in these conditions.

The system was also tested in presence of resampling. Due to the sampling frequency offset detection and correction transmission was robust against sampling frequency shift of 100 Hz. It is sufficient to compensate for the sampling frequency drift in DAC and ADC converters.

The most difficult conditions are in an acoustic channel. For testing purposes the microphone receiving the audio signal with watermark was located in 3 positions in a reverberant office room. The first one was at the distance about 50 cm from the loudspeaker but there was no direct sound propagation between both devices. The second one was at the distance about 30 cm from the loudspeaker and there was a direct sound propagation between the loudspeaker and the microphone. The third position was at the distance about 3 m from the loudspeaker and there was no direct sound propagation between both devices. In should be noted that a sampling frequency shift between DAC and ADC converters was present.

Results of tests in acoustic channel show that for system transmitting watermark as a bit stream a direct sound propagation is required (microphone position – 2). Figure 14 shows the average BER (for 6 phrases) as a function of watermark attenuation (offset) for a system without and with LDPC decoding (microphone position 2). The use of LDPC codes significantly decreased the BER for offset ≤ 18 dB. In simulations no single error was observed.

Digital signature, added to audio signal, should be difficult to detect by an unauthorized person and robust against malicious attacks of persons aiming at its removal or modification. Detection algorithm should not be vulnerable to the time shift, as it is based on amplitude spectrum of correlations shown in Fig. 12. This was fully confirmed in the simulation studies.

Resampling in a small range shouldn't affect digital signature detection, as it only affects the quasi-period of a series

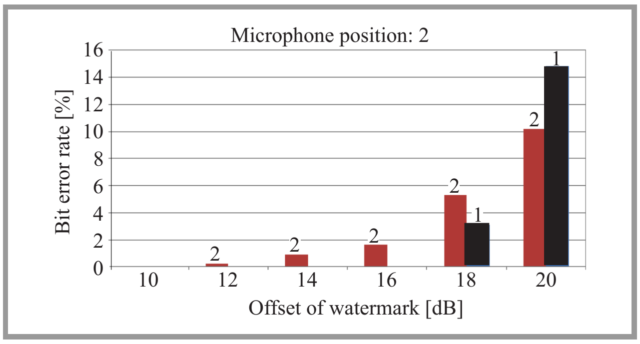

Fig. 14. BER as a function of watermark attenuation for six audio files same as in Fig. 1 for a system transmitting watermark as a bit stream – with (1) and without (2) LDPC decoding.

of these correlations. The results of simulations performed for several musical phrases and speech confirmed the authors presumptions. The correct reception of transmitted watermark is possible at sampling frequency deviation in range ⟨−200 Hz, +350 Hz⟩ from the 44.1 kHz nominal frequency. Higher sampling frequency changes may cause a shift of some subbands in the frequency scale and decrease of correlations. This prevents the proper detection of the watermark, even after accumulation of a large number of correlations.

The capacity of system hiding a digital signature, i.e. number of different logos recognized by the system, depends on the size of the window in which the spectrum is calculated. In a window containing 512 samples is is possible theoretically to put 256 Walsh functions, but this would require the use of the whole bandwidth (22.05 kHz for 44.1 kHz sampling frequency). Due to the influence of audio signal compression (which usually limits the band), in audio watermarking we use narrower bandwidth. Of course, it results in decreasing of system capacity. In our system we use up to 64 Walsh functions, which are active in band 1030–6550 Hz. It means that we can insert up to 64 different signatures. The greater system capacity, the greater receiver sensitivity to the factors hindering proper reception of inserted watermark (quantization noise, etc.). For this reason, the system real capacity was determined by simulation studies – Fig. 15. In Figs. 15–17 synchronization time is defined as a time after which the envelope of maximum of the amplitude spectrum for the correlation series calculated with a proper Walsh function permanently exceeds the other ones. If the proper signature is not detected, the synchronization time is not plotted. Capacity depends on the type of audio material. The most "demanding" in this regard is the sound of a violin – it only allows the insertion of 32 digital signatures.

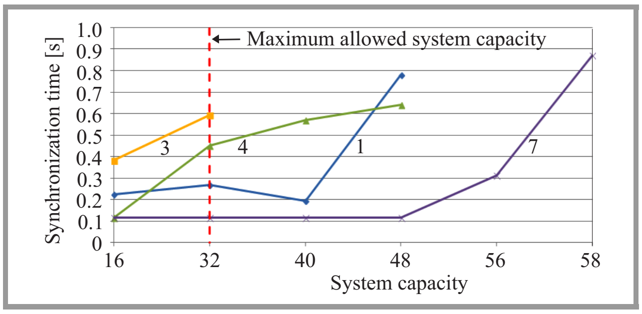

Fig. 15. Synchronization time as a function of system capacity for four phrases (1, 3, 4 – the same as in Fig. 1, 7 – speech) for a system embedding a digital signature in an audio signal.

As far as the audio signal compression (MPEG1-Audio) is concerned an attention to the fact that the compression with rates lower than 96 kb/s introduces a large distortion of music signal should be paid (Fig. 2). The results of simulations (Fig. 16) confirm the robustness of the system hiding digital signature for MPEG-1 Audio bit rate of 96 kb/s for all tested audio files. The least resistance

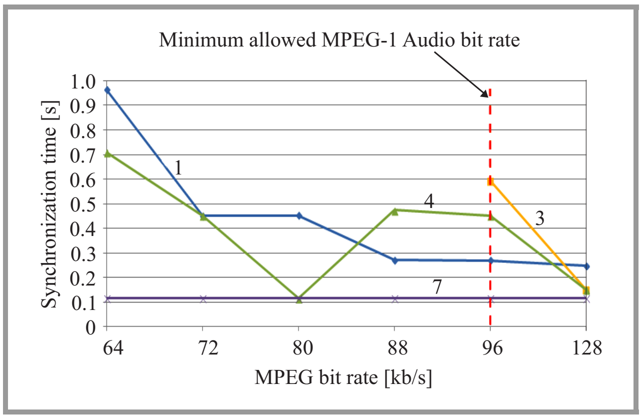

Fig. 16. Synchronization time as a function of MPEG-1 Audio bit rate for four phrases (1, 3, 4 – the same as in Fig. 1, 7 – speech) for a system embedding of a digital signature in an audio signal.

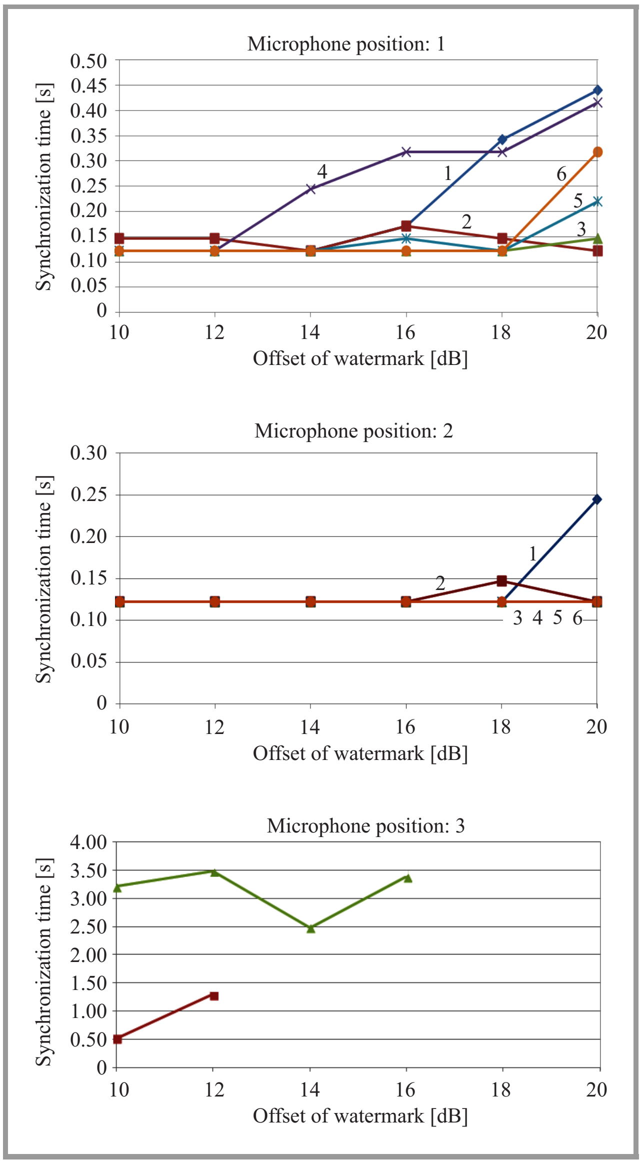

Fig. 17. Synchronization time as a function of watermark attenuation for six audio files same as in Fig. 1 for a system embedding of a digital signature in an audio signal.

exhibited the sound of the violin, the most resistant was the speech signal.

The signature embedding system was also tested in acoustic channel. The synchronization time as a function of watermark attenuation for six audio files is shown in Fig. 17. For microphone position 2 – direct sound propagation between speaker and microphone – there was no problem with proper decoding of watermark in relatively short time. The SER averaged for 6 audio files was about 0.33% for the biggest watermark attenuation (20 dB). The system was also robust against moderate reverberations – in microphone position 1 the synchronization time was less than 0.5 s for the biggest watermark attenuation (with SER $\sim 5\%$). For microphone position 3 only in two audio files the signature was properly detected after a few seconds of signal duration (with max offset 12 dB).

6. Conclusion

Two audio watermarking systems are proposed and tested. First one is devoted for annotation of acoustic files. The second one embeds a digital signature in the audio signal, e.g. for copyright protection. Both systems use similar watermark embedding algorithm, operating in frequency domain. The digital signature (logo) is robust to malicious attacks – it is hard to detect and remove by an unauthorized person. It must be noted, that its robustness is confirmed on a rather small number of signals and further testing should be effectuated. The main features of both systems are compared in Table 1. Computational complexity is moderate, the most complex is a detector of digital signature, because of overlapped windows shifted by only 20 samples.

Table 1
Comparison of algorithms for binary streaming and hiding a digital signature

Feature	Transmission of a bitstream	Hiding a digital signature
Bit rate/capacity	43 b/s	32 signatures
Watermark audibility	Inaudible in most cases	Inaudible ODG > -0.2
Robustness against symbol shift	Up to 50 samples, symbol synchronization required	Robust, no synchronization required
Robustness against resampling	Robust with sampling frequency offset correction	–200 Hz, +350 Hz, no offset correction required
Robustness against audio compression (MPEG1-Audio)	Satisfactory at 128 kb/s	Robust at bit rate 96 kb/s
Using acoustic channel	Direct sound propagation required	Robust against moderate reverberations
Computational complexity (transmitter/ receiver)	Without LDPC: 2.2/1.5 Mflops with LDPC 3/5.5 Mflops	2.2/50 Mflops

The proposed audio watermarking methods may be used for description of the recording (names of singers, lyrics of songs, etc.), for copyright protection, authentication of spoken messages, etc.

Acknowledgements

The research has been co-financed with the European Union funds by the European Social Fund.

References

[1] H. Kaur and U. Kaur, "Blind audio watermarking schemes: a literature review", *IRACST – Engin. Sci. Technol. an Int. J. (ESTIJ)*, vol. 3, no. 2, pp. 288–295, 2013.

[2] S. Xiang, "Audio watermarking robust against D/A and A/D conversions", *EURASIP J. Adv. Signal Proces.*, vol. 2011:3, 2011.

[3] C. Baras, N. Moreau, and P. Dymarski, "Controlling the inaudibility and maximizing the robustness in an audio data hiding system", *IEEE Trans. Audio, Speech Lang. Proces.*,vol. 14, no. 5, pp. 1772–1782, 2006.

[4] D. Megias, J. Serra-Riuz, and M. Fallahpour, "Efficient self-synchronized blind audio watermarking system based on time domain and FFT amplitude modification", *Sig. Proces.*, vol. 90, no. 12, pp. 3078–3092, 2010.

[5] R. Tachibana, S. Shimizu, T. Nakamura, and S. Kobayashi, "An audio watermarking method robust against time- and frequency fluctuation", in *Proc. SPIE Int. Conf. Secur. Watermark. Multim. Contents. III*, San Jose, USA, 2001, vol. 4314, pp. 104–115.

[6] X. Wang, P. Wang, P. Zhang, S. Xu, and H. Yang, "A norm-space, adaptive, and blind audio watermarking algorithm by discrete wavelet transform", *Sig. Proces.*, vol. 93, pp. 913–922, 2013.

[7] B. Y. Lei, I. Y. Soon, and Z. Li, "Blind and robust audio watermarking scheme based on SVD-DCT", *Sig. Proces.*, vol. 91, pp. 1973–1984, 2011.

[8] P. Gajewski, J. Łopatka, and Z. Piotrowski, "A new method of frequency correction using coherent averaging", *J. Telecommun. Inform. Technol.*, no. 1, pp. 142–146, 2005.

[9] Z. Piotrowski, "Drift correction modulation scheme for digital audio watermarking", in *Proc. Int. Conf. Multim. Inform. Netw. Secur. MINES 2010*, Nanjing, China, 2010, pp. 392–397.

[10] P. Dymarski, R. Markiewicz, and A. Kopertowska, "Sygnatury cyfrowe sygnałów fonicznych", Krajowe Sympozjum Telekomunikacji i Teleinformatyki, Gdańsk, Poland, *Telecom. Rev. and Telecom. News*, vol. 86, no. 8–9, pp. 650–656, 1013 (in Polish).

[11] P. Dymarski and R. Markiewicz, "Time and sampling frequency offset correction in audio watermarking", in *Proc. 18th Int. Conf. Syst. Sig. Image Proces. IWSSIP 2011*, Sarajevo, Bosnia and Herzegovina, 2011.

[12] P. Dymarski, "Watermarking of audio signals using adaptive subband filtering and Manchester signaling", *Proc. 14th Int. Conf. Syst. Sig. Image Proces. IWSSIP 2007*, Maribor, Slovenia, 2007.

[13] P. Dymarski and R. Markiewicz, "Informed algorithms for watermark and synchronization signal embedding in audio signal", in *Proc. 20th Eur. Sig. Proces. Conf. EUSIPCO 2012*, Bucharest, Romania, 2012, pp. 2699–2703.

[14] P. Dymarski and R. Markiewicz: "Audio watermarking in frequency domain using Walsh functions and LDPC codes", in *Proc. Int. Conf. Sys. Sig. Image Proces. IWSSIP 2013*, Bucharest, Romania, 2013, pp. 155–158.

[15] "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s", Part 3: Audio, ISO/IEC Standard IS 11172-3 Information technology, 1993.

[16] "Method for Objective Measurements of Perceived Audio Quality (PEAQ)", ITU-R Rec. BS 1387, 1998.

[17] Y. Zhao, W. Yang, D. Chang, and W. Guo, "A robust audio sonic watermarking algorithm oriented air channel", in *Proc. Int. Conf. Comp. Inform. Sci. ICCIS 2011*, Chengdu, China, 2011, pp. 53–57.

[18] M. H. M. Costa, "Writing on dirty paper", *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, 1983.

[19] M. L. Miller, G. J. Doerr, and I. J. Cox, "Applying informed coding and embedding to design a robust high capacity watermark", *IEEE Trans. Image Proces.*, vol. 13, no. 6, pp.792–807, 2004.

[20] A. Abrardo and M. Barni, "Informed watermarking by means of orthogonal and quasi-orthogonal dirty paper coding", *IEEE Trans. Sig. Proces.*, vol. 53, no. 2, pp. 824–833, 2005.

[21] R. Gallager, "Low density parity-check codes", *IRE Trans. Inform. Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[22] K. G. Beauchamp, *Walsh Functions and Their Applications*. London: Academic Press, 1975.

[23] "LDPC Code", Feb. 2014 [Online]. Available: https://sites.google.com/site/bsnugroho/ldpc/

[24] M. Sliskovic, "Sampling frequency offset estimation and correction in OFDM systems", in *Proc. 8th IEEE Int. Conf. Elec., Circ. Syst. ICECS 2001*, Malta, 2001, vol. 3, pp. 437–440.

Przemysław Dymarski received the M.Sc. and Ph.D. degrees from the Wrocław University of Technology, Poland, in 1974 and 1983, respectively, both in Electrical Engineering. In 2004 he received the D.Sc. degree in Telecommunications from the Faculty of Electronics and Information Technology of the Warsaw University of Technology. Now he is with the Institute of Telecommunications, Warsaw University of Technology. His research includes various aspects of digital signal processing, particularly speech and audio compression for telecommunications and multimedia, audio watermarking and applications of Hidden Markov Models.

E-mail: dymarski@tele.pw.edu.pl
Institute of Telecommunications
Faculty of Electronics and Information Technology
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Robert Markiewicz received the M.Sc. degree in Telecommunications from the Faculty of Electronics and Information Technology of the Warsaw University of Technology, Poland, in 2009. Since 2009 he has been a Ph.D. student. His interests focus on audio watermarking and new telecommunications access networks.

E-mail: r.markiewicz@tele.pw.edu.pl
Institute of Telecommunications
Faculty of Electronics and Information Technology
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland