jet INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2010, VOL. 56, NO. 3, PP. 257-262
Manuscript received July 10, 2010: revised September, 2010. DOI: 10.2478/v10177-010-0033-0

Methods of Time-Frequency Analysis in Authentication of Digital Audio Recordings

Rafał Korycki

Abstract—**This paper describes the problem of tampering detection and discusses the main methods used for authenticity analysis of digital audio recordings. For the first topic, two frequency measurement algorithms based on electric network frequency criterion are applied. Time-frequency analysis is used and improved with reassignment method for purpose of visual inspection of modified recordings. The algorithms are shortly described and exemplary plots are presented with interpretation. The last described method, recently proposed, is based on checking frame offsets in compressed audio files.**

Keywords—**Tampering detection, audio authenticity analysis, media authentication, electric network frequency, time-frequency analysis, digital audio forensics, inverse decoder, MP3, compressed audio, perceptual audio coding.**

I. Introduction

AUTHENTICATION is one of primary domains of forensic audio analysis. Audio recordings remain useless as evidence when there is no proof that they are original and they have not been tampered. In the case of analog recordings on audio tape, efficient, accurate and non-destructive is magneto-optical method based on Faraday or Kerr effect [7]. By using this method various types of information can be detected. Erase head marks can indicate if a passage has been deleted or if a recording has been copied. On the other hand, unusual track position due to misaligned audio heads can help to individualize the recorder that was used [8]. Time-frequency (TF) analysis of digitalized tapes can also be a suitable tool for tampering detection. Visual inspection of waveform and spectrogram allows to indicate start, stop and pause marks and can be helpful in detection of copied recording. This method can be also applied to establish, which tape recorder was used to produce specified recording [14].

Since digital audio recordings appeared, audio authentication has become more difficult and in most cases impossible. The currently available technologies and free editing software allow the forger to cut or paste any single word without audible artifacts. Nowadays the most accurate and frequently used method is the Electric Network Frequency (ENF) detection method [16]. This approach is based on random fluctuations of the 50 Hz or 60 Hz frequency emitted by the electric network and induced in electronic circuits of recording devices. The assumption of the method is a relatively high correlation of electric frequency changes within the areas connected to the same network grid. The main problems are: proper extraction of induced electric network signal, accurate measurement of electric network frequency and need of searching and comparison of these fluctuation patterns with a reference database.

This paper focuses on methods used in tampering detection of digital recordings. In section 2, frequency measurement methods are described and compared with a simple Fourier transform generally used in forensic ENF extraction. The use of time-frequency analysis plots is presented in section 3. In this section, reassignment method is also introduced for improvement of spectrogram readability and applied to selected transforms. Using these solutions, one can analyze minimal changes of background sounds, which can indicate tampering. In section 4, methods used for video and image analysis are described briefly with respect to perceptual audio coding techniques. Investigation of compressed audio files based on computation of a number of active coefficients (NAC) and measurement of frame offset is also presented.

II. ENF Analysis

European power system operators including PSE-Operator S.A from Poland are associated in the Union for the Coordination of Transmission of Electricity (UCTE). In any electric system the active power has to be generated at the same time as it is consumed. This is because of very limited possibility of storing electric energy. Disturbance in balance between total demand and production of electric power causes network frequency deviation from its default value. Increase of frequency value is caused by decrease in the total demand, whereas decrease of ENF is caused by decrease in power production, respectively. Normally, the system frequency has to be maintained within strict limits. The set-point frequency for power systems in UCTE is 50 Hz [23]. Changes of frequency values are random and are highly correlated in the range of synchronized area. Frequency values are measured by power system operators with high accuracy and are stored in dedicated database. Polish operator leads their database since 1997 [5].

Authenticity investigation of audio recordings based on ENF criterion depends on possible induction of electric network signal in electronic circuits of recording devices. There is also need for proper extraction and accurate measurement of the signal across the frequency range of interest. Three general methods to obtain ENF were described [17]. Time-frequency analysis using Short-time Fourier

R. Korycki is a member of ENFSI (European Network of Forensic Science Institutes) Speech and Audio Analysis Working Group (FSAA-WG) and a PhD student from Institute of Radioelectronics, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland (e-mail: R.Korycki@ire.pw.edu.pl).

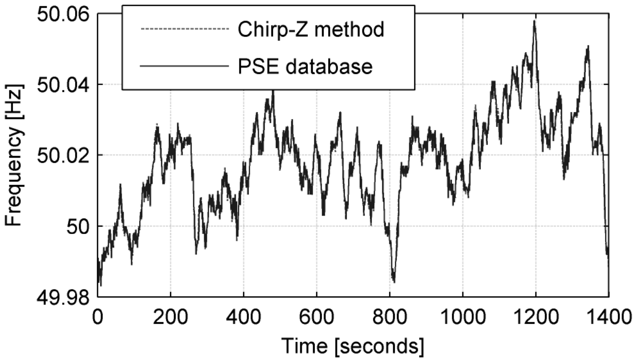

Fig. 1. ENF measurement using Chirp-Z method.

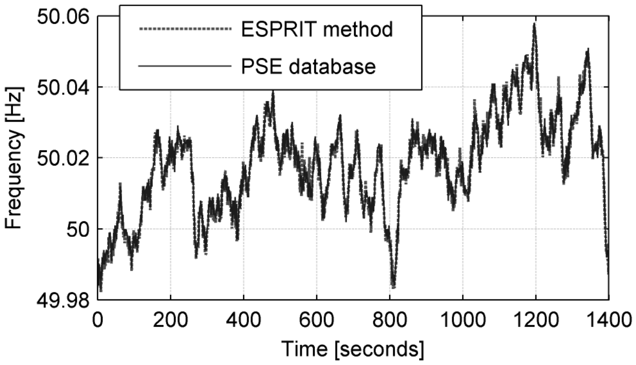

Fig. 2. ENF measurement using ESPRIT method.

transform (STFT) is used for visual comparison of extracted electric network signal with the plot determined based on the ENF database. The questioned recording should be previously down sampled to e.g. 120 Hz., which can allow using popular editing software to compute and observe spectrograms. This method is useful in general assessment whether ENF is present in the recording and in assessing disturbances introduced to the signal. Visual analysis can also help to determine if there is more than one ENF, which indicates that the recording could have been copied using analogue path.

Zero-crosses measurement method can also be used to obtain frequency values. It is recommended to use original sampling frequency rate and to remove DC component. Narrow bandpass filter is required as well. This method is used in power system frequency measurement [20] and is applied to pure sinusoidal signal. However in real case examples clear ENFs are not observed. Unwanted noise, amplitude fluctuations and higher order harmonics cause problems with implementation of this method.

In frequency domain, one can compute FFT over short time windows, measure the maximum value around the 50 Hz and compare questioned samples with database [17]. Electric network frequency variation is about 200 mHz and the measurement accuracy in PSE-Operator S.A. database is in range of 1-1,5 mHz [5]. There is a problem with achieving such accuracy and time resolution using Fourier transform of down sampled signal with e.g. 240 samples per window (with 50% overlapping).

Due to described circumstances, the use of Chirp-Z Transform (CZT) was proposed. This method, also called frequency zoom technology, was widely explored and applied in frequency measurement for power systems [10]. While Discrete Fourier Transform (DFT) algorithm evaluates signal z-transform on a circular way with unitary radius, the CZT uses a spiral path. This transform does not require analysis of the entire spectrum with the same resolution and is not dependent on the size of window. The main parameter is the ratio between analyzed spectral windows and the sampling frequency [1]. This means that the spectral resolution can be improved without increasing the length of observation window. Disadvantage of Chirp-Z Transform method is the computational effort (N^2), which is higher than required by the Fast Fourier Transform (FFT) ($N\log_2N$), and this may limit its application. However it is possible to compute CZT using algorithm based on three FFTs: two direct and one inverse [24].

Another way to obtain accurate frequency measurement from evidence recording is using parametric spectral estimation. Frequency measurement problem in power systems has been described and two methods: MUSIC (MUltiple SIgnal Classification) and ESPRIT (Estimation of Signal Properties through Rotational Independent Transformation) have been chosen as a good alternative to traditional Fourier-based techniques [19]. Both methods are based on eigendecomposition of a sample data covariance matrix. ESPRIT method is found to be more accurate and sophisticated than MUSIC according to real forensic cases, where measured signal could be modeled as a sinusoid damped in white noise [22]. ESPRIT is a signal subspace method where a sinusoid corresponds to a rotation of a complex phasor in the time domain. Identifying the frequencies is achieved by identifying this rotation [6].

The above methods were implemented in a real case example. Recording was made on April 24, 2007 between 3:37 PM and 4:00 PM. The battery powered HHB PDR-1000 digital audio tape recorder and the series 4000 DPA miniature microphone were used. The 23 minutes of conversation has been stored without compression with 16-bit resolution and 48 kHz sampling rate, followed by sample rate conversion to 120 Hz. Electric network frequency values were obtained from PSE-Operator S.A. database for a day when recording has been made. To avoid the influence of disturbances, signal was filtered using Butterworth band pass filter in a range from 48 Hz to 52 Hz. Two described methods were used to measure frequency values: CZT and ESPRIT. Short window size of 240 samples with 50% overlapping were used to achieve a good time resolution. According to these assumptions, it was unable to obtain accurate frequency values using Fourier transform only due to theoretical resolution for that window at about 250 mHz.

Comparing the outcomes with frequency values obtained from PSE-Operator SA database gives the mean absolute error of 0,81 mHz for CZT method and 1,32 mHz for ESPRIT method. As can be seen on Fig. 1 and Fig. 2, the ENF patterns are almost identical, when comparing the results obtained by using CZT and ESPRIT with the one given from power system operator database.

III. Visual Inspection of Time-Frequency Plots

Graphical spectral analysis can also be used for tampering detection in digital audio recordings. Although, short-time Fourier transform was applied in visual inspection of ENF, as mentioned in section 2, it can be used in case of integrity analysis of sinusoidal and harmonic signals with slow changing frequency. Disturbances like fan hum, engine sound or horizontal synchronization frequency in video cameras are accidently present in forensic recordings. Cutting out a part of evidential conversation also breaks the continuity of these interfered signals, which causes phase change and is visible on a spectrogram [2]. Speech signal integrity can be investigated basing on temporal spectral measurement, especially in the case of vowels. Slight changes within single words or phrases causing modification of the meaning could be detected. But the key point is readability of spectrogram, which means a good resolution in time and frequency as well as good concentration of the signal component and no misleading interference terms.

As it is known in the case of short-time Fourier transform, there is a trade-off between time and frequency resolutions. A good time resolution requires a short window, but a good frequency resolution requires a narrow-band filter so in this case a long window is required. These two conditions cannot be simultaneously granted. The above limitations are a consequence of the Heisenberg-Gabor inequality [3]. Overlapping windows make STFT segments mutually dependent. Using different types of windows to minimize the effect of cutting analyzed signal has an influence on readability of spectrogram as well.

Time-frequency transform, which can give better resolution is Wigner-Ville distribution (WVD) that satisfies a number of desirable mathematical properties. Although, the signal term is well localized in the time and frequency plane, other terms are present at positions where the energy should be null. The interferences appear due to the bilinearity and quadratic structure of the WVD. These terms oscillate perpendicularly to the line joining the two points interfering, with a frequency proportional to the distance between these two points [21]. Nevertheless, they can often be reduced, while preserving the time and frequency shift invariance property by a two-dimensional low-pass filtering. This smoothing operation is a tradeoff between time-frequency resolution and good interference attenuation. Smoothed Wigner-Ville distributions are also applied to multi component analysis like a speech signal [11]. Smoothed-pseudo WVD allows a continuous passage to the spectrogram, under the condition that the smoothing functions are Gaussian.

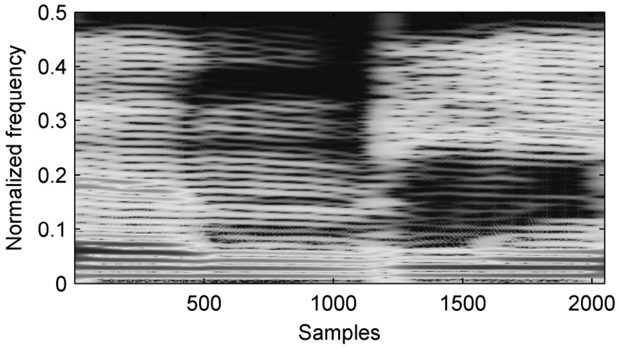

Fig. 3. Spectrogram of a part of Polish spoken tampered phrase “I am guilty”.

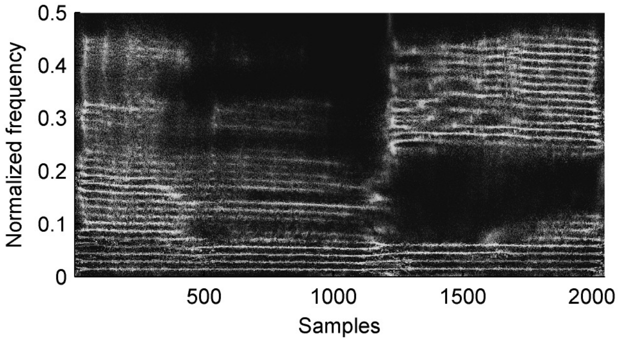

Fig. 4. Reassignment spectrogram of a part of Polish spoken tampered phrase “I am guilty”.

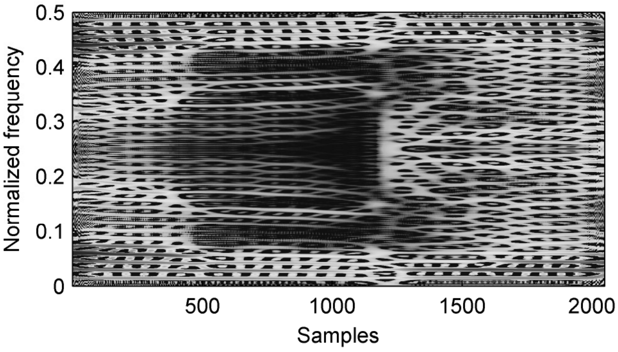

Fig. 5. Reassignment method applied to smoothed-pseudo Wigner-Ville representation of a part of Polish spoken tampered phrase “I am guilty”.

As a complement to solutions above, a reassignment method can be used to improve the readability of a signal representation. According to the previous distributions, time-frequency domains are delimited to the vicinity of the TF point, which constitutes geometrical center of the domain. Reassignment method changes the coordinates of this point to the center of gravity of the domain, which is more representative of the local energetic distribution of the signal. In addition to retaining the squared modulus, as in the case of spectrogram, this method also preserves the phase information of the short-time Fourier transform. The reassignment method can be considered as a smoothing, with a main purpose to reduce oscillatory interferences, but it causes the smearing of the localized components and a squeezing of the contributions that survived the smoothing [3], [4].

To visualize the use of time-frequency analysis in authentication of digital recordings, a controlled forgeries were made. From recorded phrase "I am not guilty" spoken in Polish, a word "not" has been removed, which completely changes the meaning of the sentence. Time duration of the recording was 2048 samples with 8 kHz sample rate. Short-time Fourier transform and smoothed-pseudo Wigner-Ville transform were computed with reassignment versions respectively.

By using the STFT it is hard to distinguish a tampering mark, which occurs near the 1200th sample, from impulse-like disturbances (Fig. 3). Applying the reassignment method improves readability of the figures and allows detecting the positions, where speech signal components are deformed slightly, which might be an indication of tampering (Fig. 4). Smoothed-pseudo Wigner-Ville distribution with applied reassignment method reveals remarkable interferences. However the traces of tampering still remain visible (Fig. 5).

This kind of analysis could be helpful in authentication of recordings, however it is not recommended to rely solely on this investigation.

IV. Analysis of Compressed Recordings

Recently much attention is paid to authenticity analysis of video and images. Several solutions have been proposed for detection of double quantization in digital video that results from double MPEG compression or from combining two videos of different qualities [13]. Other described methods are: blocking periodicity analysis in JPEG compressed images where the periodic blocking artifacts, due to differences in quantization errors between neighboring blocks, can be seen as an inherent feature in JPEG compressed images [9], and detecting traces of JPEG image resampling [12].

Unfortunately, there is no possibility to adapt these methods in analyzing compressed audio files. This is due to differences between compression algorithms. Perceptual audio coding techniques like MPEG Layer 3 (MP3) divide samples in time domain into frames with 50% overlapping and use floating point operations in quantization process, whereas JPEG compression is without overlapping and uses quantization table with integer numbers [18]. Although, it is impossible to convert tampering detection algorithms from image and video to audio recordings analysis, there is a way to adapt the idea of encoders investigation by implementing "inverse decoder" [15].

In MPEG Layer 3 encoder a sequence of 1152 input samples are poliphase filtered into 32 equally spaced frequency subbands depending on the Nyquist frequency of compressed signal. After that modified Discrete Cosine Transform (DCT) is applied to each time frame of subband samples and the 32 subbands are split into 18 thinner subbands creating a granule with a total of 576 frequency coefficients. To reduce artifacts caused by the time-limited operation on the signal, the windowing is applied. Depending on the degree of stationarity, the psychoacoustic model determines which of four types of window is used.

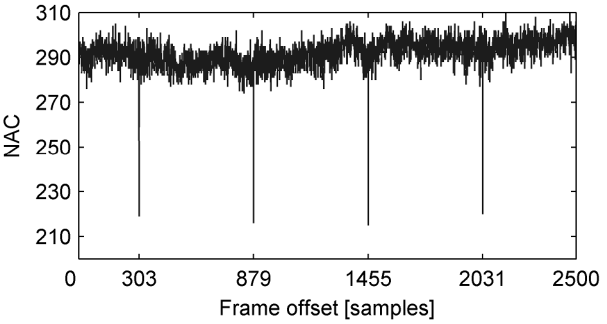

Fig. 6. Number of active coefficients in a function of frame offset.

When encoding process is applied, a number of spectral coefficients are quantized which causes an assignment to zero value. This is the key point in authentication of digital audio recordings, because the troughs in the spectral representation are only visible, when identical offset for spectral analysis is used [15]. There is also need to apply a correct spectral decomposition with almost the same analysis filterbank and window shape as used during the encoding process. That is because MPEG Layer 3 specification does not define the exact steps for encoding input data. The algorithm in encoders can therefore function quite differently and still satisfy the standard.

As an example of the above properties, a number of active coefficients were computed related to the frame offset as described in [18]. A free MPEG Layer 3 encoder was used with 128 kbps constant bit rate and 44,1 kHz sampling rate frequency. Normal window was applied to compute the spectral coefficients. It is shown in Fig. 6 that the synchronization occurs only for the multiples of 576 samples. Any modification of compressed audio file, including cutting off or pasting a part of audio recording would cause a disturbance in this regularity and could be easily detected. It is important to analyze recordings with every one of four types of windows and to know what algorithmic solutions were implemented in the encoder.

V. Summary

The methods presented in this paper can be successfully applied by forensic experts to detect tampering of digital audio recordings. Electric network frequency analysis approach is nowadays well known however, it could be improved by using more accurate and noise insensitive frequency measurement methods than just a simple Fourier transform. The spectrogram analysis is also a tool commonly used by forensic experts but in many cases one should not rely on it; especially when visual observations indicate no traces of tampering. In spite of these drawbacks, visual time-frequency analysis could still be used during authenticity analysis, and readability improvement methods could be implemented. The third described solution which regards analysis of compressed audio files is a very reliable tool. However widespread usage of this method requires more thorough research, especially in case of algorithmic differences between particular types of encoders.

Further work shall be done in the area of spectral distance measurement in the time-frequency plane using described transforms. The analysis of encoding algorithms implemented in commercial portable recorders in case of authenticity investigation will be a goal as well. The final results are planned to be described in author's dissertation.

REFERENCES

[1] M. Aiello and A. Cataliotti, "Chirp-Z Transform-Based Synchronizer for Power System Measurements", *IEEE Transactions on Instrumentation and Measurement*, vol. 54 (3), pp. 1025–1032, 2005.

[2] J. Apolinario and D. Nicolalde, "Evaluating digital audio authenticity with spectral distances and ENF phase change", in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Apr. 19-24, 2009, pp. 1417-1420.

[3] F. Auger, E. Chassande-Mottin, and P. Flandrin, *Time-Frequency Reassignment: From Principles to Algorithms*, Tempe, Arizona: CRC Press, 2003.

[4] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method" *IEEE Transactions of Signal Processing*, vol. 43 (5), 1068-1089, 1995.

[5] I. Biernacka, R. Korycki, and J. Rzeszotarski, "Analiza wahań częstotliwości prądu sieciowego w badaniach autentyczności nagrań cyfrowych", *Problemy Kryminalistyki*, vol. 258, pp. 36-40, 2007.

[6] M. Bollen, I. Gu, S. Ronnberg, and A. Tjader, "Performance evaluation for frequency estimation of transients using the ESPRIT: Measured noise versus white noise", in *Proc. 13th International Conference Of Harmonics and Quality of Power*, Wollongong, Sept. 28, 2008, pp. 1-8.

[7] D. Boss, S. Gfroerer, and N. Neoustroev, "A new tool for visualization of magnetic features on audiotapes", *International Journal of Speech, Language and Law*, vol. 10, pp. 255-276. 2003.

[8] J. Bouten, S. Donkers, and M. van Rijsbergen, "Derivation of transfer function for imaging polarimetry used in magneto-optical investigations of audio tapes in authenticity investigations", *Journal of Audio Engineering Society*, vol. 55, pp. 257-265, 2007.

[9] Y. Chen and C. Hsu, "Image tampering detection by blocking periodicity analysis in JPEG compressed images", in *Proc. IEEE 10th Workshop on Multimedia Signal Processing*, Cairns, Queensland, Oct. 8-10, 2008, pp. 803-808.

[10] L. Dai, J. He, C. Zang, and H. Zhen, "Using frequency zoom technology to realize high precision and adaptive frequency measurement for power system", in *Proc. Power System Technology, International Conference – PowerCon 2004*, Singapore, Nov. 21-24 2004, vol. 1, pp. 155-159.

[11] D. Devedeux, J. Duchene, and C. Marque, "Use of synthetic uterine signals for an optimum choice of time/frequency representation", in *Proc. 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Engineering Advances: New Opportunities for Biomedical Engineers*, Baltimore, Nov. 3-6, 1994, vol. 2, pp. 1246-1247.

[12] H. Farid and A. C. Popescu, "Exposing Digital Forgeries by Detecting Traces of Re-sampling", *IEEE Transactions on Signal Processing*, vol. 53 (2), pp. 758-767, 2005.

[13] H. Farid and W. Wang, "Exposing Digital Forgeries in Video by Detecting Double Quantization" in *Proc. ACM Multimedia and Security Workshop*, Princeton, 2009.

[14] J. A. Garcia Sanchez-Molero, "Establishment of the individual characteristics of magnetic recording systems for identification purposes", *Problems of Forensic Science*, vol. 47, pp. 20-39, 2001.

[15] R. Geiger, J. Herre, and S. Moehrs, "Analysing decompressed audio with the Inverse Decoder - towards an operative algorithm", in *Proc. 112th AES Convention*, Munich, May 10-13, 2002.

[16] C. Grigoras, "Digital Audio Recording Analysis: The Electric Network Frequency (ENF) Criterion", *The International Journal of Speech Language and the Law*, vol. 12 (2), pp. 63-76, 2005.

[17] C. Grigoras, "Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis", *Forensic Science International*, vol. 167, pp. 136-145, 2007.

[18] J. Huang, Z. Qu and R. Yang, "Detecting digital audio forgeries by checking frame offsets", in *Proc. of the 10th ACM workshop on Multimedia and security table of contents, International Multimedia Conference,* Oxford, 2008, pp. 21-26.

[19] Z. Leonowicz and T. Lobos, "Parametric Spectral Estimation for Power Quality Assessment" in *Proc. The International Conference on "Computer as a Tool"*, Warsaw, Sept. 9-12, 2007, pp. 1641-1647.

[20] R. Mikhael and Z. Salcic, "A new method for instantaneous power system frequency measurement using reference points detection", *Electric Power Systems Research*, vol. 55 (2), pp. 97-102, 2000.

[21] A. Papandreou-Suppappola, *Application in Time-Frequency Signal Processing*. Tempe, Arizona: CRC Press, 2003.

[22] T. Soderstrom and P. Stoica, "Statistical analysis of MUSIC and ESPRIT estimates of sinusoidal frequencies", in *Proc. The IEEE International Conference on Acoustics, Speech and Signal Processing,* Toronto, Apr. 14-17, 1991, vol. 5, pp. 3273-3276.

[23] UCPTE, "Summary of the Current Operating Principles of the UCPTE", Text Approved by the Steering Committee, Oct. 28, 1998.

[24] P. Zieliński, *Cyfrowe Przetwarzanie Sygnałów*, Warsaw: WKŁ, 2005, pp. 239-241.