Composite Usage of Local Magnitudes and Invariant Segmentation for Speech Signals Watermarking

Yuriy Pelekh[1], Dmytro Peleshko[2], Marta Peleshko[3], Igor Malets[4], Oleg Bashynskyy[5]

1. PIT Department, Lviv Polytechnic National University, UKRAINE, Lviv, S. Bandery street 12, E-mail: yuriy.pelekh@outlook.com
2. PIT Department, Lviv Polytechnic National University, UKRAINE, Lviv, S. Bandery street 12, E-mail: dpeleshko@gmail.com
3. MPA Department, Lviv State University of Life Safety, UKRAINE, Lviv, Kleparivska street 35, E-mail: dpeleshko@gmail.com
4. MPA Department, Lviv State University of Life Safety, UKRAINE, Lviv, Kleparivska street 35, E-mail: igor.malets@gmail.com
5. MPA Department, Lviv State University of Life Safety, UKRAINE, Lviv, Kleparivska street 35, E-mail: dpeleshko@gmail.com

Abstract - **Based on the combined usage of developed methods of sustainable segmentation and local magnitudes proposed a new scheme of watermark embedding into audio signals.**

The proposed scheme of audio signals watermarking ensures the reliability of several types of attacks, such as noise, cropping, resampling, re-quantization, compression and low frequency filtering.

Keywords - **initial segmentation, audio signal, digital watermarking, watermark, local magnitudes.**

I. INTRODUCTION

Data protection methods by its hiding can be split into three processes: cryptography, steganography and watermarking. Watermarking largely resembles steganography, but the difference is that the information is usually hidden in the object. Moreover, the digital watermark can be used not only to protect information from unauthorized copying, but also it may be used for the identification of individual data.

In recent years, many watermarking methods have been developed with the aim to create robust and invisible watermarks for audio information. The most popular methods are based on the discrete cosine transformation or other operational transformation [24, 25]. For example Zeng presented in [25] a "blind" method of watermarking, which integrates watermark into coefficients of discrete cosine transformation (DCT) using technique of quantization-modulation factor.

Recently intensively began to develop methods based on algebraic transformations that are either completely not or minimally affecting the frequency area of audio signal. For example Lee proposed method of embedding a watermark in the time domain of audio signal based on the usage of differential amplitude in each audio sample group to represent one bit of information. The basis of this technique is changes of low amplitude values for scaling amplitudes in selected segments of the samples, so signal time-domain can be almost completely preserved.

The main disadvantage of frequency based approach of the audio signal watermarking is the difficulty to provide the necessary robustness and stability to unauthorized attacks in real time. This is caused by many reasons among which the main one requires to know whole signal's time period for the implementation of operational transformation, non-invariant signal segmentation methods for some types of algebraic operations and others.

The main purpose of this paper is to develop a watermarking method which will be resistant to certain types of attacks and modifications of audio signal. Resistance to modifications provided by developed segmentation scheme, which is invariant to the choice of the source origin and affine transformations of audio signal.

II. AUDIO SIGNAL SEGMENTATION

It is known that one of the most common problems that arise in the analysis of speech signals, in the modern artificial intelligence systems development, is to determine their temporal and frequency characteristics as any non-determined signal is nonlinear object [3]. However in such signal with determined period of discretization it is always possible to select some time interval, on which values of these characteristics have very low deviation. Than signal characteristics on this interval considered as constants, and process of getting such interval called – audio signal segmentation [2, 4].

When highlight qualitative characteristics of the speech signal, then it is possible to build quite accurate parametric model of the speech signal on the selected segment, which, in turn, can be successfully used in solving a wide range of problems, including problems of watermarking.

So, the first task in watermarking method is developing of an audio signal segmentation method. Proposed to use pseudo rotation of symmetric matrix of distances and criterial factor for auto detection of edge values.

Let in space R1 defined speech signal which may be considered as continuous subjective representation of x: $\tau \rightarrow$ R1, which is defined on compact $\tau = [0; T]$, $T \in R1$, $T > 0$. If speech signal x(t) presents as a union of elementary areas

$$x(t)=\bigcup_{i=1}^{n} X_i,\ 0<n<\infty, \quad (1)$$

where $Xi = \{x(t) |\ t \in Ti\}$ – item of elementary coverage of the signal x(t), built based on disjunctive coverage on the range [0; T]. All items Xi of elementary coverage are equally powered and in discrete representation have the same measurements.

Taking into account (1) the segmentation task may be solved as a creating of a coverage {Yi}

$$x(t) = \bigcup_{i=1}^{m} Y_i,\ m \le n, \quad (2)$$

where m – count of quasistationary areas Y_i, which exists from union of consequence items Xi

$$Y_i = \bigcup_{j=I_i}^{m_i+I_i-1} X_j, \quad (3)$$

here I_i – start index of a union (3) in coverage η, in this case always $I_1 = 1$; m_i – count of elementary areas X_i in the union

$$n = \bigcup_{i=1}^{m} m_i \quad (3)$$

The first step of solving speech signal segmentation task is initial segmentation (1).

Next step after initial segmentation is building of aggregate matrix of divergence values, which in case of Euclid metric usage will look like

$$\forall i \in [1;n]: \nabla_i = \begin{pmatrix} 0 & \Delta\frac{dx_{1,1}}{dt} & \dots & (l-1)\Delta\frac{dx_{1,1}}{dt} \\ \Delta\frac{dx_{1,1}}{dt} & 0 & \dots & (l-2)\Delta\frac{dx_{1,2}}{dt} \\ \dots & \dots & \dots & \dots \\ (l-1)\Delta\frac{dx_{1,1}}{dt} & (l-2)\Delta\frac{dx_{1,2}}{dt} & \dots & 0 \end{pmatrix} \quad (4)$$

where i – index of elementary area X_i, $x_{i,j} = x(t_j)$.

To solve task finding vectors of characteristics $g_i = (g_i,1, \dots, g_i,l)$ of the elementary area of the speech signal for each item in Xi take a look at next equation

$$g_i = \nabla_i^{-1} x_i. \quad (5)$$

where $x_i = \{x(t_p) \in X_i \mid p = 1..l\}$ – l-dimensional vector of amplitude values of the speech signal x(t) on the interval Xi; Equation (12), because of invertible matrix ∇i solving as proposed in [1]

$$g_i = \nabla_i^{+} x_i + \left(1 - \nabla_i^{+} \nabla_i\right) r_i, \quad (6)$$

where ∇_i^{+} – generalized inverse Moore–Penrose matrix (pseudoinverse for matrix ∇i [1]); r_i – random vector of size l, which in iterative process defined by residuals:

$r_i^{j+1} = \|x_i - \nabla_i g_{ij}\| l$, here $\|\cdot\| l$ – l-norm [1]. First addition in (13) is pseudoinverse solution, and second – is a solution of solid system $\nabla_i g_i = 0$.

Moore–Penrose matrix itself ∇_i^{+} defined by singular value decomposition of matrix ∇_i in the next way

$$\nabla_i^{+} = V_i \Sigma_i^{+} U_i^{T}, \quad (7)$$

where U_i, V_i – unitary matrix of power l×l of the singular decomposition of matrix ∇_i; Σ_i^{+} – matrix of power l×l, which is pseudoinverse matrix to diagonal matrix Σ_i singular value decomposition [3] of matrix ∇_i.

In vectors space $\{g_i\}$ introduce metric for two values X_i and X_j by Chebyshev's distance of corresponding vectors g_i and g_j.

$$\mu(X_i, X_j) = \max_{k \in [1;l]} \{|g_{i,k} - g_{j,k}|\}, \quad (8)$$

Using metric (16) it is possible to perform initial segmentation in the way of detecting belonging condition of the elementary item X_j to quasistationary Y_i'

$$X_j \in Y_i' \Leftrightarrow \forall z \in [a_i; b_i] : \mu(X_j, X_z) \le \varepsilon, \quad (9)$$

where $\varepsilon \in R_{1,+}$ – edge value, which is a epsilon of defining if item X_j belongs to Y_i'.

Task (17) has a solution if edge value ε is defined. However in case of detecting of criteria feature K detection mode ε may be solved by using some extreme task.

Implementation of such approach may be method of automated finding of ε, in the basis of which criteria K is deviation optimization of segmentation results Y obtained by developed method from results Y_{er} obtained by "etalon" method

$$K: \|Y - Y_{er}\| \to \mathrm{opt}(\varepsilon). \quad (10)$$

During the practical implementation, the described method for automated determination of threshold ε based on maximizing the value of the selected coefficient of similarity (similarity measure) [6]. In general case instead of one coefficient it is possible to select few and by its help define integral parameter. With this aim from common equation of continuum measures of Somkin [6]

$$K_{\tau,\iota}(Y, Y_{er}) = \left(\frac{K_{\tau,\iota}(Y \mid Y_{er}) + K_{\tau,\iota}(Y_{er} \mid Y)}{2}\right)^{\frac{1}{\tau}}; -1 < \tau < \infty, -\infty < \iota < \infty \quad (11)$$

with τ = 0 (measure of similarity next objects according with common equation of Kolmogorov [6]) i ι = +∞, 1, 0, 1, -∞ selected ordered by ι list of most used measures of similarity, in particular Kulchynsky K0,1[12], Ohai K0,0[16], Sørensen K0,-1[21], Braun-Blanquet K0,-∞[8], Shymkevych-Simpson K0,+∞[20]. From list of elements it is possible to define average value

$$K_{\Sigma} = \frac{K_{0,+\infty} + K_{0,1} + K_{0,0} + K_{0,-1} + K_{0,-\infty}}{5}. \quad (12)$$

Then, accordingly to (18), to calculate value ε solving task of looking for max:

$$\|Y - Y_{er}\| \to \max_{0<\varepsilon\le 1} K_{\Sigma}. \quad (13)$$

Software implementation of the proposed algorithm is tested on the example of segmentation of word "миша". Speech signal characteristics are next: word length – 1,03 seconds, frequency 11 025 Hz, elementary item length – l = 120 samples.

Note, in the practical implementations, for distances calculations Chebyshev and cosine metrics were used. In a result of determining edge value ε in case of Chebyshev metric obtained value ε = 0.15, in case of cosine metric – ε = 0.01.

As the evaluation of criteria for automatic deviation detection for ε selected results of the same signal segmentation by algorithm DELCO [7] with edge value of 1.6 at the same length of elementary area (numerical values of segmentation by DELCO method is in the table). Herewith extreme task was solving according to integral index (12). In addition to mentioned metrics also was calculated Yurtsev K0,+∞[9] and Jacquard K1,-1[14] metrics. It was not included into integral factor (12) because Yurtsev metric is double measure to Braun-Blanquet, and measure of overlay Jacquard is equal to Sørensen.

Results of segmentation by this method are presented in table 1.

TABLE I

SEGMENTATION RESULTS OF WORD "МИША" USING DIFFERENT MEASURES.

Interval (ms:ms)	Segment by index	Count of elements	Interval (ms:ms)	Segment by index	Count of elements
Segmentation By Matrix of Divergences #Euclidean metric			Segmentation by Asymmetric Convergence Matrix		
(120:480)	[1;4]	3	(480:1200)	[4;10]	6
(480:5760)	[4;48]	44	(1200:1680)	[10;14]	4
(5760:6000)	[48;50]	2	(1680:2160)	[14;18]	4
(6000:6240)	[50;52]	2	(2160:2760)	[18;23]	5
(6240:6840)	[52;57]	5	(2760:3000)	[23;25]	2
(6840:7920)	[57;66]	9	(3000:3600)	[25;30]	5
(7920:10800)	[66;90]	24	(3600:3960)	[30;33]	3
(10800:11160)	[90;93]	3	(3960:4680)	[33;39]	6
#Cosine metric			(4680:5160)	[39;43]	4
(240:1920)	[2;16]	14	(5160:5760)	[43;48]	5
(1920:5760)	[16;48]	32	(5760:6240)	[48;52]	4
(5760:6000)	[48;50]	2	(6240:6600)	[52;55]	3
(6000:6240)	[50;52]	2	(6600:7200)	[55;60]	5
(6240:6600)	[52;55]	3	(7200:7920)	[60;66]	6
(6600:6840)	[55;57]	2	(7920:8400)	[66;70]	4
(6840:7080)	[57;59]	2	(8400:9000)	[70;75]	5
(7080:7320)	[59;61]	2	(9000:9240)	[75;77]	2
(7320:7920)	[61;66]	5	(9240:9840)	[77;82]	5
(7920:11040)	[66;92]	26	(9840:10440)	[82;87]	5
(11040:11160)	[92;93]	1	(10440:11160)	[87;93]	6
Delco Segmentation					
(120:3000)	[1;25]	24	(6240:6840)	[52;57]	5
(3000:3600)	[25;30]	5	(6840:7920)	[57;66]	9
(3600:5760)	[30;48]	18	(7920:9000)	[66;75]	9
(5760:6240)	[48;52]	4	(9000:9840)	[75;82]	7
			(9840:10200)	[82;85]	3

III. WATERMARKING ALGORITHM BASED ON LOCAL MAGNITUDES AND EXAMPLE OF ITS USAGE

Watermarking is embedding watermark bits into audio signal samples. To solve this problem it is necessary to find the corresponding samples for watermarking. The main criteria of such samples selection is the complexity of its revealing by third party methods and tools; the possibility of its calculation after signal modifications, reconstruction or filtering, changes intangibility to the human ear, data loss minimization.

Searching process of required signal's samples is to calculate the maximum distance between two next points:

$$d = \max(\|xi - xi+1\|), xi \in Yi. \tag{14}$$

When points for watermark embedding are calculated, watermark bits, it might be text, other speech signal, image, whatever that can be digitized and transformed in a bits sequence, continuously embedding into audio signal by next scheme:

In case 0-bit – reset (set 0) less meaningful bits of three signal samples at the left from key sample; In case embedding 1-bit – reset less meaningful bits of three samples to the right.

In case of watermark decoding compare sum of three less significant bits to the left and to the right from key point of watermarking. Based on the result decode 0 or 1 bit.

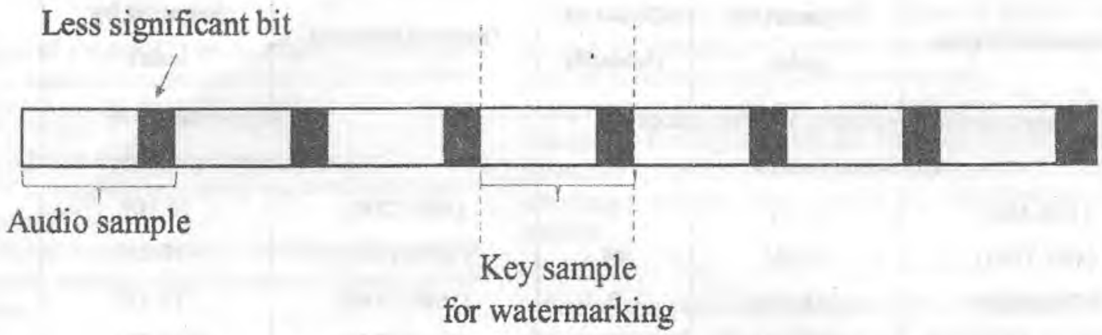

Fig.1 Scheme of digital watermark encoding/decoding into/from speech signal

In Table 2 provided results (in percentage) of the watermarking method. The first number represents the percentage of similarity for encoded and decoded watermark, the second value represents the percentage of correctness for points which were used for watermark encoding/decoding.

TABLE 2

RESULTS OF SPEECH SIGNAL DIGITAL WATERMARK ENCODING AND DECODING

Stationary split		Quasistationary split		Test word
Signal	Spectrum	Signal	Spectrum	
85 / 87	95 / 100	89 / 92	97 / 100	Тест
80 / 80	93 / 100	79 / 79	96 / 100	Миша

III. CONCLUSION

The obtained results of audio signal segmentation show that Kulchinsky, Ohai, Sørensen and Braun-Blanquet measures approximately equally represent the similarity of the results according to DELCO method. At that time Chebyshev's metric provides better results of proximity measure according to the segmentation results by DELCO method.

Values of all defined coefficients show that segmentation results based on the developed method by Chebyshev's metric are most similar to the results by DELCO. Herewith the deviation of values is not sufficiently large even when taking into account the value of Simpson's measure. This proofs that developed method based on Chebyshev's distance is resistant enough to proximity measures and allows to use the calculation of only one of the Kulchinsky, Ohai, Sørensen, or Braun-Blanquet or Simpson coefficients instead of using the integral index (20) of convergence matrix.

In the case of cosine metric usage measures of similarity values are smaller than values, but deviation is also small.

Therefore, usage of cosine metrics assumes stability of the segmentation results and automatic determination of deviation threshold accordingly to selected proximity measure.

Based on the experimental results of audio signal digital watermarking with using method of local magnitudes combined with developed method of segmentation improves correctness of looking for local magnitude points on signal spectrum and improves general result of watermark encoding and decoding into speech signal.

REFERENCES

[1] Алберт А. Регрессия, псевдоинверсия и рекуррентное оценивание/ А.Алберт// пер.с англ. – М.: Наука. – 1977. - 224с.

[2] Дорохин О.А., Старушко Д.Г., Федоров Е.Е., Шелепов В.Ю. Сегментация речевого сигнала. «Искусственный интеллект». – 2000. – №3. – С. 450- 458.

[3] Кагановский Ю. Д. Применение модели линейного предсказания для анализа стохастических сигналов [Текст] / Ю. Д. Кагановский // Технические науки: традиции и инновации: материалы междунар. науч. конф. (г. Челябинск, январь 2012 г.). — Челябинск: Два комсомольца, 2012. — С. 12-14.

[4] Маркел Дж.Д., Грэй А.Х. Линейное предсказание речи: Пер. с английского / Под редакцией Ю.Н.Прохорова и В.С.Звездина. - Москва:

Издательство «Связь», 1980
[5] Матрица расстояний. [Электронный ресурс] – Режим доступу до журналу: http://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0_%D1%80%D0%B0%D1%81%D1%81%D1%82%D0%BE%D1%8F%D0%BD%D0%B8%D0%B.
[6] Сёмкин Б.И. Об эквивалентности мер сходства и различия/ Б.И.Семкин, В.И.Двойченков// Исследование систем. Т. 1. Анализ сложных систем. Владивосток: ДВНЦ АН СССР, 1973. С 95-104.
[7] Bassia P., Pitas I., Nikolaidis N. Robust Audio Watermarking in the Time domain, IEEE Transaction on Multimedia, vol. 3, no. 2, pp. 232-241, June, 2001.
[8] Braun-Blanquet J. Pflanzensoziologie Grundzüge der Vegetationskunde. — Berlin: Verlaq von Julius springer, 1928. — 330 s.
[9] Cox I., Killian J., Leighton F., Shamoon T. Secure Spread Spectrum Watermarking for Multimedia. IEEE Transactions on Image Processing, vol. 6, no. 12, pp. 1673-1687, 1997.
[10] Huang J., Wang Y., Shi Y. Q. A Blind Audio Watermarking Algorithm with Self-Synchronization. Proceedings of IEEE International Symposium on Circuits and Systems, (ISCAS□02), vol. 3, pp. 627–630, 2002.
[11] Jaccard P. Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines // Bull. Soc. Vaudoise sci. Natur. 1901. V. 37. Bd. 140. S. 241—272.
[12] Kulczinsky S. Zespoly róslin w Pienach // Bull. intern. acad. polon. sci. lett. Cl. sci. math. natur. Ser. B. 1927. S. 2. P. 57-203.
[13] Lee S. K., Ho Y. S. Digital Audio Watermarking in the Cepstrum Domain. IEEE Transactions on Consumer Electronics, vol. 46, no. 3, pp. 744–750, 2000.
[14] Lie W. N., Chang L. C. Robust and High-Quality Time-Domain Audio Watermarking Based on Low-Frequency Amplitude Modification, IEEE Transaction on Multimedia, vol. 8, no. 1, pp. 46-59, February, 2006.
[15] Liu S. C., Lin S. D. BCH Code Based Robust Audio Watermarking in the Cepstrum Domain. Journal of Information Science and Engineering, vol. 22, pp. 535–543, 2006.
[16] Ochiai A. Zoogeographical studies on the soleoid fishes found Japan and its neighboring regions. II // Bull. Jap. Soc. sci. Fish. 1957. V. 22. № 9. P. 526—530.
[17] Penrose R. A generalized inverse for matrices. Proceedings of the Cambridge Philosophical Society 51, 406-413 (1955).
[18] Pooyan M., Delforouzi A. Adaptive and Robust Audio watermarking in Wavelet Domain. Proceedings of International Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP 2007), vol. 2, pp. 287-290, 2007.
[19] Rashkevych Y., Peleshko D., Kovalchuk A., Kupchak M., Pelekh Y. Speech signal pseudoinvariants. Computer Science and Information Technologies: Materials of the Vith International Scientific and Technical Conference CSIT 2011. - Lviv: Publishig House Vezha&Co, 2011. 2011, c.21-22.
[20] Simpson G.G. Holarctic mammalian faunas and continental relationship during the Cenozoic // Bull. Geol. Sci. America. 1947. V. 58. P. 613—688.
[21] Sörensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content // Kongelige Danske Videnskabernes Selskab. Biol. krifter. Bd V. № 4. 1948. P. 1-34.
[22] Szymkiewicz D. Une contribution statistique a la géographie floristique // Acta Soc. Bot. Polon. 1934. T. 34. № 3. P. 249—265.
[23] Wang X. Y., Zhao H. A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT. IEEE Transaction on Signal Processing, vol. 54, no. 12, pp 4835-4840, December 2006.
[24] Xie L., Zhang J., He H. Robust Audio Watermarking Scheme Based on Nonuniform Discrete Fourier Transform. Proceedings of IEEE International Conference on Engineering of Intelligent System, pp. 1-5, 2006.
[25] Zeng G., Qiu Z. Audio Watermarking in DCT: Embedding Strategy and Algorithm. Proceedings of 9th International Conference on Signal Processing (ICSP□09), pp. 2193-2196, 2008.