Volume 4, No. 2, February 2013
Journal of Global Research in Computer Science
ISSN-2229-371X
RESEARCH PAPER
Available Online at www.jgrcs.info

CONSTRUCTING OF PSEUDOINVARIANTS AND DIGITAL WATERMARKING OF SPEECH SIGNALS BASED ON A SINGULAR VALUE DECOMPOSITION

Dmytro Dmytrovych Peleshko[*1], Yuriy Myronovych Pelekh[2], Ivan Viktorovych Izonin[3]

[*1] Publishing Information Technologies, Lviv Polytechnic National University, Lviv, Ukraine
dpeleshko@gmail.com[1]
[2] Publishing Information Technologies, Lviv Polytechnic National University, Lviv, Ukraine
Yuriy.Pelekh@hotmail.com[2]
[3] Publishing Information Technologies, Lviv Polytechnic National University, Lviv, Ukraine
IvanIzonin@gmail.ru[3]

Abstract: Presented digital watermarking method of speech signals based on a constructing of quasistationary intervals, pseudoinvariants and singular value decomposition. Proposed a solution to the problem of building quasistationary intervals of speech signal. This method is based on the characteristics of singular value decomposition of square matrices of operators defined on the elementary intervals. Feature of these methods is their independence from the model voice tract. Watermarking method not dependent on input signal and watermark itself.

Keywords: digital watermarking, singular value decomposition, speech signal, quasistationary intervals, invariant, topology, the operator.

INTRODUCTION

Task of the presplitting of the speech signal into quasistationary intervals is one of the key problems in complex with speech analysis, transformation and synthesis. The successful solution of this task makes the major impact on the efficiency of solving the problem in general. The speech signal is non-stationary and invariant by its nature, although due to the inertia of articulators (lips, tongue, etc.) it is possible to allocate intervals, spectral compositions of which are very similar. Such intervals called quasistationary. In each case, the results of dividing the signal into quasistationary intervals (segmentation) are subjective, because depends on the selected method of further processing. The variety of audio elements of the speech signal and the tasks of its processing leads to the fact that there is not the optimal, and even, the conventional method of segmentation.

Since the speech signal has the property of invariance means that it does not depend on the coordinate system and is often distorted or noisy. The problem of data mining can be reduced to the construction of the fundamental spaces of invariants through which can be divided two non-equal objects.

Analysis of the literature [1] shows that for each of the selected models of voice tract and for each signal processing task can be recommended various methods of initial splitting. Typically, these techniques are closely related to the model voice tract and, accordingly, are not universal.

TASK

The main goal is to develop methods of digital watermarking and segmentation based on characteristics of singular value decomposition using the operator specified in the topological space of speech signal. A feature of these methods is their independence from the signal model. Under the segmentation means the construction of quasistationary intervals.

Construction of the operator in the form of a square matrix determined by float values

Let have time span $\tau = [0; T]$, $T \in R^{1,+}$. On the period τ by the system of open sets

$$\Gamma = \{T_i\}_{i=1,2,\ldots},\ T_i = [t \mid t_{i-1} \le t < t_i,\ t_{i-1}, t_i \in \tau], \quad (1)$$

with diameter $|T_i| = \sup_{t_a, t_b \in T_i} d(t_a, t_b)$ (where d – metric of the space R^1) define topology Γ. Since the period τ is closed bounded set, then in the topology of Γ is always presents disjunctive finite covering $\chi = \{T_i \mid i = 1..n\}$, that means

$$\forall i, j \in [1; n]: T_i, T_j \in \chi \wedge i \ne j \rightarrow T_i \cap T_j = \varnothing; \quad (2)$$

$$\tau = \bigcup_{i=1}^{n} T_i . \quad (3)$$

Assume that all elements of the covering χ have the same diameter equal to ε: $\forall i \in \overline{1,n} : |T_i| = \varepsilon$. It defines χ as ε -covering of the period τ.

Let have on the interval τ defined the speech signal $x(t)$, as an infinite surjection projection

$$x: \tau \rightarrow R^1. \quad (4)$$

Since $x(t)$ is infinite projection, then by the χ existence condition we obtain coverage η (optionally disjunctive) of the range of values $x(t)$

$$\eta = \{X_i \mid i = 1..n\}, \quad (5)$$

where $X_i = \{x(t)|t \in T_i\} = \{x(t)|t \in T_i\}$ – element of the coverage η. By the definition of coverage, obtain

$$x(t) = \bigcup_{i=1}^{n} X_i,\ X_i \in \eta . \quad (6)$$

In the discrete case τ is finite, so it is the compact. Moreover, all the elements T_i of the covering χ being equally powerful $\forall i \in [1,n] : |T_i| = \dim T_i = l$ are homeomorphic to each other. Then power of the discrete space τ of the coverage χ is determined by the Grassmann's formula based on the disjunctive covering η

$$T+1 = |\tau| = \left|\bigcup_{i=1}^{n} X_i\right| = \sum_{i=1}^{n} |X_i| - \left|\bigcap_{i=1}^{n} X_i\right| = nl \,. \qquad (7)$$

Value of l defines power of element X_i: $|X_i| = l$.

Let shift signal $x(t)$ on the range of τ into the positive side $x(t) = x(t) + \max x(t) - \min x(t)$ and normalize shifted signal $x(t)$ by the scheme of the min-max multiplier K

$$K = \frac{1}{\sup\limits_{i \in [1,n]} x(t)} \,. \qquad (8)$$

Using (8) considering $x(t)$ as a discrete normalized signal $x'(t) = K(x(t)),\ t \in \tau$.

For each element X_i of the η covering constructing operator $\nabla_i : \mathrm{R}^l \times \mathrm{R}^l \to \mathrm{R}^1$

$$\nabla_i = \left(\delta_i(p,q) = \frac{x(t_q)}{x(t_p)} \,\middle|\, x(t_q), x(t_p) \in X_i \right)_{p=1,q=1}^{l,l} \,. \qquad (9)$$

Here p and q are internal indexes of the discretized values of the signal $x'(t)$, which belonging to the corresponding element X_i of the covering η. The dimension of the operator (9) is equal to $\dim\nabla_i = l \times l$.

Operator ∇_i is positive defined square matrix. Let apply singular value decomposition (SVD, [2]) to the operator ∇_i

$$\nabla_i = U_i \Sigma_i V_i^* \,, \qquad (10)$$

where U_i, V_i – unitary matrix of order $l \times l$; V_i^* – conjugate transpose of the V_I; Σ_I – diagonal matrix of the order $l \times l$ which consists of the singular values $\{\sigma_{i,q} \mid q = 1..l\}$ such as $\sigma_{i,1} \geq \sigma_{i,2} \geq \ldots \geq \sigma_{i,l} \geq 0$. Since ∇_i is a matrix of the float values then $V_i^* = V_i^{\mathrm{T}}$.

Selecting the pseudoinvariants of the speech signal

To solve the task of finding the pseudoinvariants $\{y_i \mid i = 1..n\}$ for each element of the covering χ consider the equation

$$\nabla_i y_i = x_i \,, \qquad (11)$$

where $x_i = \{x(t_p) \in X_i \mid p = 1..l\}$; $y_i = (y_{i,1}, \ldots, y_{i,l})$ – l-dimensional vector which acting as a pseudoinvariant of i^{th} element of covering χ. By (11) pseudoinvariant y_i is defined as

$$y_i = \nabla_i^{-1} x_i \,, \qquad (12)$$

as ∇_i is degenerated matrix ($\det(\nabla_i) = 0$), then invertible matrix ∇_i^{-1} does not exist. In replacement, consider pseudoinvertible matrix ∇_i^{+}, which by the SVD (10) is defined as

$$\nabla_i^{+} = V_i \Sigma_i^{+} U_i^{T} \,, \qquad (13)$$

where Σ_i^{+} – matrix of order $l \times l$, which is pseudoinverted matrix to the diagonal matrix Σ_i. Since matrix Σ_i is also degenerated then matrix Σ_i^{+} obtained from matrix Σ_i by replacing all nonzero elements $\sigma_{i,q}$ to inversed $1/\sigma_{i,q}$.

Since matrix $\nabla_i^{+}\nabla_i$ is not degenerated, then pseudoinvariant of the i^{th} element of the covering χ is determined as [3]

$$y = \nabla_i^{+} x_i + (1 - \nabla_i^{+}\nabla_i)\, r_i \,, \qquad (14)$$

where $(1 - \nabla_i^{+}\nabla_i)$ – projection operator of the ∇_i operators core; r_i – same length random vector.

Note that the finding of pseudoinverted matrix can be accomplished by Moore–Penrose pseudoinverse [3, 4], which generalizes the definition of the inverse matrixes.

Speech signals quasistationary intervals construction

The essence of the problem of determining the quasistationary interval Y_i is to build a new representation of the signal $x(t)$

$$x(t) \leftrightarrow \bigcup_{i=1}^{m} Y_i \,. \qquad (15)$$

where Y_i – quasistationary interval instance presented as union of sequential elementary parts X_i.

$$Y_i = \bigcup_{j=I_i}^{m_i + I_i - 1} X_j \,. \qquad (16)$$

Here m – number of quasistationary intervals Y_i, m_i – number of elementary parts of X_i in the union (16). Obviously $m \leq n$ and $n = \bigcup_{j=1}^{m} m_i$. Power of quasistationary interval Y_i is equal to $|Y_i| = m_i l$ and is dependent on the characteristics of the speech signal; I_i - initial index of the union (16) in the η covering, while $I_1 = 1$. In fact, the set of indexes $\{I_i\}_{i=1..m}$ defines a new topology η' which, in its turn, induces a new topology χ', as appropriate topology modification of χ.

Then the main task of the problem of building quasistationary intervals is to determine the parameters m_i and I_i for specified covering χ and η. In general, this task can be solved as constructing of the topology mapping conversion f: $\chi \to \chi'$ or F: $\eta \to \eta'$.

Since the index of the first quasistationary interval Y_1 is always equal to 1, then

$$\forall i \in [2,m] : I_i = I_1 + \sum_{k=1}^{i-1} m_k = 1 + \sum_{k=1}^{i-1} m_k \,, \qquad (17)$$

and the main task reduced to only determine parameters m_i of the corresponding quasistationary interval instance Y_1 and equation (16) takes a form of:

$$Y_i = \bigcup_{j=1+\phi_1}^{\phi_2} X_j \,, \quad \phi_1 = \sum_{k=1}^{i-1} m_k \,, \phi_2 = \phi_1 + m_i \,. \qquad (18)$$

Algorithm for the m_i is the following. Let some condition defines range of similarity between the elementary parts. If this condition is executed between the part that defines the beginning of the quasistationary instance and the part next to the current included part, then increase the value of m_i, and mark the current part as included. Otherwise, start building a new quasistationary instance.

Usages of SVD characteristics for solving the task of quasistationary intervals constructing

Further, for solving the task of speech signal quasistationary intervals constructing lets define the condition value of the operator ∇_i

$$\mu_i = \mu(\nabla_i) = \|\nabla_i\| \cdot \|\nabla_i^{-1}\|, \tag{19}$$

where $\|\cdot\|$ – norm of operator ∇_i. Since Frobenius norm of matrix has form of

$$\|\nabla_i\| = \max_{1 \le q \le l} (\sigma_{i,q}), \tag{20}$$

then by the definition (14) obtain $\|\nabla_i\| = \sigma_{i,1}$. If take into account that matrix ∇_i is degenerated, then by (17) obtain

$$\|\nabla_i^{-1}\| = \max_{\substack{1 \le q \le l \\ \sigma_{i,q} \ne 0}} \left(\frac{1}{\sigma_{i,q}} \right), \tag{21}$$

then corresponding condition value to the i^{th} element of the coverage χ will be determined by

$$\mu_i = \max_{\substack{1 \le q \le l \\ \sigma_{i,q} \ne 0}} \left(\frac{\sigma_{i,1}}{\sigma_{i,q}} \right), \tag{22}$$

Further, in the space of elements X_i of the covering η can produce a metric space by Euclidean metric

$$\forall i, j \in [1, n] : d_{\mathrm{SVD}}(X_i, X_j) = |\mu_i - \mu_j|, \tag{23}$$

Using metric (23) define the condition of belonging elementary part X_j to the quasistationary interval Y_i

$$\forall j \in [I_i, I_i + m_i - 1] : d_{\mathrm{SVD}}(X_{I_i}, X_j) \le \varepsilon, \tag{24}$$

where $\varepsilon \in R^{1,+}$ – suitable difference of including elementary part X_j to the quasistationary interval Y_i or not. Number of sequential intervals X_j (starting from index I_i) defines value of m_i.

Usages of pseudoinvariants for solving task of quasistationary intervals constructing

In case of using pseudoinvariants (14), they will act as a representation of elementary part X_i

$$X_i \leftrightarrow \{y_{i1}, \ldots, y_{il}\}. \tag{25}$$

By representation (25) can produce metrical space using p-metric

$$\forall i, j \in [1, n] : d_{\mathrm{PI}}(X_i, X_j) = \sqrt[p]{\sum_z^l |y_{iz} - y_{jz}|^p}, \quad p > 0 \tag{26}$$

Then similarly to (24) belonging condition will be defined as

$$\forall j \in [I_i, I_i + m_i - 1] : d_{\mathrm{PI}}(X_{I_i}, X_j) \le \varepsilon. \tag{27}$$

Digital watermarking method

On the first step, we should split the input signal onto the same length segments [5, 6]. The single segment length depends on the speech signal parameters. In our experiments, we use segment length values of 2^8-2^{10} samples. Then determine the signal frequency band for a watermarking by digital filters. Digital filters can be implemented based on Fast Fourier transformation (FFT) or Haar wavelets. There is the most common solution.

To provide better stability watermark should be embedded to a lowest frequency as possible (more stable to converting and compressing).

Within the selected band, find the amplitude maximum, which will be a destination for the watermarked bit. In case of 0-bit is embedded the less significant bits of three left samples should be reset, in case of 1-bit – reset less significant bits of the three right samples. To extract the digital watermark all operations until finding the amplitude maximum should be repeated. Then perform sums comparison of the less significant bits of three left and right samples. Extract the marked bit based on the result.

The quality of encoding and decoding of the digital watermark is calculated using the Ratio of Correct Bits Recovered formula:

$$RoCBR = \frac{100}{K} \sum_{n=0}^{K-1} \begin{cases} 1, w_n = w_n' \\ 0, w_n \ne w_n' \end{cases}; \tag{28}$$

where w_n – n^{th} watermarked bit, w_n' – n^{th} extracted bit, K – total number of embedded bits.

The average result achieved is around 90%.

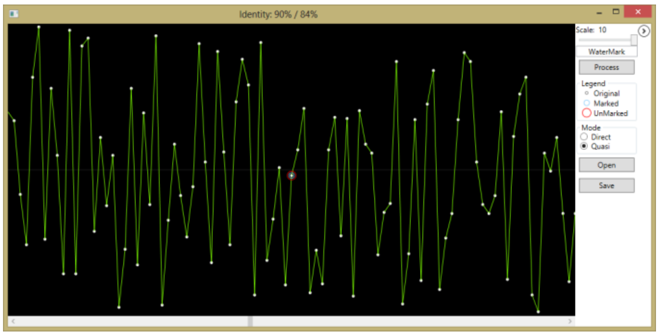

Figure 1. Example of encoding (embedding) and decoding (extracting) digital watermark (bit).

There is a white spot (speech signal item) in the center circled by green and red circles what means that specific bit was marked by the watermark and then the same (correct) bit was unmarked.

CONCLUSION

To the decision of segmentation method selection also affects subsequent task of speech signal processing. It is possible that further processing algorithms do not need such detailed segmentation.

Method described in this paper was tested on different words and phrases of speech signal. The experimental results confirm the possibility of using the proposed method for the initial segmentation of speech signal. This method can effectively extend other methods of segmentation based on selected method of voice tract. Building of quasistationary intervals are more efficient then just direct splitting or segmentation of speech signal. In addition, this method improved quality of digital watermarking and made it more robust.

ACKNOWLEDGMENT

We would like to thank our colleagues Yuriy Rashkevych and Anatoliy Kovalchuk for their inputs.

REFERENCES

[1] Wall, Michael E., Andreas Rechtsteiner, Luis M. Rocha, "Singular value decomposition and principal component analysis". In D.P. Berrar, W. Dubitzky, M. Granzow. A Practical Approach to Microarray Data Analysis. Norwell, MA: Kluwer. – 2003. – C. 91–109.

[2] D. Peleshko, A. Kovalchuk, Y. Pelekh, M. Peleshko Singular decomposition in a speech signal processing. Computer Science And Information Technologies: Materials of the Vith International Scientific and Technical Conference CSIT 2001. - Lviv: Publishig House Vezha&Co, 2011. 2011, c.19-20.

[3] Y. Rashkevych, D. Peleshko, A. Kovalchuk, M. Kupchak, Y. Pelekh Speech signal pseudo invariants. Computer Science And Information Technologies: Materials of the Vith International Scientific and Technical Conference CSIT 2001. - Lviv: Publishig House Vezha&Co, 2011. 2011, c.21-22.

[4] Roger Penrose, “A generalized inverse for matrices.” Proceedings of the Cambridge Philosophical Society 51, 406-413 (1955).

[5] Dmytro Peleshko, Yuriy Pelekh, Ivan Izonin, "Digital Watermarking of Speech Signals", CSIT’2012, p.108, November 2012.

[6] Dmytro Peleshko, Yuriy Pelekh, Ivan Izonin, "Digital Watermarking of Speech Signals Based On Quasistationary Areas", CADS’2013, p.283, 19-23 February 2013.

SHORT BIODATA OF ALL THE AUTHOR

Dr. Dmytro Peleshko is Professor at Lviv Polytechnic National University, Ukraine. He has published more than 100 papers in international and national scientific issues and journals. His research work based on digital signal processing. In addition, he is an author of monograph by the topic of his doctoral dissertation. He also is a supervisor of four aspirants.

Yuriy Pelekh is a Ph.D. student at Lviv Polytechnic National University, Ukraine. He has B.Sc. degree in Computer Science and M.Sc. degree in Automated Control Systems. His research area is speech signal reconstruction based on singular value decomposition. Currently he has published more than 10 papers in scientific journals and has presented 8 research papers at international conferences. His interests are digital photography and software engineering.

Ivan Izonin is a Ph.D. student at Lviv Polytechnic National University, Ukraine. He has B.Sc. degree in Computer Science and M.Sc. degree in Automated Control Systems. His research work based on changing image scale without quality degradation based on Zernike moments and invariants. He has published several papers in scientific journals and has participated in various international and national conferences. Currently he has been working as an Assistant at Department of Information Technology Publishing.

