UDC 057.087.1:621.391.26

ESTIMATION OF FORMANT INFORMATION USING AUTOCORRELATION FUNCTION OF VOICE SIGNAL

Pastushenko M. S. – PhD, Professor, Professor of V.V. Popovskyy Department of Infocommunication Engineering, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Pastushenko M. A. – Student of Department of Mathematical Methods for System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.

Faizulaiev T. A. – Student of Department of Theoretical Mechanics, Mechanical Engineering and Robotic Systems, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine.

ABSTRACT

Context. The current scientific problem of extracting biometric characteristics of a user of a voice authentication system, which can significantly increase its reliability, is considered. There has been performed estimation of formant information from the voice signal, which is a part of the user template in the voice authentication system and is widely used in the processing of speech signals in other applications, including in the presence of interfering noise components. The work is distinguished by the investigation of a polyharmonic signal.

Objective. The purpose of the work is to develop procedures for generating formant information based on the results of calculating the autocorrelation function of the analyzed fragment of the voice signal and their subsequent spectral analysis.

Method. The procedures for generating formant information in the process of digital processing of voice signal are proposed. Initially, the autocorrelation function of the analyzed fragment of the voice signal is calculated. Based on the results of the autocorrelation function estimation, the amplitude-frequency spectrum is calculated, from which the formant information is extracted, for example, by means of threshold processing. When the signal-to-noise ratio of the analyzed voice signal fragment is low, it is advisable to iteratively calculate the autocorrelation function. The latter allows increasing the signal-to-noise ratio and the efficiency of formant information extraction. However, each subsequent iteration of the autocorrelation function calculation is associated with an increase in the required computational resource. The latter is conditioned by the doubling of the amount of processed data at each iteration.

Results. The developed procedures for generating formant information were investigated both in the processing of model and experimental voice signals. The model signals had a low signal-to-noise ratio. The proposed procedures allow to determine more precisely the width of the spectrum of extracted formant frequencies, significantly increase the number of extracted formants, including cases at low signal-to-noise ratio.

Conclusions. The conducted model experiments have confirmed the performance and reliability of the proposed procedures for extracting formant information both in the processing of model and experimental voice signals. The results of the research allow to recommend their use in practice for solving problems of voice authentication, speaker differentiation, speech and gender recognition, intelligence, counterintelligence, forensics and forensic examination, medicine (diseases of the speech tract and hearing). Prospects for further research may include the creation of procedures for evaluating formant information based on phase data of the processed voice signal.

KEYWORDS: autocorrelation function, authentication, voice signal, speech recognition, formant information, spectrum width.

NOMENCLATURE

$\varepsilon(\tau)$ – error that occurs when calculating the autocorrelation function;

τ – delay time;

$\overline{C}_{nn}(\tau)$ – component of the autocorrelation function conditioned by noise correlation;

$\overline{C}_{ns}(\tau)$ – component of the autocorrelation function conditioned by the correlation between the noise and the useful signal;

$\overline{C}_{sn}(\tau)$ – component of the autocorrelation function conditioned by the correlation of the useful signal and noise;

$\overline{C}_{ss}(\tau)$ – component of the autocorrelation function conditioned by the correlation of the useful signal;

$C_{XX}(\tau)$ – autocorrelation function;

k – number of elements in the processed sample;

$n(t)$ – noise component, which is a stationary process;

$s(t)$ – polyharmonic voice signal;

T – signal registration period;

t – signal registration time;

$X(t)$ – additive mixture of useful and interfering signals;

x_1 – first element of the processed sample.

INTRODUCTION

In today's digital world, where information protection is one of the most important tasks, authentication methods in information systems are becoming key components of security. When using computer systems, mobile devices, online services and electronic document management, authentication plays a crucial role in confirming users' access rights to systems and their resources.

However, the existing types of authentication have their advantages and disadvantages, and the practice of their usage in network technologies shows their low reliability. Therefore, in recent years, intensive research has been carried out on the application of biometric user at-

DOI 10.15588/1607-3274-2024-3-12

tributes in authentication systems. Among biometric authentication systems a special place belongs to voice systems, which are preferable by the criterion of cost/efficiency. Obviously, that is why the Ukrainian state-owned bank Privat is implementing voice authentication systems.

The procedures of voice authentication were based on the works on speech recognition, which dates back to the middle of the last century. The whole process of voice authentication can be divided into three stages: signal preprocessing; formation of user features; decision making about the belonging of features to a given user.

Among the features of the user voice signal at authentication the following are used: pitch frequency, formant information, cepstral and mel-frequency coefficients and others. A special place belongs to formant information, which is used not only in authentication, but also in solving a number of other tasks.

Formant refers to the resonant frequency in the human speech tract that contributes to the unique timbre and quality of a speech sound. These frequencies are created by the shape and position of the tongue, lips, and other structures of the speech tract during speech production. Formants are essential in phonetics and speech processing because they play a crucial role in distinguishing between different vowels and consonants. It is known that the computation of formant frequencies is used to recognize letters or syllables (strictly speaking – phonemes) pronounced by humans. Formant information is widely used in intelligence, counterintelligence, forensic examination, medicine (diseases of the speech tract and hearing), gender recognition. In music, formants also play an important role as they contribute to the characteristic tonal qualities of various musical instruments and vocal performance.

Formant analysis has practical applications in a variety of fields, including voice identification and authentication systems, speaker recognition, speech synthesis, virtual instrument design, and audio effects processing.

Formants play a crucial role in voice synthesis by shaping the timbre and articulation of synthesized voices, contributing to the naturalness and intelligibility of the generated speech. In voice recognition systems, formant analysis helps to identify and distinguish spoken sounds, facilitating accurate speech-to-text conversion and voice authentication.

Formants play a significant role in the formation of acoustic characteristics of speech, including regional accents. Variations in formant frequencies and resonance patterns in different languages and dialects help to distinguish regional accents, affecting the perceived pronunciation and intonation of spoken speech. Formants are also used in identifying the gender characteristics of the user.

To determine formants, frequency filters, spectral analysis, wavelet transform, neural networks, etc. are used. At the same time, spectral methods are the most widely used, but they do not always meet the requirements.

The object of study is the process of digital processing of a voice signal when generating formant information.

The subject of study is the sampling methods methods for estimation of formant information from the user's voice signal using correlation and spectral analysis.

Known estimation methods are not always effective when there is a noise component.

The purpose of the work is development and study of procedures that allow to improve the quality of formant information extraction from the voice signal of the authentication system user, especially in the presence of interference.

1 PROBLEM STATEMENT

To estimate the formant information of a voice signal in the presence of a noise component, we use the properties of the correlation function [1]. To reduce the stochastic noise component of the original signal, its autocorrelation function is used in [2, 3]. Let us consider the application of the above method to extract formant data from the voice signal information. Let the original recorded speech signal $X(t)$ be an additive mixture of a polyharmonic voice signal $s(t)$ and noise $n(t)$, which is a stationary process

$$X(t) = s(t) + n(t), \qquad (1)$$

where t is the time of signal registration. We assume that at the stage of preprocessing the analyzed signals are centered.

By definition, the autocorrelation function of the analyzed signal is calculated as follows [1]

$$C_{XX}(\tau) = \lim_{T\to\infty}\int_0^T X(t)\cdot X(t-\tau)dt, \qquad (2)$$

where T is signal recording period; τ – delay time.

Due to the distributivity of the correlation operator, we can represent $C_{XX}(\tau)$ in the form of the following summands

$$\overline{C}_{XX}(\tau) = \overline{C}_{ss}(\tau) + \overline{C}_{nn}(\tau) + \overline{C}_{sn}(\tau) + \overline{C}_{ns}(\tau). \qquad (3)$$

Let us analyze the components of the autocorrelation function. We start the analysis with the second summand $\overline{C}_{nn}(\tau)$, which, taking into account the stochasticity of noise, tends to zero, and at the point $\tau = 0$ tends to the noise variance. The third and fourth summands ($\overline{C}_{sn}(\tau)$, $\overline{C}_{ns}(\tau)$) tend to zero as the processed signals are centered. The component $\overline{C}_{ss}(\tau)$ will have a higher signal-to-noise ratio [2]. As a result, we obtain

DOI 10.15588/1607-3274-2024-3-12

$$X^{'}(\tau) = \overline{C}_{ss}(\tau) + \varepsilon(\tau), \quad (4)$$

where $\varepsilon(\tau)$ is some error, which decreases with increasing integration time T.

Here we note that the autocorrelation function of the sum of periodic signals is a periodic function with the same frequency [1, 2], i.e., it contains all frequency components of the formant signals.

2 REVIEW OF THE LITERATURE

The fundamental provisions of the acoustic spectral theory of speech were formulated in the XIX century by the outstanding German scientist G. Helmholtz, which are widely used today. Let us pay attention to the works of the famous Swedish acoustician G. Fant, who also proposed a theory of distinguishing features – a universal acoustic classification of sounds. Formant information plays a significant role in this classification.

Various methods are used to obtain individual parameters of the speaker's voice, but formant analysis provides the most robust identification characteristics. It is empirically proved that four formants are sufficient to characterize speech sounds. In most cases, the first two formants are sufficient to distinguish vowel sounds, but almost always the number of formants in the sound spectrum is greater than two, indicating more complex relationships between articulation and acoustic characteristics of the sound than if only the first two formants are considered. It is the third and fourth formants that give an idea of the individual features of the speaker's pronunciation, as they capture side resonant frequencies. Formants together with other characteristics of the speech signal represent a qualitative dynamic evaluation of the speaker [4].

Determination of formant frequencies as local peaks in the amplitude spectrum of a speech signal still has significant difficulties. These difficulties are related both to the peculiarities of sound generation in the speech path and to external conditions [5].

All modern works in the field of formant analysis can be divided into several classes. The first class includes works that consider methods of formant analysis based on a generalized mathematical apparatus. In this case, the works lack a phonetic approach and do not take into account the physiological features of the human speech tract. The main attention is paid to the procedures of digital processing of nonstationary polyharmonic (polyfrequency) signal. The procedures are mostly based on the Fourier transform. Recently, cepstral analysis, wavelet transforms [6], etc. have been used for this purpose.

Another class of methods for estimating formant information is based on taking into account the peculiarities of the human speech path. For example, [7] proposes to construct the state vector of an a priori specified dynamic system (in this case, the speech path) based on the application of a recursive filter (Kalman filter). For the LPC-transformation method, [8] additionally introduces a procedure for "smoothing" formant spikes using the Newton-Raphson fast convergence algorithm.

Recently, many works have been published, in which the issues of diagnostics of diseases of the speech tract and hearing are considered [9].

The works in which neural networks are used, are considered separately [10, 11].

3 MATERIALS AND METHODS

To explain the processing procedures, let us further proceed to digital (discrete) signals. For simplicity of explanation, we will assume that a fragment of the speech signal (line vector) X containing five samples is processed. Obtaining the first half of the elements of the autocorrelation function is connected with multiplication of this vector by a matrix, namely

$$(x_1, x_2, x_3, x_4, x_5) \cdot \begin{pmatrix} x_5 x_4 x_3 x_2 x_1 \\ 0 \;\; x_5 x_4 x_3 x_2 \\ 0 \;\; 0 \;\; x_5 x_4 x_3 \\ 0 \;\; 0 \;\; 0 \;\; x_5 x_4 \\ 0 \;\; 0 \;\; 0 \;\; 0 \;\; x_5 \end{pmatrix}. \quad (5)$$

The second half of the autocorrelation function is obtained using the following matrix operations

$$(x_1, x_2, x_3, x_4, x_5) \cdot \begin{pmatrix} x_2 x_3 x_4 x_5 0 \\ x_3 x_4 x_5 0 \;\; 0 \\ x_4 x_5 0 \;\; 0 \;\; 0 \\ x_5 0 \;\; 0 \;\; 0 \;\; 0 \\ 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \end{pmatrix}. \quad (6)$$

In (6), to obtain a square matrix, the latter is supplemented with five columns consisting of zeros. As a result, we obtain an autocorrelation function containing nine elements, which is normalized accordingly. If necessary, we can calculate the autocorrelation function from the obtained samples. Naturally, the initial vector and the structure of the used matrices are changed. The number of processed elements at each processing cycle is doubled $(2 \cdot k)$, and the number of computational operations is proportional to the value of k^2.

After calculating the autocorrelation function, a fast Fourier transform is applied to the results. As a result, we obtain the amplitude-frequency spectrum from the autocorrelation function of the voice signal. In the simplest case we can perform threshold processing of the obtained spectrum and select formant frequencies. If necessary (low signal-to-noise ratio) processing procedures are repeated.

DOI 10.15588/1607-3274-2024-3-12

4 EXPERIMENTS

To verify the performance and efficiency of the considered methodology, we perform processing of a model signal. The model signal included two harmonics: the first component had a frequency value of 250 Hz and a unit amplitude, and the second component had a frequency value of 1000 Hz and a halved amplitude. The generated signal is shown in Fig. 1.

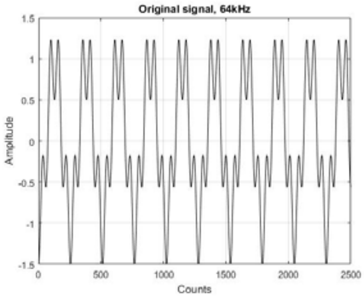

Figure 1 – Processed model signal

The sampling frequency of the considered signal is 64 kHz, the number of processed signal samples is 2500.

Then the model signal was additively supplemented with Gaussian noise. As a result, the signal-to-noise ratio by power of the processed mixture was approximately 0.7 dB. The signal received for processing is shown in Fig. 2.

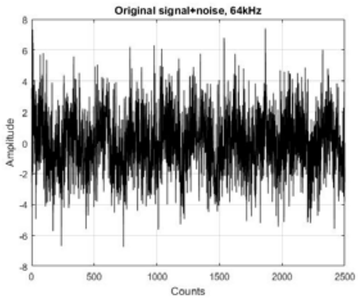

Figure 2 – Mixture of model signal and noise

Traditionally, filter-based methods have been used to determine formant information, but recently spectral procedures have been favored.

Fig. 3 shows the amplitude-frequency spectrum of the processed mixture.

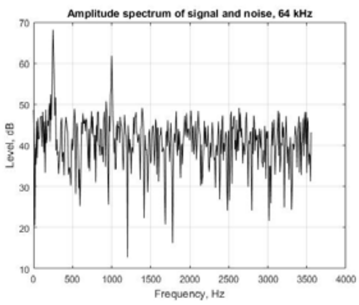

Figure 3 – Amplitude-frequency spectrum of the processed mixture

The analysis of the presented spectrum allows us to determine the frequencies of the model signal. The maxima of the spectrum correspond to the frequencies of the model signal, namely 250 and 1000 Hz. However, the width of formant frequencies is rather difficult to determine.

Further, the same mixture was processed according to the proposed methodology. The processing results after five cycles of calculating the autocorrelation function and calculating its spectrum are presented in Fig. 4. The maxima of the spectrum also correspond to the frequencies 250 and 1000 Hz. The width of the formant frequency spectrum is determined quite accurately. Thus, the proposed technique allows us to determine the frequencies of the components of the polyharmonic signal.

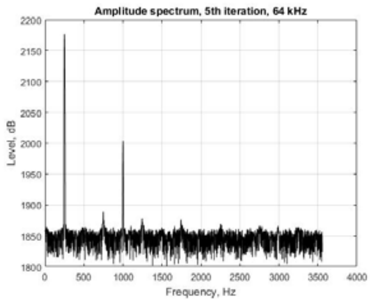

Figure 4 – Amplitude-frequency spectrum from the autocorrelation function

Analysis of the obtained spectrum makes it possible to draw a conclusion about the increase in the signal-to-noise ratio of the selected harmonics and their narrower spectrum width. The latter is essential for accurate determination of the values of format frequencies. Further we proceed to the similar processing of the experimental voice signal of the authentication system user.

5 RESULTS

Let us illustrate the results on the example of voice signal processing. The user pronounced “one” as an experimental signal. Fig. 5 shows the experimental voice signal containing 37000 samples recorded from a microphone with a sampling frequency of 64 kHz.

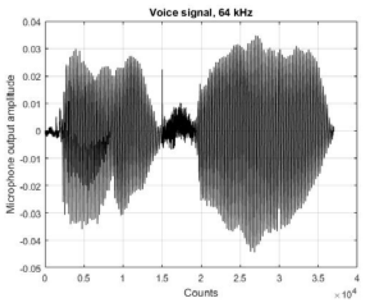

Figure 5 – Experimental voice signal

The amplitude-frequency spectrum of the analyzed signal is shown in Fig. 6.

DOI 10.15588/1607-3274-2024-3-12

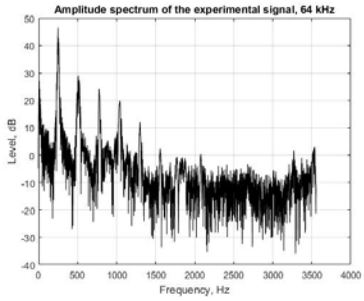

Figure 6 – Amplitude-frequency spectrum of the analyzed signal

Analysis of the spectrum shows that six maxima of formant frequencies can be identified. However, there are difficulties in determining the width of the formant spectrum and determining the frequencies of the maxima. Naturally, there will be problems with the construction of the envelope of the analyzed signal spectrum.

The results of threshold processing of this spectrum are presented in Table 1.

Table 1.

Formant number	F1	F2	F3	F4	F5
Spectrum level, dB	65	53	51	48	40
Frequency, Hz	258	524	782	1040	1297

The results of processing according to the proposed method will be illustrated during the processing of a fragment of the considered signal. Fig. 7 shows a fragment of the investigated voice signal including a part of the phoneme "o", and its spectrum is shown in Fig. 8.

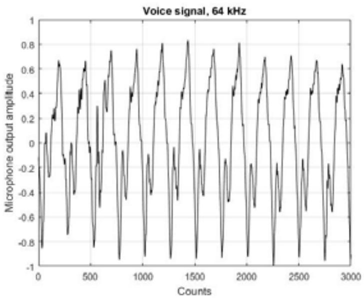

Figure 7 – Fragment of the experimental voice signal

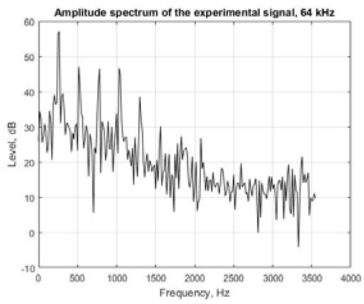

Figure 8 – Amplitude spectrum of the experimental voice signal fragment

Threshold processing (threshold value 35 dB) of the obtained spectrum allows to identify five formant frequencies. The results are presented in Table 2.

Table 2 – Spectrum thresholding results, threshold is 35 dB

Formant number	F1	F2	F3	F4	F5
Spectrum level, dB	57.1	47.1	46.5	46.7	38.57
Frequency, Hz	266	516	782	1032	1297

The results of processing a fragment of the experimental signal using the method proposed above are presented in Figs. 9–11.

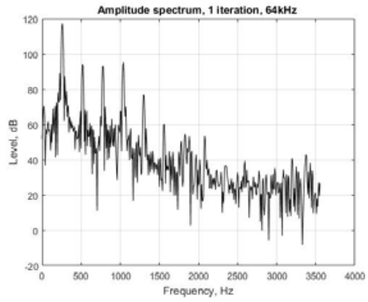

Figure 9 – Amplitude spectrum after first iteration

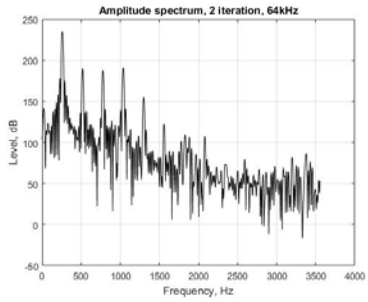

Figure 10 – Amplitude spectrum after the second iteration

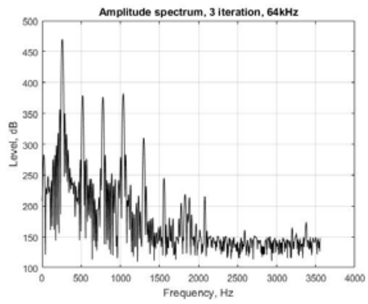

Figure 11 – Amplitude spectrum after the third iteration

The analysis of the presented spectra gives grounds to conclude that the proposed processing technique allows to increase the signal-to-noise ratio.

The results of the autocorrelation function spectrum processing after the first iteration with a threshold of 75 dB are presented in Table 3.

Table 3 – Results of spectrum thresholding, threshold 75 dB

Formant number	F1	F2	F3	F4	F5
Spectrum level, dB	117	94.2	93.1	93.3	77
Frequency, Hz	258	516	774	1032	1297

Table 4 shows the results of the autocorrelation function spectrum processing after the first iteration with a threshold of 50 dB.

DOI 10.15588/1607-3274-2024-3-12

Table 4 – Results of spectrum thresholding, threshold 50 dB

Formant number	F1	F2	F3	F4	F5	F6	F7	F8
Spectrum level, dB	117	94	93	93	77	60	55	54
Frequency, Hz	258	516	774	1032	1297	1563	1829	2079

As evidenced by the results, the proposed methodology allows to determine up to eight formant frequencies, which can positively affect the quality of voice authentication systems.

6 DISCUSSION

The use of the spectrum after the second and subsequent iterations does not give better results, however, it requires a significant computational resource. Therefore, when estimating formant frequencies, it is reasonable to limit ourselves to one iteration of the autocorrelation function calculation.

Let us briefly analyze the results of the model experiment. We begin the analysis by considering Tables 1 and 2. The first two formant frequencies are different. It is known that they are determined by the content of the voice signal. Therefore, this distinction of frequencies is justified. The third formant frequency depends on the speech apparatus of the user of the authentication system. This explains that these frequencies of the third formant are equal. The fifth formant frequency is also equal.

Now let us analyze the contents of Tables 2 and 3, namely, the results of traditional spectral methods for determining formant frequencies and the proposed methodology. The normalized correlation coefficient of both spectral density and formant frequencies is more than 0.99. However, the first three formant frequencies have a small difference. The fourth and fifth formant frequencies are the same. The second, third and fourth formant frequencies calculated by the proposed method are multiples of the first formant frequency. This fact deserves attention and requires further investigation.

CONCLUSIONS

The relevant scientific problem of development and research of procedures that allow to significantly improve the quality of formant information extraction from the voice signal of the authentication system user at low signal-to-noise ratio was solved. The object of the research is the process of generating formant information in various applications.

The scientific novelty of the obtained results lies in the fact that for the first time a method of formant information refinement is proposed on the basis of procedures for calculating the autocorrelation function of the analyzed fragment of the voice signal (mismatch function), to which traditional spectral methods of formant frequencies determination are subsequently applied. The use of the autocorrelation function as initial information, first of all, allows to increase the signal-to-noise ratio and more accurately determine the width of the formant frequency spectrum.

For the investigated calculation methods, the normalized correlation coefficient of both spectral density and formant frequencies is more than 0.99.

The results of the model experiment indicate the following.

In the process of formant frequencies estimation, calculation of one iteration of the autocorrelation function can be enough. Traditional spectral methods for determining formant frequencies do not have high accuracy, first of all, in determining the width of the formant frequency spectrum.

The proposed technique allows to determine more formant frequencies more accurately, which can significantly affect the determination of the envelope spectrum of the investigated voice signal.

The obtained results can be used in speech recognition, forensics, forensic examination, voice authentication.

Prospects for further research include the study of the proposed method for extracting formant information from the phase data of the voice signal of the authentication system.

REFERENCES

1. Van Trees H. Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory. Wiley, NewYork, 1968, 716 p. DOI:10.1109/9780470544198.ch1
2. Ifeachor Emmanuel C., Jervis Barry W. Digital Signal Processing: A Practical Approach Second Edition. Addison-Wesley Publishing Company, 1993, 779 p. URL: https://seemabaji1.wordpress.com/wp-content/uploads/2019/01/jervis-book.pdf
3. Zhuravl'ov V. Korelyatsiynyy analiz frahmentiv fonem movnoho syhnalu, *Pravove, normatyvne ta metrolohichne zabezpechennya systemy zakhystu informatsiyi v Ukrayini : naukovo-tekhnichnyy zbirnyk,* 2005, Vyp. 11, pp. 13–19. URL: https://ela.kpi.ua/handle/123456789/11249
4. Beigi H. Fundamentals of Speaker Recognition. NY, Springer, 2011, 942 p. DOI:10.1007/978-0-387-77592-0
5. Leng Y. R., Tran H. D., Kitaoka N., Li H. Selective gammatone filterbank feature for robust sound event recognition, *IEICE Transactions on Information and Systems E95.D(5)*, 2012, pp. 1229–1237. DOI:10.1587/transinf.E95.D.1229
6. Chaari S., Ouni K., Ellouze N. Wavelet ridge track interpretation in terms of formants, *Ninth International Conference on Spoken Language Processing,* 2006, pp. 1017–1020. URL: https://www.isca-archive.org/interspeech_2006/chaari06_interspeech.pdf
7. Özbek Arslan I., Demirekler M. Tracking of Visible Vocal Tract Resonances (VVTR) Based on Kalman Filtering, *9th International Conference on Spoken Language Processing/INTERSPEECH 2006, Pennsylvania, Amerika Birleşik Devletleri, 01 Ocak,* 2006, pp. 1013–1016 URI:https://hdl.handle.net/11511/53260
8. Mellahi T., Hamdi R. LPC-based formant enhancement method in Kalman filtering for speech enhancement, *AEU-International Journal of Electronics and Communications,* 2015, Vol. 69, № 2, pp. 545–554. DOI:10.1016/j.aeue.2014.11.007
9. Panek D., Skalski A., Gajda J., Tadeusiewiczacoustic R. Analysis assessment in speech pathology detection, *Int. J. Appl. Math. Comput. Sci.*, 2015, Vol. 25, № 3, pp. 631–643 DOI: 10.1515/amcs-2015-0046
10. Tran D., Wagner M., Zheng T. A Fuzzy approach to Statistical Models in Speech and Speaker Recognition, *1999 IEEE Inter-*

DOI 10.15588/1607-3274-2024-3-12

national Fuzzy Systems Conference Proceedings. Korea, 1999, pp. 1275–1280. DOI:10.1109/FUZZY.1999.790085
11. Lippmann R. Neural Network Classifiers for Speech Recognition, *The Lincoln Laboratory Journal*, 1988, Vol. 1, № 1. pp. 107–124. URL: https://archive.ll.mit.edu/publications/journal/pdf/vol01_no1/1.1.7.neuralnetworks.pdf

Received 20.06.2024.
Accepted 23.08.2024.

УДК 057.087.1:621.391.26

ОЦІНКА ФОРМАНТНОЇ ІНФОРМАЦІЇ З ВИКОРИСТАННЯМ АВТОКОРРЕЛЯЦІЙНОЇ ФУНКЦІЇ ГОЛОСОВОГО СИГНАЛУ

Пастушенко М. С. – канд. техн. наук, професор, професор кафедри інфокомунікаційної інженерії ім. В.В. Поповського, Харківського національного університету радіоелектроніки, Харків, Україна.

Пастушенко М. О. – студент кафедри математичних методів системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Файзулаєв Т. А. – студент кафедри теоретичної механіки, машинознавства та роботомеханічних систем Національного аерокосмічного університету ім. М. Є. Жуковсього «Харківський авіаційний інститут», Харків, Україна.

АНОТАЦІЯ

Актуальність. Розглядається актуальне наукове завдання отримання біометричних ознак користувача системи голосової автентифікації, які дозволяють істотно підвищити її надійність. Виконано оцінка формантної інформації з голосового сигналу, яка входить в шаблон користувача системи голосової автентифікації і широко використовується при обробці мовних сигналів в інших додатках, у тому числі і за наявності шумових складових, що заважають. Особливістю роботи є те, що дослідженню піддається полігармонійний сигнал.

Мета роботи – розробка процедур формування формантної інформації на основі результатів розрахунку автокореляційної функції аналізованого фрагмента голосового сигналу та подальшого їх спектрального аналізу.

Метод. Пропонуються процедури формування формантної інформації у процесі цифрової обробки голосового сигналу. Спочатку розраховується автокореляційна функція аналізованого фрагмента голосового сигналу. На основі результатів оцінки автокореляційної функції розраховується амплітудно-частотний спектр, з якого вилучається формантна інформація, наприклад, за допомогою порогової обробки. При низькому відношенні сигнал/шум аналізованого фрагмента голосового сигналу розрахунок автокореляційної функції доцільно виконувати ітераційно. Останнє дозволяє підвищити співвідношення сигнал/шум та ефективність виділення формантної інформації. Однак кожна наступна ітерація розрахунку автокореляційної функції пов'язана зі збільшенням необхідного обчислювального ресурсу. Останнє зумовлено подвоєнням кількості даних, що обробляються при кожній ітерації.

Результати. Розроблені процедури формування формантної інформації досліджено як при обробці модельних, так і експериментальних голосових сигналів. При цьому модельні сигнали мали низьке відношення сигнал/шум. Запропоновані процедури дозволяють більш точно визначити ширину спектра вилучаємих формантних частот, значно збільшити кількість формант, що виділяються, в тому числі і при низькому відношенні сигнал/шум.

Висновки. Проведені модельні експерименти підтвердили працездатність і достовірність запропонованих процедур отримання формантної інформації як при обробці модельних, так і експериментальних голосових сигналів. Результати досліджень дозволяють рекомендувати їх до використання на практиці для вирішення завдань голосової автентифікації, розрізнення дикторів, розпізнавання мови та статі, розвідки, контррозвідки, криміналістики та судової експертизи, медицини (хвороби мовного тракту та слуху). Перспективи подальших досліджень можуть включати створення процедур оцінки формантної інформації на основі фазових даних, оброблюваного голосового сигналу.

КЛЮЧОВІ СЛОВА: автокореляційна функція, автентифікація, голосовий сигнал, розпізнавання мови, формантна інформація, ширина спектру.

ЛІТЕРАТУРА

1. Van Trees H. Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory / H. Van Trees. – Wiley : New York, 1968. – 716 p. DOI:10.1109/9780470544198.ch1
2. Digital Signal Processing: A Practical Approach Second Edition / Emmanuel C. Ifeachor, Barry W. Jervis. – Addison-Wesley Publishing Company, 1993. – 779 p. URL: https://seemabaji1.wordpress.com/wp-content/uploads/2019/01/jervis-book.pdf
3. Zhuravl'ov V. Korelyatsiynyy analiz frahmentiv fonem movnoho syhnalu / Volodymyr Zhuravl'ov // Pravove, normatyvne ta metrolohichne zabezpechennya systemy zakhystu informatsiyi v Ukrayini : naukovo-tekhnichnyy zbirnyk. – 2005. – Vyp. 11. – P. 13–19. URL: https://ela.kpi.ua/handle/123456789/11249
4. Beigi H. Fundamentals of Speaker Recognition / H. Beigi. – NY : Springer, 2011. – 942 p. DOI:10.1007/978-0-387-77592-0
5. Selective gammatone filterbank feature for robust sound event recognition / [Y. R. Leng, H. D. Tran, N. Kitaoka, H. Li] // IEICE Transactions on Information and Systems E95.D(5). – 2012. – P. 1229–1237. DOI:10.1587/transinf.E95.D.1229
6. Chaari S. Wavelet ridge track interpretation in terms of formants / S. Chaari, K. Ouni, N. Ellouze // Ninth International Conference on Spoken Language Processing. – 2006. – P. 1017–1020. URL: https://www.isca-archive.org/interspeech_2006/chaari06_interspeech.pdf
7. Özbek Arslan I. Tracking of Visible Vocal Tract Resonances (VVTR) Based on Kalman Filtering / I. Özbek Arslan, M. Demirekler // 9th International Conference on Spoken Language Processing/INTERSPEECH 2006, Pennsylvania, Amerika Birleşik Devletleri, 01 Ocak, 2006. – P. 1013–1016 URI:https://hdl.handle.net/11511/53260
8. Mellahi T. LPC-based formant enhancement method in Kalman filtering for speech enhancement / T. Mellahi, R. Hamdi // AEU-International Journal of Electronics and Communications. – 2015. – Vol. 69, № 2. – P. 545–554. DOI:10.1016/j.aeue.2014.11.007
9. Analysis assessment in speech pathology detection / [D. Panek, A. Skalski, J. Gajda, R. Tadeusiewiczacoustic] // Int. J. Appl. Math. Comput. Sci. – 2015. – Vol. 25, № 3. – P. 631–643. DOI: 10.1515/amcs-2015-0046
10. Tran D. A Fuzzy approach to Statistical Models in Speech and Speaker Recognition / D. Tran, M. Wagner, T. Zheng // 1999 IEEE International Fuzzy Systems Conference Proceedings. – Korea, 1999. – P. 1275–1280. DOI:10.1109/FUZZY.1999.790085
11. Lippmann R. Neural Network Classifiers for Speech Recognition / R. Lippmann // The Lincoln Laboratory Journal. – 1988. – Vol. 1, № 1. – P. 107–124. URL: https://archive.ll.mit.edu/publications/journal/pdf/vol01_no1/1.1.7.neuralnetworks.pdf

DOI 10.15588/1607-3274-2024-3-12