Analysis of Voice Signal Phase Data Informativity of Authentication System User

Mykola Pastushenko[1[0000-0003-2664-1167]], Yana Krasnozheniuk[1[0000-0001-9884-0275]], Oleksandr Lemeshko[1[0000-0002-0609-6520]]

[1] Kharkiv National University of Radio Electronics, Kharkiv, 14 Nauky Ave., UKRAINE

mykola.pastushenko@nure.ua
yana.krasnozheniuk @nure.ua
oleksandr.lemeshko@nure.ua

Abstract. Directions of improving the quality characteristics specific to voice authentication systems in various access systems are analyzed and explored in the article. One of the main directions for improving the quality characteristics of the user authentication systems is the use of phase information of a voice signal. The urgent scientific task of studying new procedures is being solved to refine the estimates of the pitch frequency obtained on the basis of the amplitude-frequency spectrum analysis. The estimates were refined using phase data of the voice signal, as well as estimates of the pitch frequency in the process of obtaining cepstral coefficients. The results are obtained in the course of statistical analyzing the simulation results using experimental voice data of the authentication system user. Phase data of a voice signal allows obtaining adequate and reliable estimates in the process of spectral analysis. However, if there are errors associated with gross errors, for example, when taking the first or second formants for estimating the pitch frequency, preference should be given to the estimate obtained in the process of calculating cepstral coefficients. The presented research results should be used in voice authentication systems, improving speech recognition systems, as well as in solving speaker identification problems.

Keywords: amplitude, authentication, voice signal, information, spectrum, phase, pitch frequency.

1 Introduction

In recent decades, the achievements of science and the latest infocommunication technologies more than ever determine the dynamics of economic growth, the level of population's well-being, the competitiveness of the state in the world community, the degree of ensuring its national security and equitable integration into the global economy. The rapid development and widespread use of modern information and tele-

communication systems marked the transition of mankind from the industrial society to the information society that is based on the latest communication systems, the reliability of which does not always comply with increasing requirements. The quantity, technical level and accessibility of information systems, their reliability and performance stability already determine the degree of country's development and its status in the world community, and in the near future they will undoubtedly become a decisive mark of this status.

At the same time, the process of informatization of the world community generates a complex of negative phenomena, first of all, theft of financial, informational and computing resources. Indeed, the high complexity and vulnerability of all the systems on which regional, national and global information spaces are based, as well as the fundamental dependence of state infrastructures on their stability, lead to the emergence of principally new threats that in some cases can be solved by improving access systems.

Due to the wide spreading of distributed systems in all spheres of human activity, the task of ensuring information security in such systems is acute. One of the main measures to protect financial resources, information data and computing resources is to ensure reliable user authentication.

Currently, there are many approaches to authentication and even more implementations of these approaches. However, not all classical solutions to the authentication problem are suitable for implementation in distributed systems. In addition, various types of systems present their unique requirements for authentication subsystems. Moreover, the active development of computer technology makes it easy to crack authentication algorithms that were considered reliable 10-15 years ago. For example, in 2019, the total estimated income of fraudsters obtained using bank cards in Ukraine increased from UAH 245.8 million to UAH 361.99 million (an increase of 47.3%), as reported by the Deputy Director of the Ukrainian Interbank Payment Systems Member Association "EMA", Olesya Dalnichenko. This is largely caused by the insecurity of the password protection of bank cards.

In this regard, continuous work is ongoing in the field of research and development of authentication methods. New algorithms are constantly appearing and existing ones are being improved to ensure secure user authentication. The problem of authentication of users with access to public and personal information resources is becoming increasingly relevant. This problem is especially important for open, mass telecommunication and information systems. One of the most promising areas for protecting such systems from unauthorized influence is biometric methods for identifying users. However, despite all the attractiveness, this approach is fraught with a number of serious problems.

Initially, the development and implementation of biometric systems was associated with static biometric attributes of the user (face image, papillary finger pattern and iris), which have proven themselves in forensics. However, to date, these hopes have been destroyed, primarily because of the simplicity of the fake.

Therefore, in recent years, a lot of research has been carried out in the field of application of dynamic (behavioral) biometric authentication systems. Among these biometric systems, voice authentication takes a special place, which is simple and convenient. However, like all biometric systems, voice authentication has low quality

characteristics. In this regard, intensive research is being carried out in the field of voice authentication, as evidenced by the works [1-4].

In modern voice authentication systems (VASs), the amplitude information of a polyharmonic non-stationary voice signal of a user is recorded. User authentication is carried out mainly in the process of analyzing the amplitude-frequency spectrum of registration materials [2]. The main efforts of researchers in this case are focused on search for new or improvement of existing procedures for the formation (estimation) of templates (a set of attributes – pitch frequency, formant data, cepstral coefficients, mel-frequency cepstral coefficients, linear prediction coefficients and their dynamic characteristics, etc.) of the user, as well as the development of decision rules. The following decision-making procedures are the most popular among the latter – the methods of Gaussian Mixture Model (GMM) and Support Vector Machine (SVM). For these purposes, artificial neural networks and Hidden Markov Models (HMM) are also used.

The aim of this work is to study the influence of modern achievements in digital information processing on the accuracy of evaluating individual characteristics of the analyzed voice signal in the process of forming a user template. The object of study is the process of digital processing of voice signals.

2 General problem statement

In our opinion, an increase in the quality indicators of VASs is connected, first of all, with a change in the paradigm of digital processing of registration materials, which is associated with the addition of the amplitude-frequency spectrum analysis with modern advances in digital information processing, including algorithms for recording phase data of voice signals.

Currently, there is another way to improve the quality of the VASs, which is based primarily on the use of phase information of the user voice signal. It has long been known [5] that the phase is a more informative parameter of the signal, however, it is traditionally ignored in the VASs [2].

This is caused by the fact that to obtain phase information, additional computational and algorithmic resources are needed, which are not always available in these applications. Note that earlier in radar and radio communications to obtain phase data, special bulky devices were used – phase shifters, which could not be used in the field of voice signal processing. Currently, there are specialized microcircuits or digital signal processors that are also applicable in the field of digital processing of voice signals.

In addition, there are some features of the estimation, pre-processing and use of phase data. It should be noted that at present there is no experience and practice of using the signal phase with respect to voice authentication tasks.

This is confirmed by the fact that there are only a limited number of known works where phase data were used in the processing of speech signals. For example, in [6] the relevance of using phase information in the processing of speech data was pointed out, and in [7] a phase was used to clarify the frequency characteristics of the processed voice data. In [8], a comparative analysis of the procedures for estimating the

phase relationships between the vibrations of the pitch and harmonics of speech signals was performed, which the authors propose to use for solving problems of recognizing speech sounds and identifying speakers.

The above emphasizes the relevance of studies estimating the effect of phase data on the quality characteristics of voice authentication procedures. Phase data in voice authentication can be used in several ways that are practically important for digital processing of voice signals:

- increasing the signal-to-noise ratio of the registration materials (a known direction of using the phase in radar and radio communications);
- improving the quality of the formation of attributes for traditionally used templates, for example, the pitch frequency, formant information, etc.;
- development of new procedures for the formation of template elements based on phase data [9].

3 Work-related analysis

Let us analyze the latest scientific works in the field of speech signal processing when the issues of voice authentication in infocommunication systems have become particularly relevant. It is obvious that voice identification technologies have come to the user authentication systems from forensics. The scientific basis for the use of voice identification technology in forensics was investigated and discussed in detail in [10].

The general conclusion is that the voice identification differs from identification of fingerprints, where the variations are very small, and there is no absolutely reliable method for determining whether speech signals belong to the same person. In forensics, speaker recognition can only be probabilistic, i.e. indicating the likelihood that two speech signals belong to the same person. Under conditions of an analog telephone channel, even recognition of gender or age is sometimes complicated. Due to the small sample of speech signals, the confidence interval for evaluating the likelihood of two speech recordings belonging to the same speaker is so large that an unambiguous solution is impossible.

The task of segmentation of speakers is rather close. The segmentation of speakers in the conversation flow of different speakers (audio-indexing, diarization) is necessary when marking up sound transcripts, newsgroups, radio and television shows, interviews, etc. However, as in forensics, the quality of speaker extraction is low and unacceptable for solving user authentication problems [11].

As shown in [12], the individuality of the acoustic characteristics of the voice is determined by three factors: the mechanics of the vocal folds vibrations, the anatomy of the speech tract and the articulation control system. Naturally, the voice signal propagation channel can have some influence on acoustic characteristics (for example, the influence of external noise), the effect of which in modern systems is eliminated by digital processing procedures and organizational measures. Acoustically, the style is realized in the form of a contour of the pitch frequency, the duration of words and its segments, the rhythmics of the shock segments, the duration of pauses, and the volume level [12].

The attribute space, in which a decision is made on the identity of the speaker, should be formed taking into account all factors of the speech formation mechanism: the voice source, resonant frequencies of the speech path and their attenuation, as well as the dynamics of articulation control. In particular, in [12, 13], the following parameters of the voice source are considered: the average pitch frequency, the pitch frequency-period contour, fluctuations in the pitch frequency and the shape of the excitation pulse. The spectral characteristics of the vocal tract are described by the envelope of the spectrum and its average slope, formant frequencies and their bands, long-term spectrum or cepstrum [13].

It was shown in [14] that the most important factor in voice individuality is the fundamental frequency ($F0$), followed by formant frequencies, the size of fluctuations $F0$, and the slope of the spectrum. In [15], it was suggested that the attributes associated with $F0$ provide the best separability of voices, followed by the signal energy and the duration of the segments.

In some works, the formant frequencies are considered the most important factor [16, 17]. In particular, the fourth formant is practically independent of the type of phoneme and characterizes the tract [17].

The speaker recognition method is dominated by the cepstral method of transforming the spectrum of voice signals, which was first proposed in [18].

Cepstrum describes the shape of the envelope of the signal spectrum, which integrates the characteristics of the excitation sources (voice, turbulent and pulsed) and the shape of the speech tract. In experiments on subjective speaker recognition, it was found that the envelope of the spectrum strongly affects voice recognition [19]. Therefore, the use of a particular method of spectrum envelope analysis for speaker recognition is justified.

Instead of calculating the spectrum of the speech signal using the discrete Fourier transform over a short time interval, the amplitude-frequency characteristic of the signal found from the coefficients of linear speech prediction can also be used [20].

In [21], three informative areas were found: 100-300 Hz (the influence of a voice source), 4-5 kHz (pear-shaped cavities) and 6.5-7.8 kHz (possibly the effect of consonants). A small area is in the region of 1 kHz.

Due to the fact that the vast majority of speaker recognition systems use the same attribute space, for example, in the form of cepstral coefficients, their first and second differences, much attention is paid to the construction of decision rules, which were discussed above.

The development and application of the GMM method was considered in [22, 23]. The GMM method can be considered as an extension of the vector quantization method [23]. Vector quantization is the simplest model in speaker recognition systems, regardless of context.

The Support Vector Machines method (SVM) is actively used in various pattern recognition systems after the publication of the monograph [24]. This method allows to build a hyperplane in a multidimensional space that separates two classes, for example, the parameters of the target speaker and the parameters of speakers from the reference base. The hyperplane is calculated using not all parameter vectors, but only

specially selected ones. These vectors are called reference vectors. Since the dividing surface in the initial parameter space does not necessarily correspond to a hyperplane, a nonlinear transformation of the space of the measured parameters into a certain attribute space a higher dimension is performed. This nonlinear transformation must satisfy the requirement of linear separability in the new attribute space. If this condition is satisfied, then the dividing surface in the hyperplane is constructed using the support vector method. Obviously, the success of applying the support vector method depends on how well the non-linear transformation is selected in each specific case when recognizing speakers.

The Support Vector Machines method is used to verify speakers often in combination with the GMM or the HMM method.

The method of Hidden Markov Models (HMM) is also applied to speaker recognition, which has proven itself in problems of automatic speech recognition [25, 26]. In particular, it is assumed that for short phrases of a few seconds duration for a contest-dependent approach, it is best to use phoneme-dependent HMMs, rather than models based on the transition probabilities from frame to frame lasting 10 to 20 ms. The Hidden Markov models method can be used together with the GMM method.

The general conclusion from the analysis of well-known literature is that the templates for authentication (speaker recognition) are formed on the basis of digital processing of the amplitude-frequency spectrum of the user voice signal. At the same time, a more informative parameter of the user voice data is ignored, namely, the phase-frequency spectrum. This could be a promising area.

4 Methods and research results

We will analyze the experimental voice signal of the authentication system user, who pronounced the word “one”. The sampling frequency is 64 kHz and the signal-to-noise ratio is more than 20 dB. The analyzed voice signal is presented in Fig. 1.

Further, as in the well-known modern VASs, we calculate the amplitude-frequency spectrum from the experimental voice signal and perform its analysis. In this case, as indicated above [21], we will focus on the low-frequency region where the attributes of the authentication system user are located, focusing mainly on the pitch frequency and the associated formant frequencies.

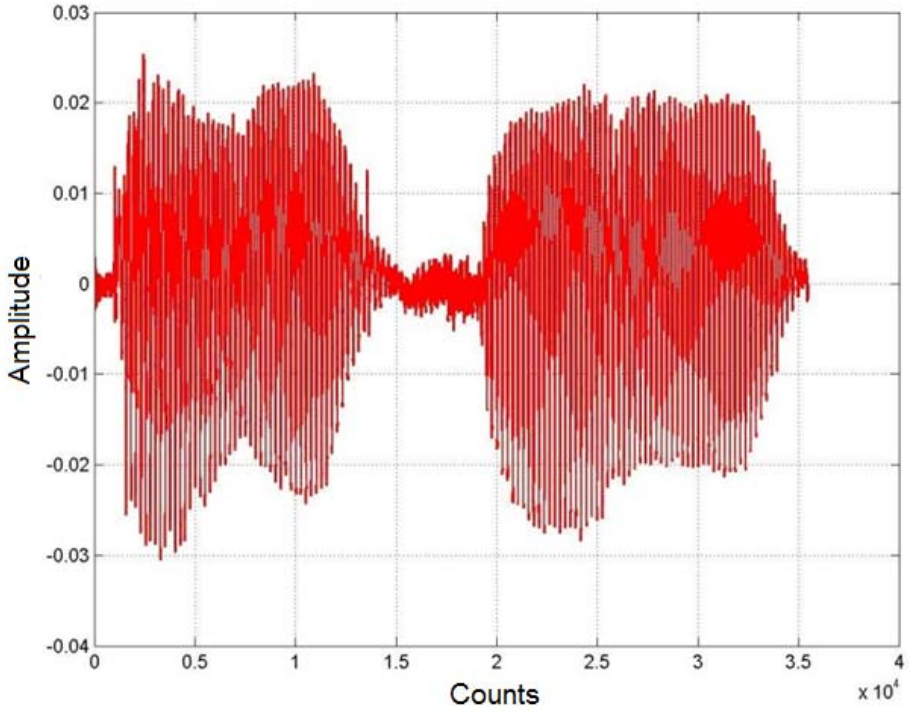

Fig. 1. Voice signal of the word “one”

It is known that the value of the pitch frequency is an individual characteristic of the speaker. It can vary depending on the emotional coloring of speech, but within fairly narrow limits. With parametric coding of speech, it is assumed that the pitch frequency of a person lies in the range of 80-400 Hz, and most formant frequencies are $F0$-fold.

The amplitude-frequency spectrum of the analyzed signal is presented in Fig. 2.

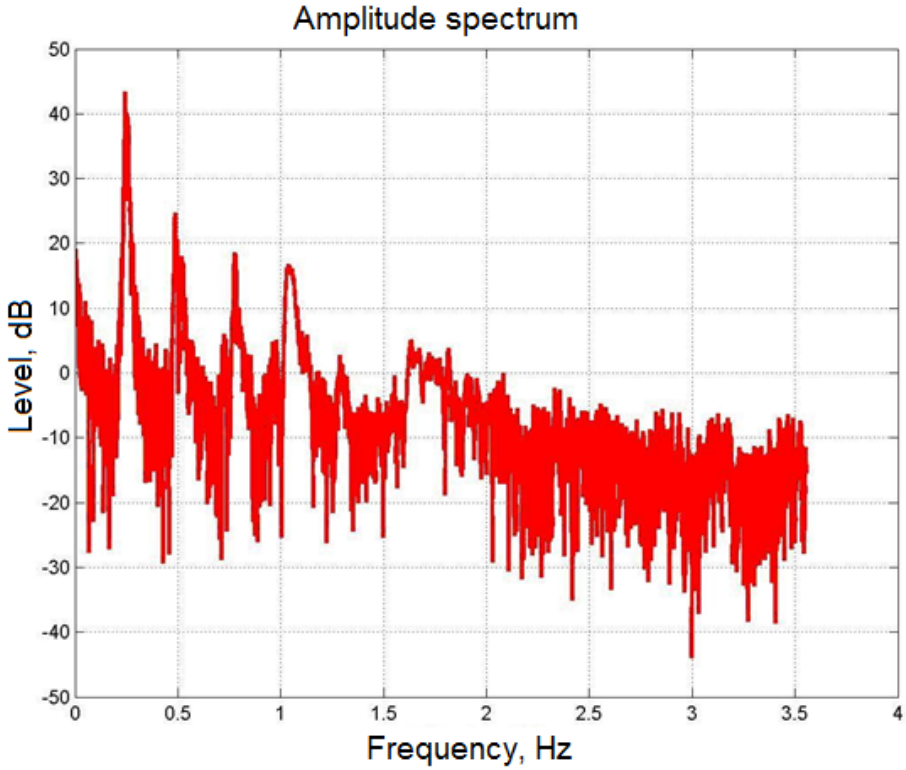

Fig. 2. A short range of voice signal “one”

Spectral analysis of the amplitude-frequency spectrum of the user's real voice signal made it possible to obtain an estimate of the pitch frequency in the region of 243 Hz. In this case, three formant frequencies are clearly pronounced (see Table 1), and the next ones have a low level of intensity.

Table 1. Characteristics of the amplitude spectrum formants

Level, dB	24.6	18.6	14.2
Frequency, Hz	243	486	776

Now we examine the characteristics considered with respect to the phase information of the voice signal of the authentication system user. For this, it is necessary to generate phase data that are not registered for a voice signal.

Therefore, phase data, as a rule, are calculated programmatically and algorithmically. To do this, it is necessary to restore the quadrature (imaginary) component of the voice signal from the registration materials. These procedures are associated with the application of the Hilbert transform [5].

$$y(t) = \frac{1}{2} \int_{-\infty}^{\infty} \frac{x(\tau)}{\pi(t-\tau)} d\tau ,$$

where $x(t)$ is the recorded voice signal; $y(t)$ is the quadrature (imaginary) component of the analytical signal; t is an independent variable that has the physical meaning of a unit of time; τ is an integration variable. Next, we can calculate the phase of the voice signal using the following ratio

$$\varphi(t) = arctg \frac{y(t)}{x(t)} .$$

Unfortunately, the function $arctg$ gives angle values ranging from $-\pi/2$ to $\pi/2$. To determine the correct value of the phase angle, which for a voice signal varies from 0 to $2 \cdot \pi$, it is necessary to adjust the angle $\varphi(t)$ accordingly, taking into account the signs of the numerator and denominator in the ratio of the function $arctg$. Otherwise, the phase spectrum will be incorrect. After correction, we obtain a phase angle, which has the form of a sawtooth signal of unknown duration.

As the results of previous studies [27] showed, after the formation of phase data, it is necessary to perform the procedures of their preliminary processing. This is due to some factors, among which we highlight the following:

- the polyharmonic nature of the voice signal, which is processed by the Hilbert transform. The latter is oriented to work with harmonic stationary data;
- incorrect data when the components $y(t)$ or $x(t)$ in the function $arctg$ are equal to zero;
- for small values of the components $y(t)$ or $x(t)$, the latter can be lost in rounding noises.

These factors lead to the fact that both random errors and anomalous measurements can occur in sawtooth phase signals. This necessitates preliminary processing of both the voice signal and the phase data. Pre-processing can be based on a priori data about the nature of the phase change of the voice signal and will improve the quality of the characteristics formation for both existing and perspective components of templates.

Now we analyze the phase spectrum of the analyzed signal. Fig. 3 shows the phase spectrum of the corrected phase data, which we will consider below.

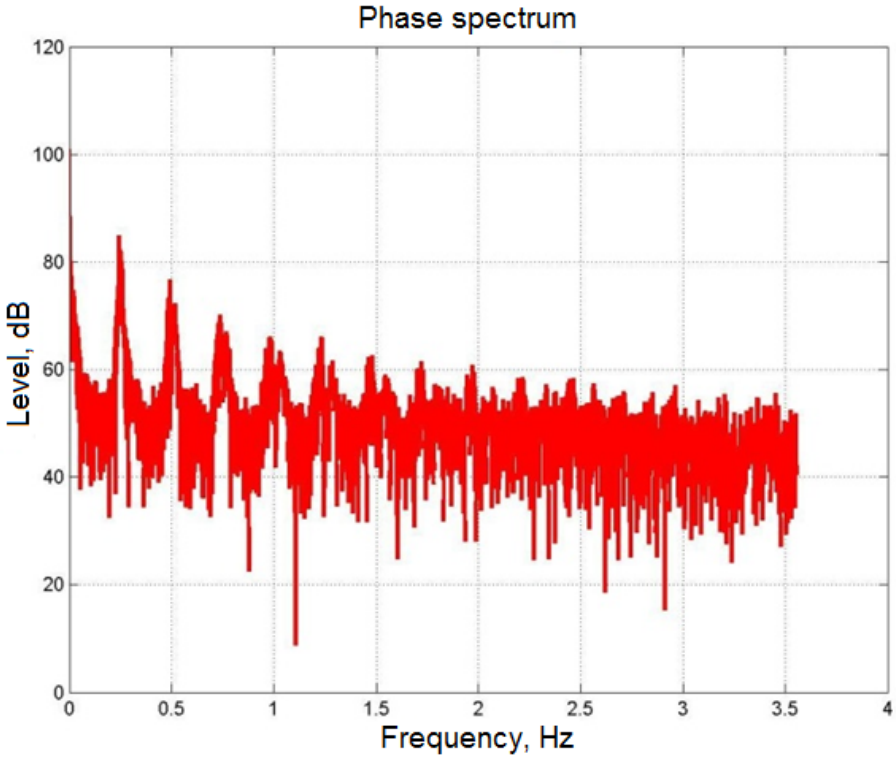

Fig. 3. A short phase spectrum of the voice signal

The results of processing the formant information of the phase spectrum are presented in Table. 2. In this spectrum, six formants can be distinguished, and the seventh and eighth have a slight energy difference. The pitch frequency, as in the amplitude spectrum, is 243 Hz.

The level of spectral density of the selected maxima is several times higher than the level of the maxima of the amplitude spectrum, which greatly simplifies the procedure for their selection. The number of selected formants in the phase spectrum is one and a half times greater. The aforementioned indicates a more informative phase spectrum of the voice signal.

Table 2. Characteristics of phase spectrum formants

Level, dB	84.9	76.7	70.3	65	64	62
Frequency, Hz	243	492	738	990	1217	1450

Another way to obtain an approximate estimate of the pitch frequency can be associated with the calculation of cepstral or mel-frequency cepstral coefficients (MFCC), which, as a rule, are included in the user template as attributes.

As it is known, cepstral coefficients are determined in accordance with the scheme presented in Fig. 4. The following notation is used in this Figure: FFT - Fast Fourier Transform block of a signal; LOG - block logarithmation spectrum; IFFT is the Inverse Fast Fourier Transform block.

Fig. 4. General scheme of cepstral signal analysis

Thus, the cepstral coefficients are the result of applying the inverse Fourier transform to the logarithmic power spectrum. The calculation of these coefficients is carried out on the samples of the signal, the duration of which are several tens of milliseconds. It is proposed to estimate the pitch frequency in each sample after performing the inverse Fourier transform.

In this case, the samples are selected with some overlap. The result of each inverse transformation allows to get an estimate of the maximum frequency in the sample. As a rule, approximately 40 coefficients are calculated, which means that we can get as many estimates of the pitch frequency. Averaging the results, we can form a more accurate estimate of the maximum pitch frequency.

The adequacy and reliability of the hypothesis put forward about a different method for estimating the pitch frequency is feasible in the process of a model experiment. To do this, we will perform digital processing according to the scheme shown in Fig. 4.

The amplitudes and phase data of the voice signal analyzed above were subjected to processing. When estimating the pitch frequency, the processed samples had an overlap coefficient of 0.75, the number of points of the discrete Fourier transform was 1024 and the Hamming smoothing width for the discrete Fourier transform.

The specifics of digital processing was the following: a sample may include vocalized or nonvocalized sounds. As we know, the pitch frequency is estimated from vocalized sounds that were extracted during the threshold processing of the spectral power level of the selected maxima.

As a result of processing the amplitude data, the following estimates were obtained: mathematical expectation is 247.5 Hz; the standard deviation is 15.7 Hz, and for phase data – 250.4 Hz and 17 Hz, respectively.

Thus, the proposed method for estimating the pitch frequency allows to get adequate and reliable results. The indicated method for estimating the pitch frequency can be useful in the presence of errors associated with gross errors. For example, taking the maximum frequencies of the first or second formants as the estimate of the pitch frequency. In this case, preference should be given to the estimate *F0* obtained in the process of calculating the cepstral coefficients.

5 Conclusion

The problem of improving the quality characteristics of voice authentication systems has been discussed in the article. As the main direction of solving this problem, it is proposed to use phase data of the analyzed voice signal in the process of digital processing. The reliability of the solution proposed for this problem and the analysis of the information content of the voice signal phase data are studied in the process of experimental evaluation of the pitch frequency and formant information, which are included in most user templates as required parameters. The cepstral or mel-frequency cepstral coefficients and a number of other attributes are additionally included in the template. The pitch frequency allows to solve the following problems: emotion recognition, gender determination, audio segmentation with multiple voices and speech separation into phrases.

In this regard, the current scientific task of studying new procedures to refine the estimates of the pitch frequency obtained on the basis of the amplitude-frequency spectrum analysis was considered in the work. The estimates were refined based on the use of phase data of the speech signal, as well as the estimates of the pitch frequency in the process of obtaining cepstral coefficients.

Therefore, the scientific novelty of the obtained results lies in the fact that for the first time a technique has been developed and experimental studies have been carried out to form an estimate of the pitch frequency (as well as formant frequencies) based on phase data of the voice signal. In addition, a new method has been developed for estimating the pitch frequency in the process of calculating cepstral coefficients. It can be performed during the analysis of both amplitude and phase data of the studied voice signal.

The results have been obtained in the process of statistical analysis of the simulation results using experimental voice data of the authentication system user.

Phase data of a voice signal allows obtaining adequate and reliable estimates in the process of spectral analysis. However, if there are errors associated with gross errors, for example, taking the maxima of the frequencies of the first or second formants as an estimate of the pitch frequency, preference should be given to the estimate obtained in the process of calculating the cepstral coefficients.

The practical significance of the results is as follows:

- a technique has been developed and features of phase information forming of the studied voice signal have been identified;
- an example of estimating the pitch frequency has shown higher informativity of the phase data, which allows selecting a larger number of formant frequencies;
- the developed method for estimating the pitch frequency eliminates gross errors in the formation of a template of the authentication system user.

It is advisable to carry out further studies in the direction of estimating the quality of the formation of attributes for traditionally used templates (for example, cepstral coefficients, mel-frequency cepstral coefficients, linear prediction coefficients, etc.) taking into account the phase of the voice signal, as well as the development of new procedures for the formation of template elements based on phase data.

References

1. Ramishvili, G.S.: Avtomaticheskoe opoznavanie govoriashego po golosu (Automatic speaker recognition over voice). Radio i sviaz, Moscow (1981).
2. Beigi, H.: Fundamentals of speaker recognition, Springer, New York (2011).
3. ISO/IEC 2382-37:2012 Information technology – Vocabulary – Part 37: Biometrics (2012).
4. Boll, R.M., Connel, J.Kh., Pankati, Sh., Ratkha, N.K., Senior E.U.: Handbook on biometry. Translated from Englsih by Agapova N. Ye. Tekhnosphera, Moscow (2007).
5. Oppenheim, A.V., Lim, J.S.: The importance of phase in signals. In: Proceeding of the IEEE, Vol. 69(5), pp. 529 - 541 (1981) doi:10.1109/PROC.1981.12022
6. Paliwal, K.: Usefulness of phase in speech processing. In: Proc. IPSJ Spoken Language Processing Workshop, Gifu, Japan, pp. 1-6 (2003).
7. Paliwal, K., Atal, B.: Frequency-related representation of speech. In: Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH-2003), pp. 65–68 (2003).
8. Borisenko, S.Yu., Vorobiev, V.I., Davidov, A.G.: Sravnenie nekotorykh sposobov analiza fazovikh sootnoshenii mejdu kvazigarmonicheskimi sostavliaushimi rechevykh signalov (Comparison of some methods for analyzing phase relationships between the quasi-harmonic components of speech signals). In: Proceedings of the First All-Russian Acoustic Conference, pp. 2-7 (2004).
9. Wu, Z., Kinnunen, T., Chng, E., Li, H., Ambikairajah, E.: A study on spoofing attack in state-of-the-art speaker verification: the telephone speech case. In: Proc. Asia-Pacific Signal Information Processing Association Annual Summit and Conference (APSIPA ASC) (2012).
10. Broeders, Ton: Forensic Speech and Audio Analysis Forensic Linguistics. In: Proceedings 13th INTERPOL Forensic Science Symposium, Lyon, France, 16-19 October 2001, D2, pp. 54-84 (2001) https://ssrn.com/abstract=2870568
11. Fergani, B., Davy, M., Houacine, A.: Speaker diarization using one-class support vector machines. Speech Communication, vol.50, pp. 355–365 (2008) doi:10.1016/j.specom.2007.11.006
12. Kuwabara, H., Sagisaka, Y.: Acoustic characteristics of speaker individuality: Control and Conversion. Speech Communication, vol.16, pp. 165-173 (1995) doi:10.1016/0167-6393(94)00053-D
13. Sorokin,V.N., Tsyplikhin, A.I.: Speaker verification using the spectral and time parameters of voice signal. Journal of Communications Technology and Electronics, v.55, N12, pp. 1561-1574 (2010) doi:10.1134/S1064226910120302
14. Matsumoto, H., Hiki, S., Sone, T., Nimura, T.: Multidimensional representation of personal quality of vowels and its acoustical correlates. In: IEEE Trans. AU, vol. AU- 21, pp. 428-436 (1973) doi:10.1109/TAU.1973.1162507
15. Shriberg, E., Ferrer, L., Kajarekar, S., Venkataraman, A., Stolcke, A.: Modeling prosodic feature sequences for speaker recognition. Speech Communication, vol.46, N3–4, pp. 455–472 (2005) doi:10.1016/j.specom.2005.02.018
16. Lavner, Y., Gath, I., Rosenhouse, J.: The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. Speech Communication vol.30, 9-26 (2000) doi:10.1016/S0167-6393(99)00028-X
17. Takemoto, H., Adachi, S., Kitamura, T., Mokhtari, P., Honda, K.: Acoustic roles of the laryngeal cavity in vocal tract resonance. J. Acoust. Soc. Am., vol.120, pp. 2228–2239 (2006) doi:10.1121/1.2261270

18. Davis, S., Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. In: IEEE Trans. Acoustics, Speech, Signal Process., vol.28, N4, pp. 357–366 (1980) doi:10.1109/TASSP.1980.1163420
19. Itoh, K.: Perceptual analysis of speaker identity. In: Speech Science and Technology, Saito S. (Ed.), IOS Press, pp. 133-145 (1992).
20. Huang, X., Acero, A., Hon, H.-W.: Spoken Language Processing: a Guide to Theory, Algorithm, and System Development. Prentice-Hall, New Jersey (2001).
21. Lu, X., Dang, J.: An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification. Speech Communication, vol.50, N4, pp. 312–322 (2007).
22. Reynolds, D.: Speaker identification and verification using Gaussian mixture speaker models. Speech Communication, vol.17, pp. 91–108 (1995).
23. Reynolds, D., Quatieri, T., Dunn, R.: Speaker verification using adapted gaussian mixture models. Digital Signal Process., vol.10, N1, pp. 19–41 (2000) doi:10.1006/dspr.1999.0361
24. Vapnik, V.N.: Statistical Learning Theory. Wiley, New York (1998).
25. BenZeghiba, M., Bourlard, H.: On the combination of speech and speaker recognition. In: Proc. Eighth European Conf. on Speech Communication and Technology (Eurospeech), pp. 1361–1364 (2003).
26. Bimbot, F., Blomberg, M., Boves, L., Genoud, D., Hutter, H.-P., Jaboulet, C., Koolwaaij, J., Lindberg, J., Pierrot, J.-B.: An overview of the CAVE project research activities in speaker verification. Speech Communication, vol. 31, pp. 155-180 (2000).
27. Pastushenko, M., Pastushenko, V., Pastushenko, O.: Specifics of receiving and processing phase information in voice authentication systems. In: 2019 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kyiv, Ukraine, pp. 621-624 (2019).