UDC 004.934

METHODS AND ALGORITHMS FOR THE ANALYSIS OF SPEECH SIGNALS

[1]*Mamatov N. S.*,[1]*Yuldoshev Yu. Sh.*, [1]*Abdullaev Sh. Sh.*, [2]*Abdurazzaqov F. B.*
m_narzullo@mail.ru; yusuf_yuldoshev@mail.ru; shash_8707@mail.ru; boss8388@mail.ru
[1]Scientific and innovation center of information and communication technologies;
[2]Department of Software Engineering, Tashkent University of Information Technology.

In recent years, many successes have been achieved in the direction of speech recognition. The results obtained make it possible to use voice dialogue instead of text dialogue, which provides speed and naturalness between man and machine. Currently, there are speech recognition systems for English, Russian, Spanish, French, Japanese Chinese and Arabic languages with high accuracy. The above systems are developed in dependence of economic and political relations. Leading speech recognition developers do not pay enough attention to other languages, in particular the Uzbek language. This article is devoted to methods and algorithms for translating Uzbek speech signals into text, the requirements of speech recognition systems, and the problems of determining the effectiveness of automatic speech recognition systems are considered. In addition, methods and algorithms for translating speech signals into text, methods for analyzing speech signals are considered. Solved the problem of analysis and processing of speech signals based on the formation of approximation structures of the correct Hadamard basis method, filtering speech signals based on the basis of discrete cosines.

Keywords: system, divided and adjacent speech, filter, smoothing, approximation, spectrum, basis, signal, polynomial.

Citation: Mamatov N.S. , Yuldoshev Yu.Sh., Abdullaev Sh.Sh., Abdurazzaqov F.B. 2019. Methods and algorithms for the analysis of speech signals. *Problems of Computational and Applied Mathematics.* 1(19): 90–99.

1 Introduction

Currently, personalized speech recognition systems are being developed by many large corporations. These systems are widely used in many areas. It is widely used in software and hardware management, stretching, teaching people with disabilities, training, and telephone conversations. The development of a speech recognition system is a very complex and high-value task. The field of artificial intelligence is rapidly developing in the direction of automatic speech recognition. In recent years, many achievements have been made in this direction. The acquired results give us opportunity to use voice communication, which is fast and natural, instead of textual communication between man and machine. At the same time, highly accurate recognition systems were developed in English, Russian, Spanish, French, Japanese, Chinese, and Arabic. The choice of these languages is related to the economic and political aspects of the development of speech technologies. Other languages are not considered properly, including the Uzbek language.

The effectiveness of existing systems is based on system requirements and many other indicators. The following is a summary of the requirements for speech recognition systems and approaches to their development.

Requirements for speech recognition systems. Speech recognition is characterized by many features, such as the properties of the voice transmission channel, the size of the dictionary, the variation of speech, the level of interference, the type of speech (divided or concatenated) [1]. The distinction between speeches allows you to slow down the conversation for a short break. It reduces the naturalness. There is no need to break a word when defining a compound word. Natural speech, unlike textual or artificial signals, does not allow an element (phoneme, word, sentence) to be simple and edfe. These elements do not have a specific physical dimension, but are separated by the understanding of the listener [2]. If the listener wants to write an unfamiliar word with a phoneme, he will make many mistakes when he divides a phrase and words. Even a person cannot distinguish speech without grammatical, vocabular and lexical knowledge. It is difficult to determine when defining word boundaries, and this boundary is determined by determining the optimal combination of optimal words with the flow of speech based on linguistic and acoustic criteria.

The complexity of the problem of speech recognition depends on many factors, the variation of the main parameters of the effect. One of them is the main component of the study of speech, which is expressed by the sole owner and expressed in many words. Changes associated with differences in individual speaking equipment are most significant. It is also affected by gender, age, dialect, emotions, and the physical condition of the speaker. In addition, the acoustic aspect, that is, switching on and replacing the microphone, as well as the acoustic state of the room, seriously deteriorate. In addition, dictionary expansion also has a significant effect on identification, which means an acoustic group of words that do not have the same acoustic value. This leads to an exponential increase in the glossary. There are several types of dictionary sizes that need to be recognized. A dictionary with one and dozens of words is called a small dictionary [3]. There are some minor issues and applications that use the dictionary. For example, you can identify numbers (telephone numbers), commands-based commands (cars, airplanes) [23], remote control systems for robots [4.5], device management systems (medical) [6] Average words usually contain hundreds of words. The numerous systems of communication or questions and answers based on these dictionaries are enormous [8]. Large dictionaries contain thousands of words [3]. These dictionaries are often used in self-help systems or automated help systems in limited areas. Extremely voluminous dictionaries contain hundreds of thousands of words that can be used when shorthand arbitrary texts [9].

The user wants to receive a quick response from the system when working in a switched system or when entering text with speech. Therefore, speech recognition requires a real-time response. There are problems identifying such characters, and the reaction time plays an important role. For example, translation of archived voice recordings into text [10]. Many of these systems are now available The most important requirements for modern automatic speech recognition systems are:

- high accuracy of recognition of complex speech;
- does not depend on the speaker;
- the presence of many words;
- high speed of work.

The choice of the optimal model and its parameters is the most urgent task of many everyday assessments of complex systems, such as speech recognition systems [11,21].

To assess the effectiveness of improved speech recognition systems, multiple criteria are used at each stage of speech processing. The accuracy of these criteria and the system response time (response) are an integral measure. An ideal automatic system should provide a good result. But at present, existing computer systems have not reached this goal, and it is important to strive to recognize speech at the human level.

2 Algorithms for speech analysis

One of the common problems of digital speech processing is the mathematical expression of this incoming signal. Information systems not only use dynamic processes, but also use a mathematical model of the incoming signal in the form of analytical functions. Efficient equipment and algorithmic processing methods are used with high accuracy to provide speech analysis, filtering, image recognition and compression.

When the signal has different interactions or the values of the table are indicated, the processing is performed by an algebraic method. Approaching a signal or parts in the form of a simple combination of simple simultaneous signals (features) or common polynomials, the subject is brought to a simplified task. For simplicity, the processed signal is characterized by the characteristic f (t), which describes the actual dynamic process, and a function of time with a limited interval. The frequency of the signal should be about 50 kHz, and the signal-to-noise ratio should not exceed one tenth. For the analytical expression of the incoming speech signal, the algebraic form (1) is somewhat favorable. It does not change the overall structure and algorithm, but only allows replacing the value of the AC coefficients and the ability to generate all the functions and numerous signals.

$$f(t) = \sum_{k}^{m} A_k t^k \tag{1}$$

The grinding of signals (polyphilic filtering) and interpolation are performed by the formula (1). To solve these problems in real time, they need methods that offer high speed, simpler and more accurate than traditional methods.

For a classical interpolation polynomial, it is not enough to solve the problem of noise signal processing. In this case, the noise generators should not exceed the approximation error of the useful signal. Otherwise, the quality of processing increases in proportion to the increase in noise.

An efficient method of average magnification for processing speech signals is a spectral algorithm for obtaining approximate speech processing signals, the least squares method. In this case (1) the mathematical model of signals is constructed as a general expression of a polynomial expression. The binary-orthogonal base system is used to convert a signal from $f\,(t)$, into the form of an algebraic polynomial (1). In practice, complex signals (in the partial polyphonic approximation) are less than one-third unavailable for polynomial signals, and polynomials greatly simplify processing and do not violate the traditional approach.

The correct method of forming approximate structures. When calculating algebraic polyphonic coefficients, the value of the undesirable variable in classical methods is used not as the value of the input signal, but as its spectral coefficient. This allows the paramagnetic shift to move from the equation to an approximate structure.

Studies have shown that Fourier analysis can be performed on all binary orthogonal (Xaara, Hartli, Discret Cosine, Fure, Adamar, Vauvlet Deboshi) basic systems. Here is an example of a kind of basic Hadamard function that will be solved. (1) Using the fast algorithm of the formula $f\,(t)$, the array of signal values is converted to the base spectrum W using an algebraic polynomial on this basis and extends to the line:

$$\sum_{j=0}^{N-1} f(t_j), W_{ij} = \sum_{j=1}^{N-1} \left(\sum_{k=0}^{m} A_k t^k \right) W_{ij} \tag{2}$$

where $i = 0, 1, 2, ..., N-1$; t $[0, 1]$; N is the number of signal values ($N = 2n$, $n = 1, 2, 3, ...$); W_ij, - matrix elements of the base function of the type± 1;

f $(t, j)$is the number of incoming signals ($j = 0, 1, 2, ..., N-1$); $k -$ polygon level. Correctly changing the two parts of the equation gives the following form:

$$P_i = \sum_{k=0}^{m} S_{k,i} \cdot A_k, i = 0, 1, 2, \ldots, N-1, \tag{3}$$

Here $P_i - W$ is the spectrum of incoming signals in the database; $S_{k,i}-$ is the spectrum of algebraic poles.

The resulting expression (3) can be depicted as a system of equations that relates algebraic polynomial coefficients with the spectral coefficients of the incoming signals.

A_k is calculated as a function of the spectrum of the incoming signals.

W (2) $-$ for the second degree

$$A_2 = 16 * P24$$

$$A_1 = -4 * P16 - 15.5 * P24$$

$$A_0 = P0 \ + 1.93 * P16 + 2.42 * P24$$

W (3) $-$ for the second degree

$$A_3 = -85.34 * P28$$

$$A_2 = 16 * P24 + 124 * P28$$

$$A_1 = -4 * P16 \ - 15.5 * \ P24 \ - 49.41 * P28$$

$$A_0 = P0 \ - 35.5 * P28 + 2.42 * P24 \ - 1.93 * P16$$

The spectral coefficients P_i, and A_k, of approximate polyphonic coefficients were fairly simple for analytical dependence. These approximation structures are distinguished by the convenient operation of signal processors. When using approximate approximation, it is necessary to take into account the limitations set above the frequency and the signal / ratio to ensure the required accuracy.

3 Editing speech signals using the Adamar database.

Adamar is based on the Adamar square matrix, where the elements are equal to "+" or "-", and the columns are displayed as orthogonal vectors. The homogeneous matrix N is calculated by the following formula.

$$H_N H_N^T = I_N \tag{4}$$

Among the orthonormal Hadamard matrices, the second-order matrix is the smallest matrix.

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{5}$$

It is known that, if N $(N > 2)$ is a homogeneous matrix, then $N4$ remains unchanged. There is no mathematical free $N-$matrix that would satisfy this requirement. But for all available N, up to 200, an appropriate matrix was created. It is often easy to create a matrix $(n - all)$ in the case when $N = 2n$.

If $HN - N$ is the order of the elementary matrix, then the matrix is the Hadamard matrix, but only the fourth and eighth order of the Hadamard matrix, which is constructed in accordance with the $2N$ order. The general rule for this is as follows:

$$H_{2^N} = \begin{bmatrix} H_N & H_N \\ H_N & -H_N \end{bmatrix} \tag{6}$$

The Hadamard matrix was proposed by Harmut as a locked structure with the fastest interpretation. When the number of changes in each row of the Hadamard matrix is divided into two, a sequence occurs. When setting the sequence $n = 2n$ of the Hadamard matrix, the change in the line takes from 0 to $N - 1$. The unitary matrices of this characteristic are called sequential matrices. Now we take any function through a framework that fits into the mathematical environment, and we change an arbitrary function by performing the following sequence.

$$\begin{gathered} H = hadamard(N) \\ x = 0 : 1/N : 1 - 1/N \\ f(x) \\ P = f * H/N \\ P = P' \end{gathered}$$

$F = H * P-$ this sequence is a general formula for the mathematical environment in speech processing using the Hadamard matrix. Here $H = Hadamard(N)$ is the order of the matrix element in the mathematical environment, and N denotes the size of the matrix;

$x = 0 : 1/N : 1 - 1/N-$ the values of the signals specified in this step;

$f(x)$ is an optional function;

$P = f * H/N$ spectral values, i.e. $f(x) * (H = Hadamard(N))/N$,

$P = P' transpose of P', P$;

$F = H * P$values of the restoration of functions or signals.

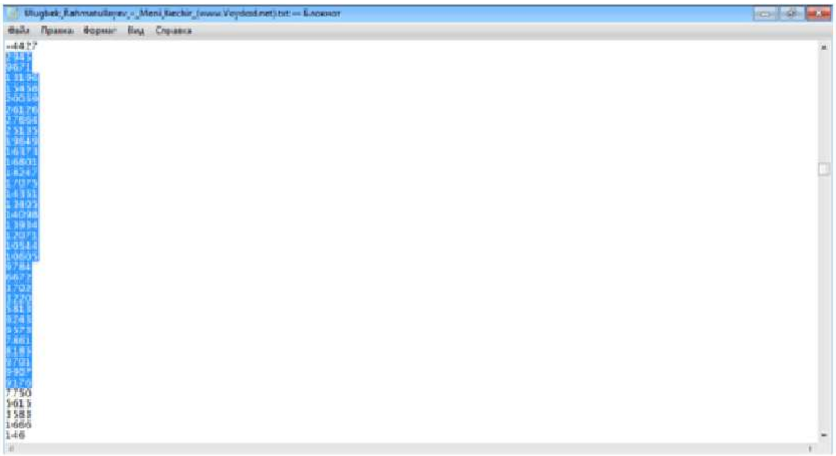
Let's now look at this sequence on an $8 * 8$ matrix, $H = Hadamard(8)$

$$\begin{gathered} x = 0 : 1/8 : 1 - 1/8 \\ f = x \\ P = f * H/8 \\ P = P' \\ F = H * P. \end{gathered}$$

Thus, the spectral values for $f = x, f = x^2, f = x^3$, are derived by a $32 * 32$ matrix. The purpose of the study is to configure the fastest algorithms and software for processing digital signals. For this, it is necessary to create a method for transmitting a fragment or its fragments to an algebraic polynomial.

To make the signal look like a polynomial, you need to use fast spectral algorithms that provide accurate and appropriate processing.

Filtering speech signals using a discrete cosine base. In most cases, spectral tactics are used to digitize speech signals. Multiplication of unwanted speech signals in the correct


Ulugbek_Rahmatullayev_-_Meni_Kechir_(www.Voydod.net).txt — Блокнот
Файл Правка Формат Вид Справка

```
-4427
3445
9671
13196
15458
[illegible]
26126
27856
23135
19649
16373
16801
14247
17075
14351
13823
14098
13934
12073
10544
10605
9784
6677
1702
3220
5813
8245
9573
7861
8185
8791
9904
9176
7750
5615
3583
1666
146
```

Figure 1 The numeral values of the audio file

and inverse exchange formulas in spectral methods by the matrix of the main functions x_i. The spectral values of the signal are generated by multiplying the signal vector by the matrix. And this is determined by the following formula

$$F_i = \sum_{m=0}^{N-1} x_i * c_{i,j} - right; \tag{7}$$

$$f_i = \sum_{m=0}^{N-1} X_i * c_{i,j} - theopposite; (eight) \tag{8}$$

where $c_{i,j}$ is the matrix of discrete change, f_i—is the vector of the incoming speech signal, F_i—is the spectral values of the speech signal. The speech signal is analyzed using formulas (7) and (8). To do this, we need to create a signal vector, which is done using the Adobe Audio program (Figure 2).

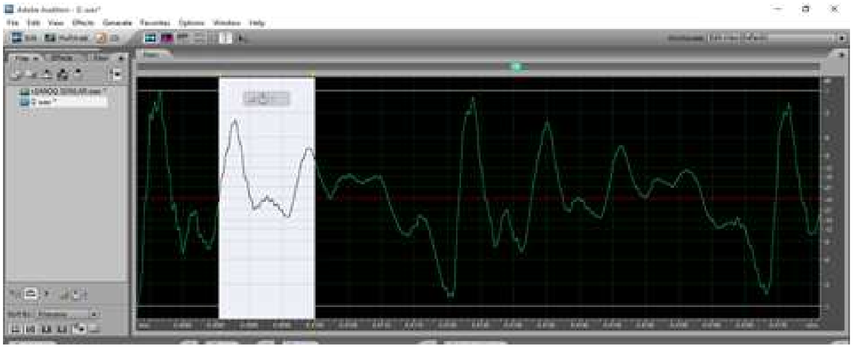


Figure 2 Signal view

For example:

SAMPLES: 64
BITSPERSAMPLE: 16
CHANNELS: 1
SAMPLERATE: 44100
NORMALIZED: FALSE

f_i =[451 4642 6029 7478 11043 12848 13531 16564 19660 20124 20555 21019 18786 15882 14654 12563 8351 5595 4929 2971 631 -138 -1385 -2993 -2783 -2043 -2310 -2033 -815 -607 -772 162 444 -641 -923 -381 -1113 -2081 -1649 -1726 -2817 -3237 -3579 -4463 -4894 -4905 -5048 -4540 -2979 -1304 596 2795 4448 6293 8653 10106 10703 11738 13081 13547 13486 13377 12477 11136].
Based on these values, we create a graphical representation of the signal (Figures 5 and 6).

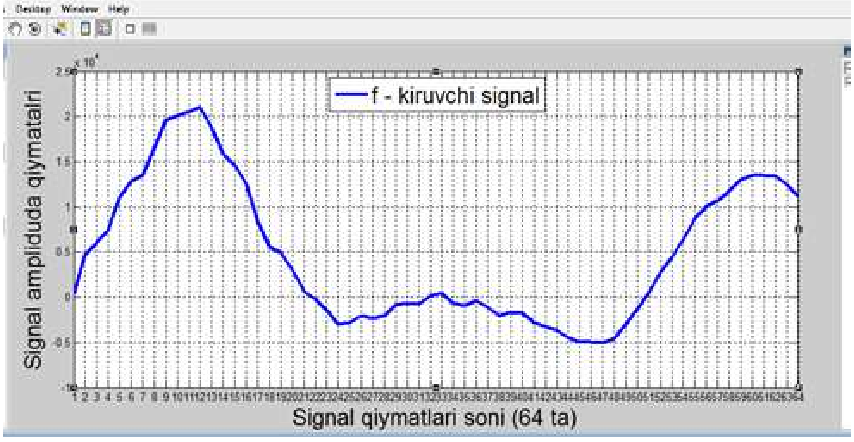


Figure 3 Graphical representation of the signal

The appearance of the speech signal is generated by the formula below

$$F_i = \sum_{m=0}^{N-1} x_i * c_{i,j}$$

where F_i is the signal spectrum

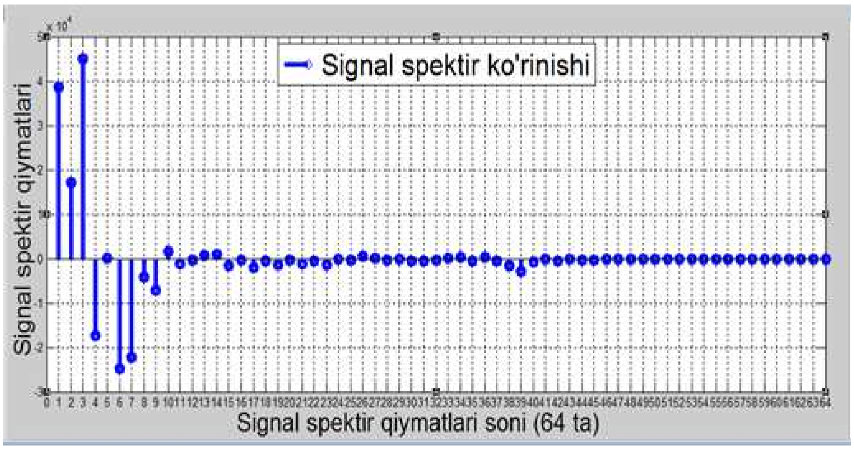


Figure 4 The signal spectrum

Below, the signal spectrum is divided into 3 parts and analyzed by spectrum. For each part, the signal is filtered out (eliminating interrupts), and the signal can be restored using formula 3.8. The refinement results are shown in Figures 8-10.

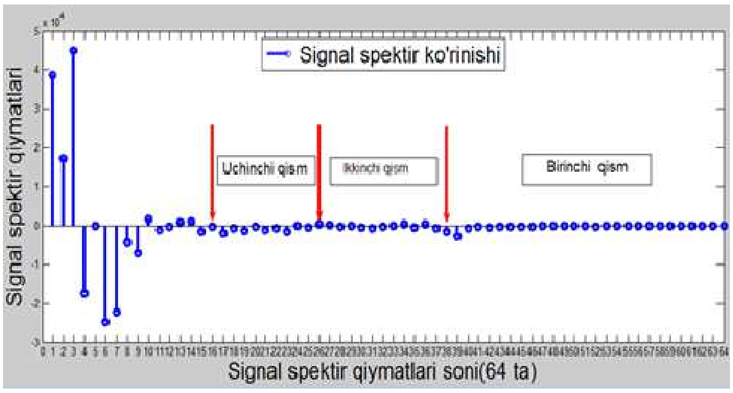


Figure 5 Breakdown of the signal spectrum

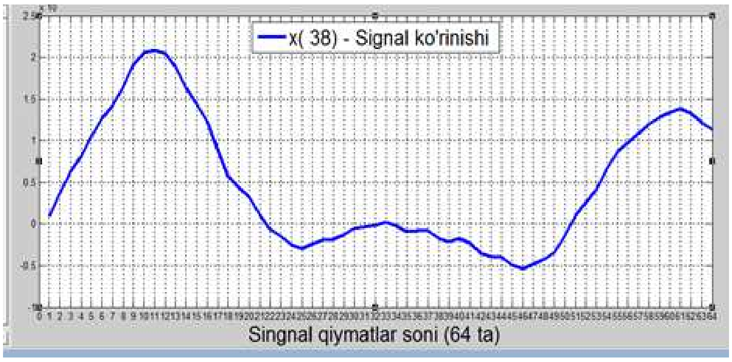


Figure 6 A reset signal for the first part

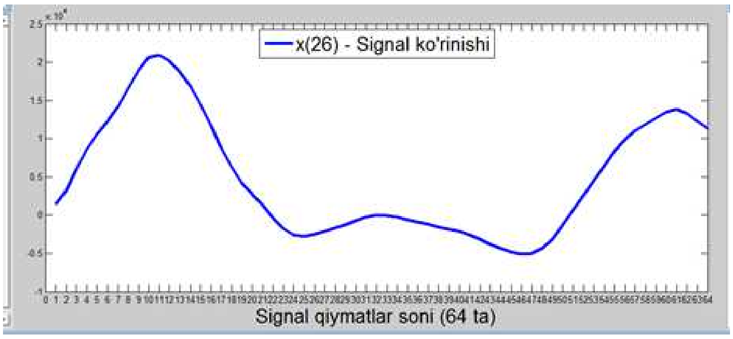


Figure 7 Image for the second part

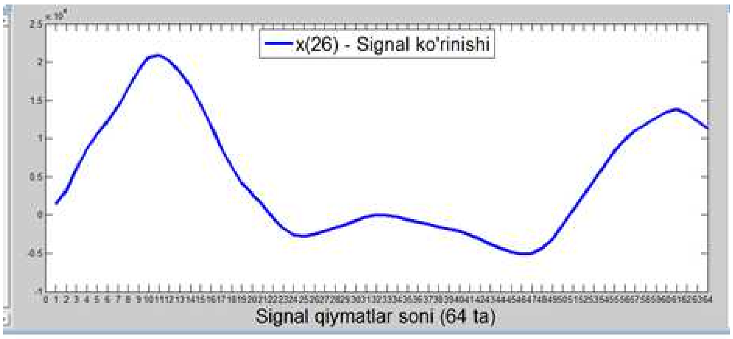


Figure 8 Signal view for the third part

4 Evaluation of the quality of algorithms

Table 6 shows the mean square error for the base of Hadamard (W) in the approximation of the properties and subtypes of the test, obtained by the least squares method and the proposed methods (N = 32).

Table 6 Function and Signal Error Levels

Function	Power	For basis W
$sin\,(\pi * /4)$	2	$1.5 \cdot 10^{-3}$
	3	$2.9 \cdot 10^{-5}$
$f = log\,(1+)$	2	$2.2 \cdot 10^{-3}$
	3	$2.5 \cdot 10^{-5}$
speech signal	2	$5.5 \cdot 10^{-5}$
	3	$3.5 \cdot 10^{-4}$

5 Concluding Remarks

The analysis of modern speech systems and their characteristics showed that these systems are necessary, and on the basis of the characteristics of the system, its speed, external interference and interference resistance were evaluated.

When analyzing speech signals, filtering, identification, compression, deletion and elimination of interruptions and quality improvements are important. These issues are solved by interpolation, approximation, discrete cosine substitution and other methods.

Vector and matrix calculations in the spectral algorithm of digital processing of speech signals allow digital processing of speech data. The study showed that the effective use of spectral methods in the expression of speech and processing of speech signals is effective.

References

[1] A. Rojin. Metod raspoznavaniya stilnoy rechi na osnove analiza signala v skolzyashem okne i teorii razmitix mnojestv., Nauchno-teoreticheskiy jurnal. "Iskusstvenniy intellekt".,- Donesk. Ukraina, 2002, S. 256-263.

[2] L. A. Stankevich Intellektualnie raboti i sistemi upravleniya. Neyro kompyuteri: razrabotka i primenenie, 8-9. 2005.

[3] I.B. Galunov. Sostoyanie issledovaniy i oblasti rechevix texnologiy i zadachi, vidigaemie gosudarstvennimi zakazchikami. doklad na seksii po avtomaticheskomu raspoznavaniyu i sentizu rechi RAN. -M., 2002.

[4] A.A. Petrovskiy. Metodi postroeniya raspoznavaniya rechi na baze gibrida neyronnaya set skretaya Markovskaya model. eyro kompyuteri: razrabotka i primenenie, 2002, 12, S. 26-36.

[5] B.V. Skolov, R.M.Yusupov. Koseptualnie osnovi otsenivaniya i analiza kachestva modeliy i polimodelnix kompleksov. Teoriya i sistemi upravleniya. -2004, 6 S. 5-16.

[6] D. Surendran, G. Levow Dialog Act Tagging with Support Vector Machines and Hidden Markov Models, Proceedings of Interspeech, Pittsburgh, PA, USA, 2006.

[7] E.A. Trentin M. Gori. Survey of hybrid ann/hmm models for automatic speech recognition. Neurocomputing, vol. 37, no. 1-4, 2001, pp. 91-126.

[8] J.P. Haton. Automatic speech recognition: Past, Present and Future. Proceedings of SPECOM'2004, St. Petersburg: “Anatoliya”, 2004, pp. 3-7.

[9] S.Potryasaev, B.Sokolov, R.Yusupov. Quality and Quantity Estimation and Analysis of Multimodal Systems for Human-Computer Interaction, Proceedings of SPECOM, St. Petersburg: “Anatoliya”, 2006, pp. 158-167.

[10] T. Hori, A. Nakamura. An extremely-large-vocabulary approach to named entity extraction from speech, Proceedings of ICASSP, Toulouse, France, 2006.

[11] A.V. Timofeev. Development of man-machine interfaces and virtual reality means for integrated medical systems, Proceedings of SPECOM, St. Petersburg: “Anatolya”, 2006, pp. 175-178.

[12] S. Osoviskiy. Neyronnie seti dlya obrabotki informatsii. -M.: Finansi i statistika, 2004, S. 344.

[13] R.K. Potapova. Rechevoe upravlenie robotom. -M.: Kom Kniga ,2005, S. 328.

Received December 16, 2018

УДК 004.934

МЕТОДЫ И АЛГОРИТМЫ АНАЛИЗА РЕЧЕВЫХ СИГНАЛОВ

[1]***Маматов Н.С.,***[1]***Юлдошев Ю.Ш.,*** [1]***Абдуллаев Ш.Ш.,***
[2]***Абдураззаков Ф.Б.***
m_narzullo@mail.ru; yusuf_yuldoshev@mail.ru; shash_8707@mail.ru;
boss8388@mail.ru
[1]Научно-инновационный центр информационно-коммуникационных технологий;
[2]Ташкентский университет информационных технологий им. М. Ал-Хоразмий.

В последние годы в направлении распознавания речевых сигналов достигнуты множество успехов. Полученные результаты дают возможность использования вместо текстового диалога голосовой, который обеспечивает быстроту и естественность между человеком и машиной. В настоящее время существуют системы распознавания речевых сигналов английского, русского, испанского, французского, японского китайского и арабского языков с высокой точностью. Вышеуказанные системы разработаны в зависимости экономических и политических отношений. Ведущие разработчики в области распознавания речи не обращают достаточное внимание другим языкам, в частности узбекскому языку. Настоящая статья посвящена методам и алгоритмам перевода речевых сигналов узбекского языка в текстовых, рассмотрены требования системам распознавания речи, а также проблемы определения эффективности систем автоматического распознавания речи. Кроме того, рассмотрены методы и алгоритмы перевода речевых сигналов в текстовых, методы анализа речевых сигналов. Решены задачи анализа и обработки речевых сигналов на основе формирования аппроксимационных структур правильного метода базиса Адамара, фильтрация речевых сигналов на основе базиса дискретных косинусов.

Ключевые слова: система, разделенная и примыкающая речь, фильтр, сглаживание, аппроксимация, спектр, базис, сигнал, многочлен.

Цитирование: *Маматов Н.С., Юлдошев Ю.Ш., Абдуллаев Ш.Ш., Абдураззаков Ф.Б.* Методы и алгоритмы анализа речевых сигналов // Проблемы вычислительной и прикладной математики. — 2019. — №1(19). — С. 90–99.