International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9, Issue-7, May 2020

Using Artificial Intelligence Algorithms for Speech Therapy Systems

Berdanov U.A, Khujayorov I.Sh, Abdurashidova K.T, Salimova Kh.R., Musadjanova D.A.

Abstract**: *This article analyzes the algorithms used to process speech signals. A structure for processing Uzbek speech signals has been proposed. Based on this structure, the effectiveness of recognizing Uzbek speech sounds, syllables and words was evaluated on the basis of intellectual algorithms. Based on the evaluated indicators, the structure of neural networks that recognize children's speech pronunciation was developed. Recognition indicators of children's speech sounds, syllables, and words were identified.***

Keywords**: *Speech, Recognition, Artificial Intelligent, Neural Network(NN), Speech Therapy System.***

I. INTRODUCTION

It is known that today speech recognition systems are used in many fields. The creation of such systems dates back to the 1950s. Today, scientists use a number of algorithms and mathematical methods in the creation of such systems, such as segmentation of speech, filtering methods, the use of spectral values, Mel Frequency Cepstral Coefficients(MFCC), algorithms of neural networks. Based on the analysis of the algorithms presented in the processing of speech signals in the Uzbek language, speech signals were divided into parameters[1]. Based on these parameters, a decision about the speech signal is required. The decision-making process is carried out by artificial intelligence algorithms. There are several types of artificial intelligence algorithms depending on the functions used.

Revised Manuscript Received on May 30, 2020.

* Correspondence Author

Berdanov Ulugbek*, Computer systems department, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan. Email: ulug-907@mail.ru

Khujayorov Ilyos, Computer systems department, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan. Email: ilyoshujayorov@gmail.com

Abdurashidova Kamola, Computer systems department, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan. Email: kamolabdurashidova@mail.ru

Salimova Khusniya, Computer systems department, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan. Email: husniyarustamovna@gmail.com

Musadjanova Dilobar, Computer systems department, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan. Email: dilobarkhon.musadzhanova.9@mail.ru

Their currently widely used algorithms for speech signals include dynamic programming algorithms, Hidden Markov models(HMM) and neural network algorithms. These artificial intelligence algorithms perform the task of making decisions based on the parameters of speech signals.

II. MODELING AND EXPERIMENTS

The main purpose of the research work is to propose a generalized model that allows to organize a set of parameters of speech signals and make decisions based on these parameters as a result of effective use in artificial intelligence algorithms and evaluate its effectiveness.

As a result of the experiments, a set of parameters based on spectral analysis of speech signals was generated. Selecting the desired parameters from this set of parameters resulted in a general model based on dynamic programming algorithms, HMM and NN. The structure of this model is shown in Figure 1.

Due to the problem posed in this model, the use of all parameters and decision-making algorithms in the model is not required[2]. That is, the parameters are selected based on the problem and the possibility of decision-making algorithms is used.

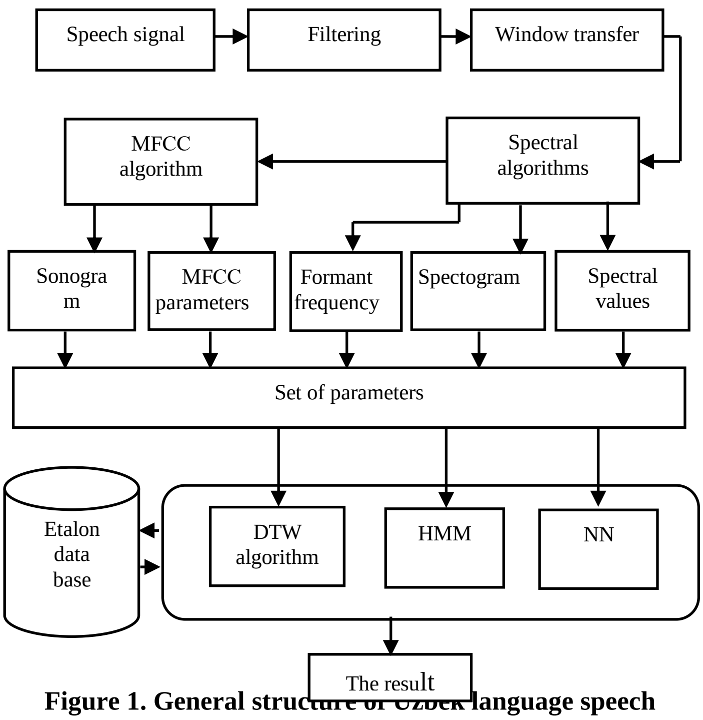

Figure 1. General structure of Uzbek language speech processing.

When recognizing speech signals, the user first performs the process of filtering the recorded speech signal in this general structure[3].

This uses the IIR filter algorithm. The filtered signal is then segmented. Segmentation performs the function of splitting the speech signal into frames. The filtered signal is then segmented. Segmentation performs the function of splitting the speech signal into frames. After that, the issue of separating the parameters from each frame is considered[4]. Parameter separation is performed using discrete Fourier transform(FT) modification from spectral algorithms. is sent to the algorithms of the block. Using spectral algorithms, the spectral values, formant frequencies, and spectrograms of a speech signal can be returned as a signal parameter[5]. It is also possible to obtain MFCC parameters using spectral algorithms and to return the sonogram as a signal parameter using the MFCC algorithm. These parameters are formed in a common set of parameters.

The selected parameters are determined by the solution of the problem based on artificial intelligence algorithms. Artificial intelligence algorithms, dynamic programming algorithms, latent marking models, and neural network algorithms were analyzed in the research process. This model has the ability to solve a number of practical problems used in human point processing[5]. As an example, identification based on human speech can be used in systems that are controlled by the content and command of human speech.

The decision-making process based on artificial intelligence algorithms in the recognition of sounds, syllables and words of the Uzbek language were the results of experiments with algorithms DTW (Dynamic time warping), HMM and neural networks.

A reference database for each artificial intelligence algorithm for Uzbek sounds has been created. The results of the analysis were obtained for 29 speech sounds (except ng) in the Uzbek language[7]. The results of artificial intelligence algorithms for vowels and consonants in the Uzbek language are given in Table 1.

Table 1. The result of artificial intelligence algorithms for sounds

№	Name	D T W	HMM	NN	№	Name	D T W	HMM	NN
		%	%	%			%	%	%
1.	**A**	85	83	90	15.	**P**	73	83	87
2.	**B**	70	76	86	16.	**Q**	73	88	92
3.	**D**	76	82	93	17.	**R**	68	75	76
4.	**E**	89	82	93	18.	**S**	60	77	86
5.	**F**	68	76	78	19.	**T**	74	66	90
6.	**G**	60	73	90	20.	**U**	87	93	90
7.	**H**	74	67	94	21.	**V**	73	77	76
8.	**I**	93	97	96	22.	**X**	67	79	87
9.	**J**	73	65	65	23.	**Y**	89	83	93
10.	**K**	68	76	97	24.	**Z**	73	76	88
11.	**L**	79	76	86	25.	**O'**	87	88	91
12.	**M**	81	80	83	26.	**G'**	68	76	83
13.	**N**	80	86	89	27.	**CH**	65	74	93
14.	**O**	86	87	98	28.	**SH**	66	61	91

The results of the analysis show that experiments on vowels and consonants have made it possible to detect 85-90% of vowels using artificial intelligence algorithms. It gave 60-85% accuracy in consonant sounds. In all of these algorithms, the MFCC coefficients of each frame were obtained as a parameter that converts to artificial intelligence algorithms. The results show that the sound recognition process yielded an average of 73% in the DTW algorithm, 85% in the HMM, and 87% in NN. This in turn proves to be effective in processing speech sounds using neural networks[8].

The results for the open and closed joints of the "R" sound when processing speech joints using these algorithms are given in Table 2.

Table 2. The result of artificial intelligence algorithms.

Sound type	Joints	Degree of detection (%)		
		D T W	HMM	NN
R	Ar	78	83	98
	Er	67	75	99
	Ir	65	76	86
	Or	71	76	87
	Ur	78	84	89
	O'r	82	92	97
	Ra	81	86	96
	Re	72	94	97
	Ri	74	68	98
	Ro	79	84	98
	Ru	76	72	96
	Ro'	75	87	99
Average:		74	81	95

In the recognition of speech joints using artificial intelligence algorithms, open and closed types of joints were obtained. The recognition accuracy of these links yielded an average of 74% in the DTW algorithm, 81% in the Markov chain and 95% in neural networks[9]. In this table, the highest result was given by neural networks. This is because of the process of training joint sounds built on the basis of neuronal tracts. The files selected for training consist of frames with a limited size.

Table 3. The result of artificial intelligence algorithms for speech words

№	Uzbek word(English translated)	D T W	HMM	NN
		%	%	%
1.	Bir(one)	84	86	96
2.	Ikki(two)	72	79	87
3.	Uch(three)	75	78	88
4.	To'rt(four)	83	83	89
5.	Besh(five)	90	96	92
6.	Olti(six)	79	58	96
7.	Yetti(seven)	87	87	87
8.	Sakkiz(eight)	85	92	83
9.	To'qqiz(nine)	81	76	92
10.	O'n(ten)	86	81	91
11.	Yigirma(twenty)	74	76	87
12.	O'ttiz(thirty)	83	90	89
13.	Qirq(forty)	76	86	92
14.	Ellik(fifty)	72	80	84
15.	Oltmish(sixty)	73	78	88
16.	Yetmish(seventy)	76	85	94
17.	Sakson(eighty)	87	92	85
18.	To'qson(ninety)	89	98	99
19.	Yuz(hundred)	92	96	99
20.	Ming(thousand)	90	95	99
21.	Million(million)	95	97	99
	Average:	82	85	91

When processing speech words based on artificial intelligence algorithms, a limited set of vocabulary-based words is required to be selected[10].

Because, based on the selected set of words, the problems of the industry are supposed to be solved as a result of research. Therefore, a set of numerical words was obtained in the experiment. Their results are presented in Table 4.

In the countdown, the results were more effective than the results based on sound and syllable. It yielded 82% in the DTW algorithm, 88% in the Markov chain, and 91% in neural networks. A general table of these results is given below (Table 4).

Table 4. The result of artificial intelligence algorithms

№	Speech information	D T W	HMM	NN
1.	Sound	73	77	87
2.	The joint	74	81	95
3.	The word	82	85	91

From the results of the analysis it can be seen that from the artificial intelligence algorithms used in Uzbek language speech recognition, it was found that neural networks are effective based on experimental results. It is also effective with computational time from other artificial intelligence algorithms due to the simplicity of the computational process in neuronal necks[11]. Experiments have shown that there are clear requirements for the training and testing process when working with neural networks. Therefore, this algorithm is more efficient than other artificial intelligence algorithms. It is recommended to use neural network algorithms in solving field problems in the processing of speech signals.

III. MATHEMATICAL ALGORITHMS AND APPLICATION

Neural network algorithms are one of the types of artificial intelligence algorithms. These algorithms have also been widely used by scientists in matters of speech recognition. To date, a number of speech recognition issues have been solved by scientists for fields based on neural crests. On the basis of neural networks, a set of speech-based speech, syllables and words based on a limited vocabulary in the recognition of sounds, syllables and words in the Uzbek language was obtained[12].

In this research paper, a speech therapist raises the issue of developing a sequence of algorithms that process children's speech in ensuring the correct pronunciation of speech in preschool and school-age children with speech impairments. It focuses on speech therapy experiments in determining the solution to the problem. Speech therapy experiments show that the repetition and pronunciation of pronounced words as a result of the occurrence of vowel and consonant sounds at the beginning, middle, and end of a word relative to words leads to the development of speech. In the processing of this type of speech data, a standard database of * .wav, mono and 22050 Hz was created based on the correct pronunciation of sounds, syllables and words by fluent children, recommended by speech therapists. The recorded data was returned as an unwanted speech signal. The main goal is to develop a structure that allows a child's speech pronunciation to be compared to sounds, syllables, and words spoken by children whose pronunciation in the database is correct. In developing such a structure, mathematical algorithms of neural networks are developed. Based on the results of the analysis and experiment, a structure based on neural networks that provide optimum high accuracy was developed. The general sequence of algorithms for this structure is shown in Fig.2.

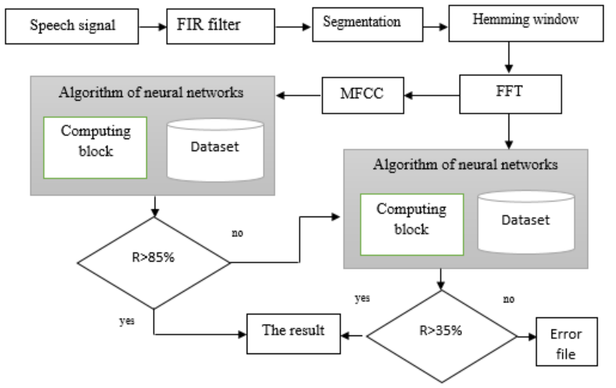

Figure 2. The general structure by which children recognize speech pronunciation.

The possibility of this structure can be used not only in the recognition of children's speech pronunciation, but also in the problems of determining the content of Uzbek speech sounds, syllables and words.

The wide range of issues that neural networks solve today limits the ability to develop a powerful, universal network that is functional with a variety of algorithms. Because each neural network is used in different resources that do not go directly to each other in the problems it solves. Therefore, in the design of neural networks, its structure is developed depending on the type of problem and the field of application. The development of this structure was organized on the basis of the following approach[13].

Speech recognition A multilayer perceptron neural network with a classical multilayer structure with the above-mentioned serial connections and sigmoidal neuron activation functions was used. The input layer of the network consists of 12 elements (MFCC parameters) X1, X2...X12, has a four-layer hidden layer and consists of a single output layer. The first hidden layer consists of F1N N = 8 elements, the second hidden layer F2N N = 6 elements, the third hidden layer F3N N = 4 elements, the fourth hidden layer F4N N = 2 elements. These dimensions are determined experimentally, and the Y dimension of the output layer always consists of 1 element. The fact that the output layer is equal to 1 is based on the child saying a single word during the speech pronunciation.

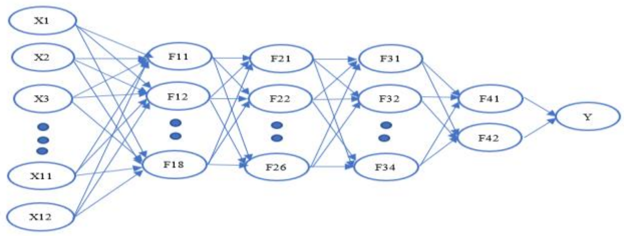

Figure 3. The general structure of the neural network used.

In the structure of a neural network, each synapse is characterized by its own synaptic connection or its weight. The current state of the neuron is determined by the randomly found sum of the values at the input.

$$F_{i,N} = \sum_{i=1}^{n} x_i w_i$$

The output of a neuron has a function for its state:

$$Y = f(F_{i,N})$$

One of the most common of these functions is the nonlinear function, i.e. the logistic function in the form f (x) or the sigmoid function.

$$f(x) = \frac{1}{1 + e^{-\alpha x}}$$

by decreasing the index a it becomes as if the sigmoid is spreading, at a = 0 a horizontal line is formed with all points equal to 0.5. and by increasing the value of a pointer, the function allows the unit to be converted to a variable by jumping the unit with the limit T at the point x = 0.

This activation function has the following advantages:

- in contrast to the logistics sigma, this function is bipolar and accordingly all signals in the neural network are also bipolar, which reduces the number of iterations of the algorithm for studying this network [14];
- Hyperbolic tangent compared to other bipolar sigmoid due to its simplicity it is computed much faster on a computer, which allows to accelerate the operation of neural networks in both recognition mode and learning mode of recording signals.

The mathematical model of the neural network used in this article is described as follows:

$$Y = f\left[\sum_{N=1}^{2} F_{4,N} f\left(\sum_{N=1}^{4} F_{3,N} f\left(\sum_{N=1}^{6} F_{2,N} f\left(\sum_{N=1}^{8} F_{1,N} X_1^{12}\right)\right)\right)\right]$$

Here F1, N - latent layer neurons, X112- MFCC coefficients, f-activation function, Y- outgoing neuron cells.

In the general structure that recognizes children's speech pronunciation, we can see that the spectra obtained based on the Fourier modification algorithm are directed directly to the neural network[15]. To do this, the structure of the neural network is constructed based on the following model.

$$Y = f\left[\sum_{N=1}^{8} F_{4,N} f\left(\sum_{N=1}^{16} F_{3,N} f\left(\sum_{N=1}^{32} F_{2,N} f\left(\sum_{N=1}^{64} F_{1,N} X_1^{128}\right)\right)\right)\right]$$

Here F1, N - latent layer neurons, X1128- spectral coefficients, f-activation function, Y- outgoing neuron cells.

In this formula we can see that the number of spectral coefficients is 128. This is because the speech signal contains 256 values when divided into frames. When spectral values are obtained by changing the Fourier transform from each frame, 128 real frequencies and 128 abstract frequencies are obtained. Only true frequencies were obtained when the spectral coefficients were directed to the neural network.

Based on the sequence of these algorithms, it is possible to determine how accurately the speech signal is recognized in children's speech development. The sequence of algorithms consists of decision-making through neural networks based on the parametric separation of the speech signal (MFCC and spectral parameters). In this case, there is an 85% high probability that the speech signal is pronounced correctly. This means that the sequence of algorithms of the Uzbek language processing model characterizes the correct organization. Experimental results show that words with a speech recognition rate of less than 85% are processed using the spectrum of the repetitive speech signal. Based on the experiments, syllables and words with less than 85% familiarity accuracy are listed in Table 5 below. We can see an increase in the level of accuracy of words as a result of the processing of syllables and words based on the signal spectrum.

Table 5. Description of speech therapy words

№	Children's speech Sound (In Uzbek language.)	Recognition (%)		Children's speech joint (In Uzbek language.)	Recognition (%)		Children's speech word (In Uzbek language)	Recognition (%)	
		MFCC+NN	Spectrum +NN		MFCC+NN	Spectrum +NN		MFCC+NN	Spectrum +NN
1.	Z	82	93	ze	55	73	Dazmol	60	88
2.	Ch	89	96	cho	42	85	Qaychi	29	83
				ich	80	99			
				uch	35	76			
3.	L	91	97	lu	70	95	Olma	67	96
							Olmaxon	37	89
4.	R	72	97	ar	78	98	Rediska	46	76
				er	57	79			
				ur	45	66	Pomidor	54	95
				or	61	87			
5.	P	92	96	ep	66	97	Paxta	77	89
6.	D	96	98	id	48	79	Daraxt	65	96
				do'	57	89			
7.	K	94	96	uk	58	86	Kartoshka	74	92
8.	G'	82	94	g'e	65	87	G'ildirak	70	98
							Qo'girchoq	61	89
9.	Q	76	97	qa	82	90	Qo'ziqorin	58	86
							Qo'zichoq	35	76
				eq	39	93	Sichqon	78	92

The research shows that when a neural network was built for the MFCC and spectral parameters of the speech signal, the neural network was able to solve this problem with high accuracy. Joints and words with a detection rate of less than 85% were re-analyzed in a neural network constructed based on spectral parameters. The results showed that a neural network built on spectral parameters was found to be more efficient than a neural network built on MFCC parameters.

IV. CONCLUSION

Given that there are several areas and areas of speech signal processing, this article presents mathematical models of algorithms used in children's speech development and their experimental results. At the same time, a model of speech recognition in the Uzbek language has been developed, the results of experimental research in the recognition of speech signals. The following problems have been solved and the results of experimental research have been obtained in solving the problems of recognizing speech signals.

1. A general model for the recognition of Uzbek speech has been developed and a general model has been developed that provides optimum high accuracy and precision in the processing of speech signals;

2. Solutions of artificial intelligence algorithms based on spectral parameters of speech were analyzed as a result of experiments.
3. The problem of recognizing speech sounds, syllables and words used in children's speech therapy for the Uzbek language was solved using artificial intelligence algorithms;
4. Speech therapy is recommended to use neural network algorithms in solving the problem of recognizing children's speech sounds, syllables and words.

APPENDIX

This research was conducted within the framework of the research project № 24-19-F "Development of computer programs for the rehabilitation of speech defects in preschool and school-age children".

ACKNOWLEDGMENT

№ 24-19-F "Development of computer programs in the rehabilitation of speech defects in preschool and school-age children" research project is to create computer programs for speech development for children with speech development or retardation among preschoolers.

REFERENCES

1. Nilanjan Dey, Intelligent Speech Signal Processing 1st Edition, Academic Press, 2019, ISBN: 9780128181300.
2. Z. Muhammad, M. Nasrun, C. Setianingsih and M. A. Murti, "Speech recognition for English to Indonesian translator using hidden Markov model," 2018 International Conference on Signals and Systems (ICSigSys), Bali, 2018, pp. 255-260.
3. Upadhyaya, S. K. Mittal, Y. V. Varshney, O. Farooq and M. R. Abidi, "Speaker adaptive model for Hindi speech using Kaldi speech recognition toolkit," 2017 International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT), Aligarh, 2017, pp. 222-226.
4. Kiss and K. Vicsi, "Investigation of cross-lingual depression prediction possibilities based on speech processing," 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen, 2017, pp. 000097-000102.
5. Lizhe Tan, Jean Jiang. Digital Signal Processing: Fundamentals and Applications, pub. "Academic Press", 2018. pp-920
6. Uday Kamath, John Liu, James Whitaker, Deep Learning for NLP and Speech Recognition, Springer, 2019
7. Xiong Xiao, Normalisaton of the speech modulation spectra for robust speech recognition , IEEE transactions on Audio, Speech and Language Processing, Vol.16,No.1 Jan 2018.
8. Musaev M, Khujayorov I, Ochilov M. Image approach to speech recognition on CNN. 2019 International Conference on Frontiers of Neural Networks (ICFNN 2019). Rome, Italy on July 26-28, 2019.
9. M.Musaev, I. Khujayorov, A. Buriboev. Accelerate the Solution of Problems of Digital Signal Processing Technology Based INTEL CILK PLUS. Asian Journal of Computer and Information Systems (ISSN: 2321-5658) Volume 03-Issue 02.
10. Musaev, M., Rakhimov, M., A Method of Mapping a Block of Main Memory to Cache in Parallel Processing of the Speech Signal // International Conference on Information Science and Communications Technologies: Applications, Trends and Opportunities, ICISCT 2019, Tashkent, 2019
11. M.M.Musaev, U.A.Berdanov, Shukurov K.E, "Hardware&Software Solution Signal Compression Algorithms Based on the Chebyshev Polynomial" International journal Information and Electronics Engineering(ISSN:2010-3719), Vol.4, No.5, September 2014, South Korea, p.380-383.
12. M.M.Musaev, U.A.Berdanov, "Parallel Speech Signal Processing on Multicore Processors" International conference on multimedia information technology and application (MITA-2015), June 30-July 2, 2015, Tashkent, Uzbekistan.
13. M.M.Musaev, U.A.Berdanov. "The Technology of Parallel Processing on Multicore Processors", International journal of Signal Processing Systems (IJSPS),Vol. 4, No. 3, June 2015, p. 252-257.
14. Rakhimov M.F., Berdanov U.A., Parallel Processing Capabilities in the Process of Speech Recognition // 2017 International conference on information science and communications technologies ICISCT, –Tashkent, 2017.
15. Ilyos Khujayorov, Ochilov Mannon. Parallel Signal Processing Based-On Graphics Processing Units. International Conference on Information Science and Communications Technologies ICISCT 2019. 4th of November 2019, Tashkent Uzbekistan. DOI: 10.1109/ICISCT47635.2019.9011976.

AUTHORS PROFILE

Berdanov Ulug'bek Abdumurodovich from Uzbekistan, was born on 22 July, 1987. He obtained Bachelor of Applied Informatics from the National University of Uzbekistan (NUU) in 2009, Master in Applied Informatics from Tashkent University of Information Technologies (TUIT) in 2012. He published many papers in different. He is currently conducting research on algorithms and models used in pediatric speech therapy. Currently, he is junior researcher and Assistant-Teacher at the department Computer systems, Tashkent University of Information Technologies (TUIT), Uzbekistan since 2012.

Khujayorov Ilyos Shiraliyevich is from Uzbekistan and was born on 28 April, 1985. He obtained bachelor of Information technologies from the Tashkent University of Information Technologies (TUIT), Uzbekistan, in 2006, master in computer science and engineering from TUIT, Uzbekistan, in 2008. Author of many scientific papers. Currently, he is a PhD student at the department Computer systems, TUIT, Uzbekistan since 2018. His current interest is in the field parallel processing, machine learning, digital signal processing and real time systems. He is currently researching digital processing of speech and digital image processing.

Abdurashidova Kamola Turgunbayevnna from Uzbekistan, was born on 26 June, 1982. She obtained Bachelor of Information Technologies from Tashkent State Technical University named after Beruni in 2003, Master in Computers, Complexes, Systems and Networks from Tashkent State Technical University named after Beruni in 2005, Docent in Mathematical and Software Support of Computers, Complexes and Computer Networks from Council of The TUIT, in 2019. She published many papers in different journals such as "International Journal of Advanced Research in Science, Engineering and Technology". Currently, she is junior researcher and docent-Teacher at the department Computer systems, Tashkent University of Information Technologies (TUIT), Uzbekistan since 2005

Salimova Khusniya Rustamovna from Uzbekistan, was born on 27 November, 1993. She is a student at the Department of Computer engineering of Tashkent university of information technologies (TUIT), Uzbekistan. Her research interests include data protecting techniques steganography, identification, authentication, face recognition, access control, parallel information processing and real time systems. She is currently researching Digital processing of speech and Digital Image processing.

Musadjanova Dilobar Abduvaxidovna from Uzbekistan, was born on 13 December, 1989. She obtained Bachelor of Tashkent State Technical University of Uzbekistan (TSTU) in 2013. Master in Tashkent State Technical University of Uzbekistan (TSTU) in 2015. Currently she is assistant -Teacher at the department Computer systems, Tashkent University of Information Technologies 2019. She is currently researching digital processing of speech and digital image processing.