

ISSN 2518-1726 (Online),  
ISSN 1991-346X (Print)

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫНЫҢ  
әл-Фараби атындағы Қазақ ұлттық университетінің

# Х А Б А Р Л А Р Ы

---

---

## ИЗВЕСТИЯ

НАЦИОНАЛЬНОЙ АКАДЕМИИ НАУК  
РЕСПУБЛИКИ КАЗАХСТАН  
Қазақстан Республикасының  
Ғылым Академиясының  
им. аль-Фараби

## NEWS

OF THE NATIONAL ACADEMY OF SCIENCES  
OF THE REPUBLIC OF KAZAKHSTAN  
Al-Farabi  
Kazakh National University

**SERIES**  
**PHYSICO-MATHEMATICAL**

**1 (335)**

**JANUARY – FEBRUARY 2021**

PUBLISHED SINCE JANUARY 1963

PUBLISHED 6 TIMES A YEAR

ALMATY, NAS RK

---

*NAS RK is pleased to announce that News of NAS RK. Series physico-mathematical journal has been accepted for indexing in the Emerging Sources Citation Index, a new edition of Web of Science. Content in this index is under consideration by Clarivate Analytics to be accepted in the Science Citation Index Expanded, the Social Sciences Citation Index, and the Arts & Humanities Citation Index. The quality and depth of content Web of Science offers to researchers, authors, publishers, and institutions sets it apart from other research databases. The inclusion of News of NAS RK. Series of chemistry and technologies in the Emerging Sources Citation Index demonstrates our dedication to providing the most relevant and influential content of chemical sciences to our community.*

*Қазақстан Республикасы Ұлттық ғылым академиясы "ҚР ҰҒА Хабарлары. Физикалық-математикалық сериясы" ғылыми журналының Web of Science-тің жаңаланған нұсқасы Emerging Sources Citation Index-те индекстелуге қабылданғанын хабарлайды. Бұл индекстелу барысында Clarivate Analytics компаниясы журналды одан әрі the Science Citation Index Expanded, the Social Sciences Citation Index және the Arts & Humanities Citation Index-ке қабылдау мәселесін қарастыруда. Web of Science зерттеушілер, авторлар, баспашылар мен мекемелерге контент тереңдігі мен сапасын ұсынады. ҚР ҰҒА Хабарлары. Химия және технология сериясы Emerging Sources Citation Index-ке енуі біздің қоғамдастық үшін ең өзекті және беделді химиялық ғылымдар бойынша контентке адалдығымызды білдіреді.*

*НАН РК сообщает, что научный журнал «Известия НАН РК. Серия физико-математическая» был принят для индексирования в Emerging Sources Citation Index, обновленной версии Web of Science. Содержание в этом индексировании находится в стадии рассмотрения компанией Clarivate Analytics для дальнейшего принятия журнала в the Science Citation Index Expanded, the Social Sciences Citation Index и the Arts & Humanities Citation Index. Web of Science предлагает качество и глубину контента для исследователей, авторов, издателей и учреждений. Включение Известия НАН РК в Emerging Sources Citation Index демонстрирует нашу приверженность к наиболее актуальному и влиятельному контенту по химическим наукам для нашего сообщества.*

Б а с р е д а к т о р ы  
ф.-м.ғ.д., проф., ҚР ҰҒА академигі  
**Ғ.М. Мұтанов**

Р е д а к ц и я а л қ а с ы:

**Асанова А.Т.** проф. (Қазақстан)  
**Бошкаев К.А.** PhD докторы (Қазақстан)  
**Байгунчеков Ж.Ж.** проф., академик (Қазақстан)  
**Quevedo Hernando** проф. (Мексика)  
**Жүсіпов М.А.** проф. (Қазақстан)  
**Ковалев А.М.** проф., академик (Украина)  
**Калимолдаев М.Н.** проф., академик (Қазақстан)  
**Михалевич А.А.** проф., академик (Белорусь)  
**Мырзакулов Р.** проф., академик (Қазақстан)  
**Рамазанов Т.С.** проф., академик (Қазақстан)  
**Такибаев Н.Ж.** проф., академик (Қазақстан), бас ред. орынбасары  
**Тигиняну И.** проф., академик (Молдова)  
**Уалиев З.Г.** проф., чл.-корр. (Қазақстан)  
**Харин С.Н.** проф., академик (Қазақстан)

**«ҚР ҰҒА Хабарлары. Физика-математикалық сериясы».**

ISSN 2518-1726 (Online), ISSN 1991-346X (Print)

Меншіктенуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.).

Қазақстан Республикасының Ақпарат және коммуникациялар министрлігінің Ақпарат комитетінде  
14.02.2018 ж. берілген № 16906-Ж мерзімдік басылым тіркеуіне қойылу туралы куәлік.

**Тақырыптық бағыты:** *физика-математика ғылымдары және ақпараттық  
технологиялар саласындағы басым ғылыми зерттеулерді  
жариялау.*

Мерзімділігі: жылына 6 рет.

Тиражы: 300 дана.

Редакцияның мекенжайы: 050010, Алматы қ., Шевченко көш., 28; 219, 220 бөл.;

тел.: 272-13-19; 272-13-18,

<http://physics-mathematics.kz/index.php/en/archive>

---

© Қазақстан Республикасының Ұлттық ғылым академиясы, 2021

Типографияның мекенжайы: «NurNaz GRACE», Алматы қ., Рысқұлов көш., 103.

Главный редактор  
д.ф.-м.н., проф. академик НАН РК  
**Г.М. Мутанов**

Редакционная коллегия:

**Асанова А.Т.** проф. (Казахстан)  
**Бошкаев К.А.** доктор PhD (Казахстан)  
**Байгунчеков Ж.Ж.** проф., академик (Казахстан)  
**Quevedo Hernando** проф. (Мексика)  
**Жусупов М.А.** проф. (Казахстан)  
**Ковалев А.М.** проф., академик (Украина)  
**Калимолдаев М.Н.** проф., академик (Казахстан)  
**Михалевич А.А.** проф., академик (Беларусь)  
**Мырзакулов Р.** проф., академик (Казахстан)  
**Рамазанов Т.С.** проф., академик (Казахстан)  
**Такибаев Н.Ж.** проф., академик (Казахстан), зам. гл. ред.  
**Тигиняну И.** проф., академик (Молдова)  
**Уалиев З.Г.** проф., чл.-корр. (Казахстан)  
**Харин С.Н.** проф., академик (Казахстан)

**«Известия НАН РК. Серия физика-математическая».**

ISSN 2518-1726 (Online), ISSN 1991-346X (Print)

Собственник: РОО «Национальная академия наук Республики Казахстан» (г. Алматы).

Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и коммуникаций Республики Казахстан № 16906-Ж, выданное 14.02.2018 г.

**Тематическая направленность:** *публикация приоритетных научных исследований в области физико-математических наук и информационных технологий.*

Периодичность: 6 раз в год.

Тираж: 300 экземпляров.

Адрес редакции: 050010, г. Алматы, ул. Шевченко, 28; ком. 219, 220; тел.: 272-13-19; 272-13-18,  
<http://physics-mathematics.kz/index.php/en/archive>

---

© Национальная академия наук Республики Казахстан, 2021

Адрес типографии: «NurNaz GRACE», г. Алматы, ул. Рыскулова, 103.

Editor in chief

doctor of physics and mathematics, professor, academician of NAS RK

**G.M. Mutanov**

Editorial board:

**Asanova A.T.** prof. (Kazakhstan)

**Boshkayev K.A.** PhD (Kazakhstan)

**Baigunchekov Zh.Zh.** prof., akademik (Kazakhstan)

**Quevedo Hemando** prof. (Mexico)

**Zhusupov M.A.** prof. (Kazakhstan)

**Kovalev A.M.** prof., academician (Ukraine)

**Kalimoldaev M.N.** prof., akademik (Kazakhstan)

**Mikhalevich A.A.** prof., academician (Belarus)

**Myrzakulov R.** prof., akademik (Kazakhstan)

**Ramazanov T.S.** prof., akademik (Kazakhstan)

**Takibayev N.Zh.** prof., academician (Kazakhstan), deputy editor in chief.

**Tiginyanu I.** prof., academician (Moldova)

**Ualiev Z.G.** prof., chl.-korr. (Kazakhstan)

**Kharin S.N.** prof., academician (Kazakhstan)

**News of the National Academy of Sciences of the Republic of Kazakhstan. Physical-mathematical series.**

ISSN 2518-1726 (Online), ISSN 1991-346X (Print)

Owner: RPA "National Academy of Sciences of the Republic of Kazakhstan" (Almaty).

The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Communications of the Republic of Kazakhstan **No. 16906-Ж**, issued on 14.02.2018.

**Thematic scope: *publication of priority research in the field of physical and mathematical sciences and information technology.***

Periodicity: 6 times a year.

Circulation: 300 copies.

Editorial address: 28, Shevchenko str., of. 219, 220, Almaty, 050010, tel. 272-13-19; 272-13-18,

<http://physics-mathematics.kz/index.php/en/archive>

---

© National Academy of Sciences of the Republic of Kazakhstan, 2021

Address of printing house: «NurNaz GRACE», 103, Ryskulov str, Almaty.

**NEWS**

OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN  
**PHYSICO-MATHEMATICAL SERIES**

ISSN 1991-346X

Volume 1, Number 335 (2021), 19 – 25

<https://doi.org/10.32014/2021.2518-1726.3>

UDC 004.89

IRSTI 28.23.37

**M. Dilmagambetova<sup>1</sup>, O. Mamyrbayev<sup>2</sup>**<sup>1</sup>Al-Farabi Kazakh National University, Almaty, Kazakhstan;<sup>2</sup>Institute of Information and Computing Technologies, Almaty, Kazakhstan.

E-mail: morkenj@mail.ru, d.m.d97@mail.ru

**DEVELOPMENT OF THE NEURAL NETWORK FOR SOLVING  
THE PROBLEM OF SPEECH RECOGNITION**

**Abstract.** The article discusses a method for solving the problem of speech recognition on the example of recognizing individual words of a limited dictionary using a forward propagation neural network trained by the error back propagation method. The goal was to create a neural network model for recognizing the solution of individual words, analyze the training characteristics and behavior of the constructed neural network. Based on the input data and output requirements, a feedback neural network selected. To train the selected neural network model, a back propagation algorithm was chosen. The developed neural network demonstrated the expected behavior associated with learning and generalization errors. It found that even if the generalization error decreases as the learning sequence increases, the errors begin to fluctuate regardless of the introduction of a dynamic learning rate. The network sufficiently trained to meet the generalization error requirements, but there is still room to improve the generalization error. Practical results of training the constructed neural network at different sizes of the training sample presented.

**Keywords:** speech recognition, neural networks, error back propagation algorithm, learning, learning rate.

**1. Introduction.**

The task of speech recognition is one of the most urgent tasks of our time. Despite the fact that now there are many ready-made speech recognition systems based on various technologies, the problem of speech recognition not completely solved, since existing systems have certain disadvantages. In particular, the dependence of the system on access to data transmission facilities and insufficient recognition accuracy.

One of the promising directions in solving speech recognition problems is the use of artificial neural networks. Neural networks are widely used in solving various classes of pattern recognition problems due to their ability to generalize.

**2. Source data of the task**

Aspects of the construction and application of neural networks for solving the problem of speech recognition on the example of the problem of recognizing numbers from 1 to 9, i.e. the words "one", "two", "three", "four", "five", "six", "seven", "eight" and "nine", respectively. Since the sounds of human speech lie in the frequency range from 100 to 4000 Hz, to solve this problem, it is enough to use a sampling frequency of 11025 Hz to digitize speech signals. Using this frequency allows you to reduce the flow of audio data, while avoiding the loss of useful components of the signal. As part of the task, audio signals are represented by sets of frames, each of which contains 512 samples.

Based on the experimental analysis of audio recordings of various pronunciation variants of the studied words, the maximum duration of the useful signal was determined (figure 1), which was 1 s. Accordingly, the minimum set of frames covering the duration of the useful signal should consist of 20 frames. Missing samples of the original signal filled with zeros.

The results of the Fourier transform performed for each analyzed frame will be used as input data for training the neural network. This approach allows you to analyze the signal both in the frequency domain (using the frame spectrum) and in the time domain - by splitting the original signal into frames. Since

significant information is contained in the real frequency spectrum, after performing the Fourier transform, the real spectrum of the signal is used, discarding the phase information (figure 2, b).

At the output of the neural network, a number is expected that is in the range from 1 to 9 and uniquely corresponds to its verbal representation submitted to the input of the neural network.

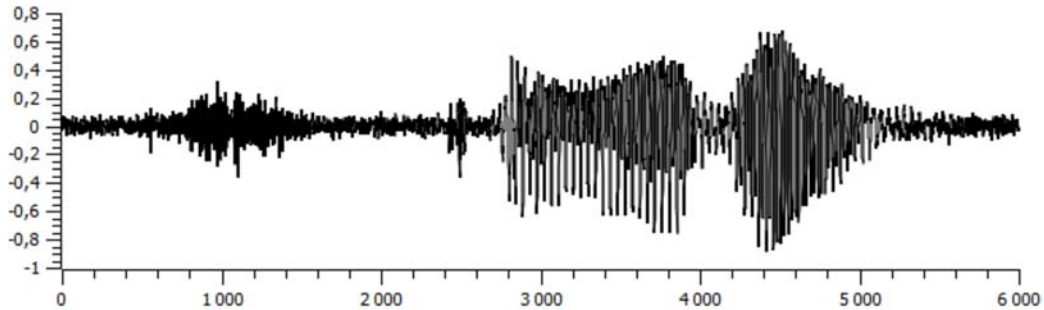


Figure 1 - Time diagram of the word «four»

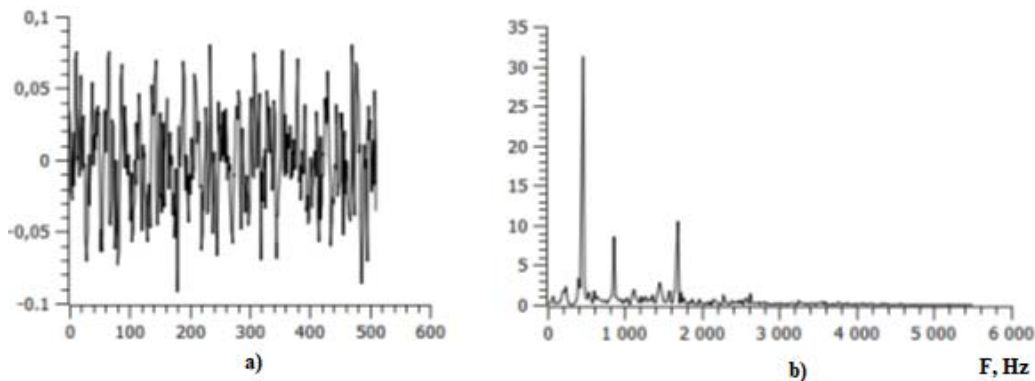


Figure 2 - The first frame of the signal "four":  
a-time diagram; b-spectrum of the selected frame

### 3. Neural network approach to solving the problem

A neural network is a collection of connected and interconnected artificial neurons that accumulate input values and generate an output signal using the activation function. The work of a single neuron can be represented by the formula

$$y_j = F(\sum w_{ij}x_i), \quad (1)$$

where  $y_j$  – the output signal of  $j$  neuron;  $w_{ij}$  – the weight of the connection between  $i$  and  $j$  neurons;  $x_i$  – the output signal of  $i$  neuron;  $F$  – the activation function of the neuron [1].

Methods of connecting neurons in artificial neural networks determine the topology of the neural network. According to the structure of interneuron connections, two types of neural networks can be distinguished: direct propagation neural networks and recurrent neural networks. In direct propagation neural networks, the communication between layers is unidirectional – each neuron connected only to the neurons of the next layer. Such networks are static due to the lack of feedbacks and dynamic elements. The output of such a network depends only on the input data. Recurrent neural networks are dynamic, due to the presence of feedbacks. The output of a recurrent neural network depends on its previous state [2].

The topology of the neural network is selected directly for the analyzed problem, taking into account the features and complexity of its solution. Optimal configurations already exist for some types of tasks. However, if the problem cannot be reduced to any of the known types, it is necessary to synthesize a new configuration of the neural network directly for the problem solved. Since there is no general method for choosing the optimal configuration of a neural network, the structure of the neural network is selected experimentally.

The most obvious structure is the network of direct signal propagation, so named because the neurons of one layer can only be connected to the neurons of nearby layers without reverse and recurrent connections [3]. Typically, such networks consist of an input layer, one or more hidden layers, and an

output layer. The simplest structure of such a network is shown in figure 3. This network has one hidden layer, an input layer consisting of  $n$  neurons and an output layer consisting of  $m$  neurons.

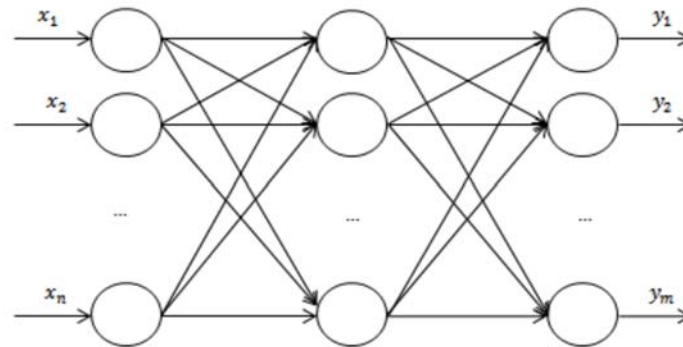


Figure 3 - Typical structure of a neural network

Using such a neural network, data are converted from an  $n$ -dimensional input space to an  $m$ -dimensional output space. The advantage of this type of neural networks is their relative simplicity and visibility, which allows you to analyze the operation of the neural network used. Based on the format of input and output data, a neural network of direct signal propagation will be used to solve the problem, the input layer of which contains such a number of neurons that corresponds to the number of analyzed features (that is, the number of frames multiplied by the number of analyzed spectral components) [4]. Recurrent neural networks cannot be used in the solution of the problem, as due to the presence of feedback output values of a recurrent neural network depends on the previous state of the network, and since spoken words within the tasks are not linked, the previous state of the network should not affect the recognition result.

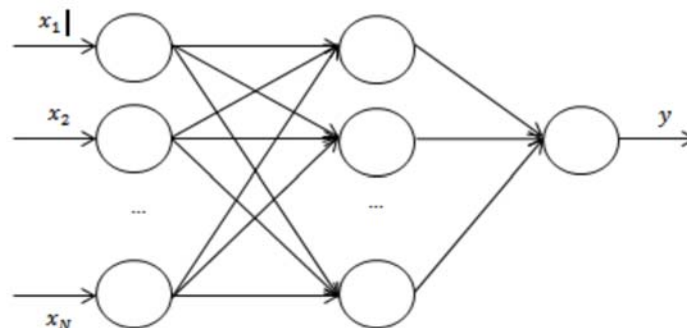


Figure 4 - The neural network structure used

To solve speech recognition problems, the most common solution is to use the number of neurons in the output layer that corresponds to the number of recognized objects. However, when solving this problem, the neural network architecture was chosen, which contains one neuron in the output layer, the output value of which is in the range from 0 to 1, which corresponds to the numbers from 1 to 9. The neural network used also has one hidden layer (figure 4).

After determining the topology of the neural network, you need to choose a learning algorithm. Neural network training can be done in two ways - with a teacher and without a teacher. When using training with a teacher, neural networks are represented by pairs of input and output data vectors, because of which the error is calculated and the weights of the neural network adjusted. The algorithm repeated until the neural network error reaches the required minimum value. In the case of unsupervised learning, the output data are not known in advance, so only the input data are used [5].

When choosing a training algorithm, it is necessary to take into account the topology of the neural network, the model of the analyzed data and the intended method of training the neural network. Since a forward propagation neural network chosen, the most well - known multi-layer perceptron learning algorithm, the error back propagation algorithm, will be used.



#### 4. Training a neural network by back propagation of an error

The error back propagation algorithm involves calculating the error of both the output layer and each neuron of the trained network, as well as correcting the weights of the neurons in accordance with their current values. In the first step of this algorithm, the weights of all interneuron connections initialized to small random values (from 0 to 1). After initializing the weights, the following steps performed in the neural network training process:

- direct signal propagation;
- calculation of the error of neurons of the last layer;
- reverse propagation of the error [6].

Direct spread signal made in layers, starting with the input layer, in this case calculates the sum of the input signals for each neuron using an activation function to generate the response of a neuron that propagates in the next layer, weighted by interneuron connection according to the formula (1). As a result of this step, a vector of output values of the neural network obtained.

The next stage of training is to calculate the neural network error as the difference between the expected and actual output values. The error is calculated for each neuron of the output layer according to the formula

$$\delta_k = (EXP_k - y_k)F'(y_k), \quad (2)$$

where  $\delta_k$  – the received error of the k neuron of the output layer;  $EXP_k$  – the expected value for k output neuron;  $y_k$  – the actual output value of k neuron;  $F'(y_k)$  – the derivative of the activation function of k neuron [7].

For subsequent layers of the neural network, the neuron error is calculated using the formula

$$\delta_k = F'(y_k) * \sum_{i=1}^M \delta_i w_{ki}, \quad (3)$$

where  $\delta_k$  – received error for the k neuron;  $\delta_i$  – error of the i neuron of the previous layer;  $w_{ki}$  – the weight connection between neuron k of the current layer and neuron i of previous layer;  $y_k$  – the actual output value of neuron k;  $F'(y_k)$  – derivative of the activation function of neuron k; M – number of neurons of the previous layer [8].

The resulting error values propagate from the last, output layer of the neural network, to the first. In this case, the values of the correction of the weights of neurons calculated depending on the current value of the link weight, the learning rate and the error made by this neuron.

After completing this step, the steps of the described algorithm are repeated until the error of the output layer reaches the required value.

When correcting the weights of interneuron connections, the concept of learning rate is used. The learning rate of a neural network is one of the most important parameters that control the learning process. This parameter determines the amount of change in the weighting coefficients of interneuron connections. For a perfect approximation to the minimum error of the neural network, the learning rate should tend to an infinitesimal value to ensure the best convergence of the learning algorithm. However, the smaller the selected value of the learning step, the longer the learning takes place online [9].

In order to overcome these problems, the so-called dynamic learning rate is used. When using this method, the learning step is not a constant value, but depends on other parameters of the learning process (time, iteration number, or neuron error in the previous step). The dynamic learning rate can be introduced for each neuron of the network individually, or for the entire network as a whole.

The functions used to calculate the learning rate must have the following properties:

- 1)  $Y(x) = 0$  for  $x = 0$ ;
- 2)  $Y(x) = MAX$  at  $x \rightarrow \pm\infty$ ;
- 3)  $Y(x) \rightarrow 0$  for  $x \rightarrow 0$ .

To work with the neural network, the following function is selected, reflecting the dependence of the learning rate of the neuron on the error value:

$$Y(x) = |MAX * (-CST * |x|)|, \quad (4)$$

where MAX – a constant that determines the maximum possible learning rate; x – the amount of error introduced by the neuron; CST – a constant that determines the degree of steepness of the resulting function. The function is represented by a graph in figure 5.

This function meets the specified requirements and provides the most optimal change in the learning rate. At the beginning of the learning process, the MAX parameter is set to the maximum value of the learning rate (in our case,  $MAX = 3$ ), because of which, for large values of the learning error, changes in the weight coefficients will be significant. As the neuron error decreases, the learning rate will decrease, and as the learning error tends to zero, the learning rate will also tend to zero [10].

Thus, when solving the problem, dynamic control of the learning rate is implemented, in which the value of the learning step is calculated for each neuron separately, depending on the error made by this neuron. The introduction of this algorithm made it possible to more accurately approach to the minimum learning error of the neural network. When comparing the learning nature of a neural network with an adaptive learning rate and a neural network with a minimum fixed learning step, the former shows a smoother tendency of the error to the minimum value without significant fluctuations.

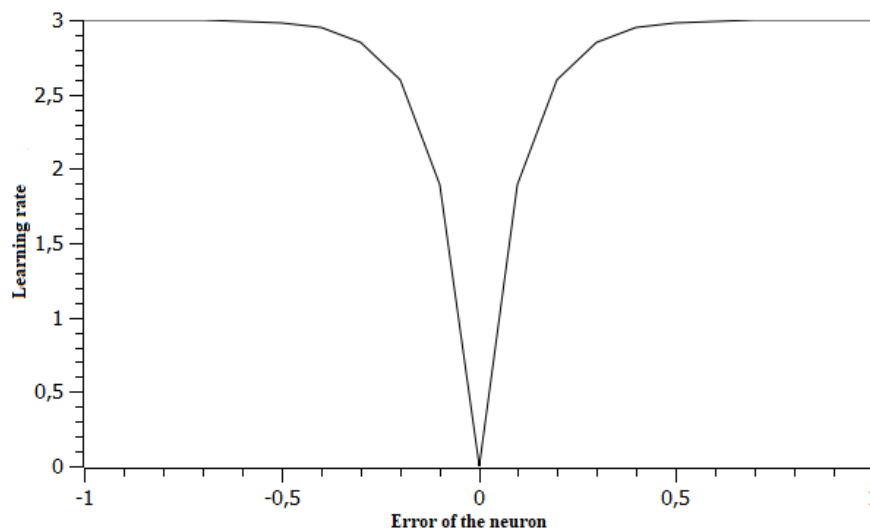


Figure 5 - Graph of the learning rate of an individual neuron

Two types of neural network errors that most fully characterize the learning process are considered. In the process of training a neural network, a training error and a generalization error are distinguished. Generalization error is an error that the neural network demonstrates in examples that were not involved in the learning process. A learning error, on the contrary, is an error that the trained neural network demonstrates on the examples of the training sample [11].

An important aspect of neural network training is the training sample. A training sample is a set of pairs of input and output data (for training with a teacher) used in training a neural network. The control sample-part of the sets that are not involved in training the neural network - used to determine the generalization error.

For correct training of a neural network, the training sample must have the representativeness property. Representativeness in this case should be understood as the presence of a sufficient number of diverse training examples that reflect the patterns that should be detected by the neural network in the learning process. The representativeness of the training sample expressed in the following aspects:

- sufficiency: the number of training examples should be sufficient for training;
- diversity: the training sample should contain a large number of different combinations of input and output data in the training examples;
- uniform representation of classes: examples of different classes should be presented in the training sample in the same proportions.

Increasing the number of examples in the training sample increases the time required for the neural network to reach the specified indicators due to a generalization error [12].

When training the constructed neural network, we obtained results confirming the theoretical dependence of the generalization error on the power of the training sample (Fig. 6, a). The dependence between the power of the training sample and the deviation of the generalization error from the steady-state value also revealed (figure 6, b).

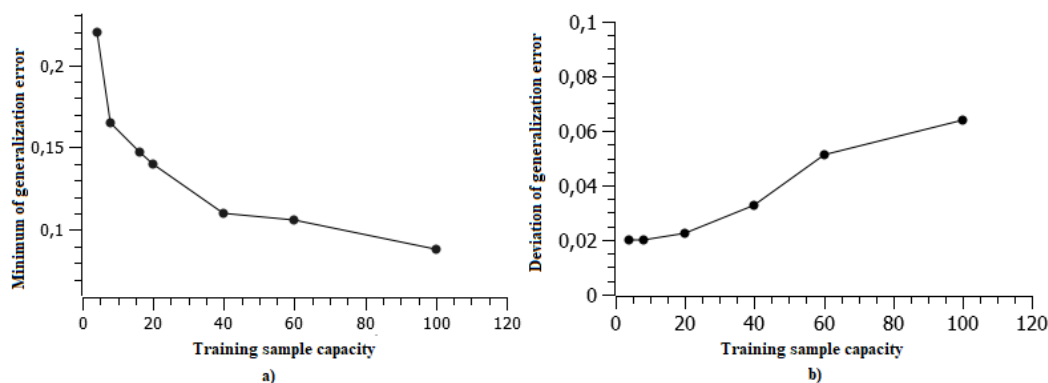


Figure 6 - Behavior of the generalization error depending on the power of the training sample: a-change in the minimum value of the generalization error; b-change in the deviation of the generalization error from the steady-state value

**5. Conclusion.** Based on the results obtained, it can be concluded that with an increase in the training sample size, the minimum possible value of the generalization error decreases, and the nature of the dependence of the generalization error on the power of the training sample coincides with the theoretical one. However, this increases the time spent on training the neural network, and increases the deviation of the generalization error from the established value.

Thus, a neural network model was implemented to solve the problem of recognizing words corresponding to the pronunciation of numbers from 1 to 9. When analyzing the behavior of the neural network, it was determined that the existing size of the training sample is not enough to achieve a zero error in the generalization of the neural network. However, the constructed network showed the ability to learn, confirmed by experimental data. When the specified generalization error is reached, the program saves the weight coefficients in the form of a header file, which makes it possible to restore the trained neural network for later use.

**М. Д. Дильмагамбетова<sup>1</sup>, О. Ж. Мамырбаев<sup>2</sup>**

<sup>1</sup>Әл-Фараби атындағы ҚазҰУ, Алматы, Қазақстан;

<sup>2</sup>ҚР БҒМ БК Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан

### СӨЙЛЕУДІ ТАҢУ МАҚСАТЫНДА НЕЙРОНДЫҚ ЖЕЛІНІ ҚҰРУ

**Аннотация.** Мақалада қатені кері тарату әдісімен оқытылған тікелей таратудың нейрондық желісін қолдана отырып, шектеулі сөздіктің жеке сөзін тану арқылы сөйлеуді тану мәселесін шешу әдісі қарастырылады. Мақсатымыз – жеке сөзді тану үшін нейрондық желі моделін құру, құрылған нейрондық желінің оқыту сипаттамалары мен әрекетін талдау.

Фурье түрлендіру нәтижелері нейрондық желіні оқыту үшін кіріс ретінде пайдаланылады және кіріс деректерін таңдауға негізделген. Мәселені шешу үшін қатенің динамикалық оқу жылдамдығымен кері таралу алгоритмі таңдалды, өйткені бұл алгоритмді енгізу нейрондық желіні оқытудың минималды қатесіне дәл жақындауға мүмкіндік берді.

Желі жалпылау қатесінің талаптарын қанағаттандыруға жеткілікті дайындалған, бірақ жалпылау қатесін жақсартуға болады. Құрылған нейрондық желіні оқытудың практикалық нәтижелері түрлі оқу үлгісінде ұсынылған.

**Түйін сөздер:** сөйлеуді тану, нейрондық желілер, қателіктерді кері тарату алгоритмі, оқыту, оқу жылдамдығы.

**М. Д. Дильмагамбетова<sup>1</sup>, О. Ж. Мамырбаев<sup>2</sup>**

<sup>1</sup>Казахский национальный университет им. аль-Фараби, Алматы;

<sup>2</sup> Институт информационных и вычислительных технологий КН МОН РК, Алматы, Казахстан

### РАЗРАБОТКА НЕЙРОННОЙ СЕТИ ДЛЯ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ РЕЧИ

**Аннотация.** В статье рассматривается метод решения задачи распознавания речи на примере распознавания отдельных слов ограниченного словаря с использованием нейронной сети прямого распространения, обученной методом обратного распространения ошибок. Цель состояла в том, чтобы создать нейросетевую модель для

распознавания отдельных слов, проанализировать обучающие характеристики и поведение построенной нейронной сети.

Результаты преобразования Фурье использованы в качестве входных данных для обучения нейронной сети и обоснован выбор входных данных. Для решения поставленной задачи выбран алгоритм обратного распространения ошибки с динамической скоростью обучения, так как введение этого алгоритма позволило более точно приблизиться к минимальной ошибке обучения нейронной сети.

Сеть достаточно обучена, чтобы соответствовать требованиям ошибки обобщения, но есть еще место для улучшения ошибки обобщения. Представлены практические результаты обучения построенной нейронной сети при различных размерах обучающей выборки.

**Ключевые слова:** распознавание речи, нейронные сети, алгоритм обратного распространения ошибок, обучение, скорость обучения.

#### Information about authors:

Mamyrbayev O. Zh., PhD, Associate Professor, head of the Laboratory of computer engineering of intelligent systems at the Institute of Information and Computational Technologies, Almaty, Kazakhstan; morkenj@mail.ru; <https://orcid.org/0000-0001-8318-3794>;

Dilmagambetova M. D., Master's degree student, Al-Farabi Kazakh National University, Almaty, Kazakhstan; d.m.d97@mail.ru, <https://orcid.org/0000-0002-8456-5417>

#### REFERENCES

- [1] Rabiner L., Juang B. Speech Recognition. Chapter in Springer Handbook of Speech Processing. NY: Springer. 2008. (In Eng.).
- [2] Hinton G. et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine. 2012. vol. 29. no. 6. pp. 82–97.
- [3] Makovkin K.A. [Hybrid models – Hidden Markov Models/Multilayer perceptron and their application in speech recognition systems. Servey]. Rechevye tehnologii – Speech Technology. 2012. vol. 3. pp. 58–83. (in Russ.).
- [4] Yu D., Deng L. Automatic Speech Recognition – A Deep Learning Approach. Springer. 2015. 322 p.
- [5] Deng L. Deep learning: from speech recognition to language and multimodal processing. APSIPA Transactions on Signal and Information Processing. 2016. vol 5. pp. 1–15.
- [6] Seide F., Li G., Yu D. Conversational speech transcription using context-dependent deep neural networks. Proceedings of Interspeech. 2011. pp. 437–440.
- [7] Dahl G., Yu D., Deng L., Acero A. Context-dependent pre-trained deep neural networks for large vocabulary speech recognition. IEEE Transactions on Audio, Speech and Language Processing. 2012. vol. 20. no. 1. pp. 30–42.
- [8] Maas A.L. et al. Building DNN Acoustic Models for Large Vocabulary Speech Recognition. Preprint arXiv:1406.7806. 2015. Available at: <http://arxiv.org/pdf/1406.7806.pdf> (accessed: 14.09.2016).
- [9] Mamyrbayev O., Toleu A., Tolegen G., Mekebayev N. Neural architectures for gender detection and speaker identification, Cogent Engineering 7 (1), 1727168, 2020. (In Eng.)
- [10] Cosi P. A KALDI-DNN-based ASR system for Italian. Proceedings of IEEE International Joint Conference on Neural Networks IJCNN'2015. 2015. pp. 1–5.
- [11] Mamyrbayev O.Zh., Turdalyuly M., Mekebaev N.O., Kydyrbekova A.S. Automatic Recognition of the Speech Using Digital Neural Networks // ACIIDS, Indonesia, Proceedings. Part II, 2019 (In Eng.)
- [12] Veselý K. et al. Sequence-discriminative training of deep neural networks. Proceedings of INTERSPEECH'2013. 2013. pp. 2345–2349.
- [13] Povey D., Zhang X., Khudanpur S. Parallel training of DNNs with natural gradient and parameter averaging. Preprint arXiv:1410.7455. 2014. Available at: <http://arxiv.org/pdf/1410.7455v8.pdf> (accessed: 14.09.2016).
- [14] Popović B. et al. Deep Neural Network Based Continuous Speech Recognition for Serbian Using the Kaldi Toolkit. Proceedings of the 17th International Conference on Speech and Computer (SPECOM-2015). Springer. 2015. LNAI 9319. pp. 186–192.
- [15] Miao Y. Kaldi+ PDNN: building DNN-based ASR systems with Kaldi and PDNN. arXiv preprint arXiv:1401.6984. 2014. Available at: <https://arxiv.org/ftp/arxiv/papers/1401/1401.6984.pdf> (accessed: 14.09.2016).
- [16] Sainath T.N., Mohamed A.R., Kingsbury B., Ramabhadran B. Deep convolutional neural networks for LVCSR. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2013. pp. 8614–8618.
- [17] Delcroix M. et al. Context adaptive neural network for rapid adaptation of deep CNN based acoustic models. Proceedings of INTERSPEECH-2016. 2016. pp. 1573–1577.
- [18] Gapochkin A. V. [Нейронные сети в системах распознавания речи]. Science Time. 2014. vol. 1(1). pp. 29–36. (In Russ.).
- [19] Peddinti V., Povey D., Khudanpur S. A time delay neural network architecture for efficient modeling of long temporal contexts. Proceedings of INTERSPEECH-2015. 2015. pp. 2440–2444.
- [20] Tampel I.B. [Automatic speech recognition – the main stages over last 50 years]. Nauchno-tehnicheskij vestnik informacionnyh tehnologij, mehaniki i optiki – Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2015. vol. 15. no. 6. pp. 957–968 (In Russ.).

**Publication Ethics and Publication Malpractice**  
**in the journals of the National Academy of Sciences of the Republic of Kazakhstan**

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct ([http://publicationethics.org/files/u2/New\\_Code.pdf](http://publicationethics.org/files/u2/New_Code.pdf)). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

(Правила оформления статьи для публикации в журнале смотреть на сайтах:

[www.nauka-nanrk.kz](http://www.nauka-nanrk.kz)

<http://physics-mathematics.kz/index.php/en/archive>

**ISSN 2518-1726 (Online), ISSN 1991-346X (Print)**

Редакторы: *М. С. Ахметова, Д. С. Аленов, А. Ахметова*  
Верстка на компьютере *А.М. Кульгинбаевой*

Подписано в печать 08.02.2021.  
Формат 60x881/8. Бумага офсетная. Печать – ризограф.  
6,75 п.л. Тираж 300. Заказ 1.