International Journal of Civil Engineering and Technology (IJCIET)
Volume 8, Issue 10, October 2017, pp. 523–532, Article ID: IJCIET_08_10_054
Available online at http://http://www.iaeme.com/ijciet/issues.asp?JType=IJCIET&VType=8&IType=10
ISSN Print: 0976-6308 and ISSN Online: 0976-6316

BUILDING ROBOT VOICE CONTROL TRAINING METHODOLOGY USING ARTIFICIAL NEURAL NET

Rami Matarneh
Department of Computer Science,
Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

Svitlana Maksymova
Department of Computer-Integrated Technologies, Automation and Mechatronics,
Kharkiv National University of RadioElectronics, Kharkiv, Ukraine

Zhanna Deineko
Department of Media Systems and Technology,
Kharkiv National University of RadioElectronics, Kharkiv, Ukraine

Vyacheslav Lyashenko
Department of Informatics,
Kharkiv National University of RadioElectronics, Kharkiv, Ukraine

ABSTRACT

The article describes using artificial neural net for speech recognition. This is necessary for the automation of construction work. We selected multi-layer perceptron for separate words recognition tasks. We researched different forms of hidden layer. And we made a conclusion that for voice commands analysis tasks solution it is expedient to use multi-layer perceptron with linearized functions (the best result was achieved using model with hidden layer with linearized hyperbolic tangent function). We use this experience to train robots in civil engineering.

Key words: Building Robot, Voice Control, Motion Commands, Artificial Neural Net

Cite this Article: Rami Matarneh, Svitlana Maksymova, Zhanna Deineko and Vyacheslav Lyashenko, Building Robot Voice Control Training Methodology Using Artificial Neural Net, International Journal of Civil Engineering and Technology, 8(10), 2017, pp. 523–532
http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=10

1. INTRODUCTION

Now modern building technologies actively use different robotic complexes. In [1] robots for welding are considered. Their using allows both to speed up different constructions welding and to protect humans against welding harmful effects. Robots using in building for safety provision, efficiency and quality improving is also considered in article [2]. Generally, robots using in building direction is connected with general concept of production and technology risks minimization [3].

At the same time in [4] the approach is represented that allows robotizing masonry processes using robots system connected between themselves by parallel connections.

In [5] building works automation using autonomous mobile robots general concept is considered. Wherein such research authors notice it is necessary to use natural-language commands for robots control.

Researchers [6] express similar ideas. They also focus their attention on natural-language command using for building robot control.

Thus robot voice control for building tasks is a perspective field both for research and for practical implementation in building practice. Nevertheless, specified task solution involves robot voice control training stage. In its turn it involves a number specify tasks solution [7, 8].

2. MATERIALS AND METHODS

2.1 Voice Information Analysis Base for Robotic Systems

During natural language processing the procedure for splitting a phrase into recognizable keywords is used by grammatical analysis. Then obtained command structure is analyzed by syntax (grammar, time), vocabulary and context. Then [7, 9, 10]:

if final result is incompatible with known for machine «rules» we can reduce model dimensions and either repeat all the process or perform command grammar analysis again;

if the command is input correctly, i.e. it is recognized and it can be performed movement command sequence is generated, it is an input to robot controller;

if commands are incompatible with current robot position (e.g. robot is in position unsuitable for necessary object taking), computer generates a feedback sound to the person (synthesizes speech) suggesting to operator to perform correcting actions.

Difficulties during speech recognition are explained by voice changings inherent in different people or in one-person speech. Thus, sentences structure, phrases meanings and speech morphology must be programmed in particular in «rules» form in order to allow robot to self-learn during conversation.

In modern robotic systems more attention is paid to interface with natural information input-output development (handwriting recognition, speech dialogue) [8, 11].

Now speech input systems are the most perspective. Most of them are based on CMU Sphinx, Google Speech API, Microsoft SpeechAPI, Siri and so on [12-15]. Speech information recognition task can be divided into two large subtasks:

- Separate words direct recognition;
- Commands meaning recognition.

Separate words direct recognition is complicated by a number of factors: the difference in languages, the specifics of pronunciation, noises, accents, accentuations [7, 16].

Now we can distinguish two main directions during speech recognition systems development [7, 17, 18]:

Standard method is based on comparison some speech characteristics (energy, spectral). As the standards in most cases whole words are used. This method is convenient for using in systems with limited vocabulary (e.g. for small command set). Standards are formed by large number of templates statistic processing. Input signal comparison with the standard is possible by fuzzy pattern matching;

Phonemes-oriented method is based on phonemes distinguishing from speech stream. Comparing speech stream recognition by whole words recognition method and phonemes recognition we can make a conclusion that with a small number of words used by the operator the higher reliability and speed can be expected from the whole words recognition method but when the vocabulary increases, the speed drops sharply.

At the same time voice information analysis may be provided by different methods. Classical method is voice information processing using discrete Fourier transform [19, 20].

Nevertheless, artificial neural nets using is a perspective method [21, 22]. Artificial neural nets by their structures are sets of interconnected adaptive nonlinear processing elements. So practically, any processing elements can be implemented by a simple sum of input signals, which is modulated by a nonlinear function.

2.2 Trained Artificial Neural Networks in Speech Recognition Systems

Neural nets using, as one of the means for intelligent data analysis implementation, allows to solve following tasks [21, 22]:

- To simulate complex nonlinear dependencies between data and target indicators;
- To identify trends in data (in the presence of time series) for forecasting;
- To work with noisy and incomplete data;
- To obtain meaningful results with a relatively small amount of the initial information with the ability to improve the model as new data become;
- To identify abnormal data that deviates significantly from «open» stable laws.

One of the neural nets types are trained nets. This nets type is used for non-formalizable tasks to the category of which speech recognition belongs. During net training such parameters as synaptic coefficients are changed automatically. In some case topology also can be changed automatically [23].

Neural nets also have the property of classifying objects by their numerical parameters. When teaching a net with a teacher, net can be taught to recognize objects belonging to a predetermined classes set. If net is taught without teacher it can group objects by classes according to their digital parameters. Thus, on the basis of neural nets it is possible to create learning and self-learning systems.

In particular, for voice control system realization multi-layer perceptron can be used. It is one of the simplest artificial neural nets models. It is realized as a net containing one input, one or more hidden and output layers. Perceptrons ability to solve recognition problems in combination with the implementation simplicity allows them to be used in many industries [24, 25].

Multi-layer perceptron has several processing elements arranged in layers. Layers that don't have direct access to the outside world, for example, connected to the input or output layer, are called hidden. Layers that receive input signal from the outside are called input layers. Layers that are in contact with outside world are called output layers. Layers are connected between themselves by synapses. Then for multi-layer perceptron implementation in robot voice control system it is expedient to use perceptron with hidden layer [26, 27].

At the same time for such realization usually describing functions of hyperbolic tangent or sigmoid function are used.

Sigmoid function has next form [28]:

$$y = 1/(1+\exp(-(\sum_i W_i x_i - \Theta))), \tag{1}$$

Where, w_i – weights (synapse coefficients), to which the input values are multiplied,

x_i – Input values to i-th layer,

Θ – Some input threshold.

Hyperbolic tangent function has next form [28, 29]:

$$y = \tanh(x), \tag{2}$$

Modifying net parameters process, aimed at improving output signals, is called learning. In artificial neural net learning process net output signal error relatively to the desired signal result is calculated. The error signal is sent to the back propagation net and provides modification of synaptic (weight) coefficients matrix connecting net layers.

Generally, for signs static recognition a multilayered perceptron with two hidden layers is considered a universal means [26].

For learning errors correction error back propagation algorithm is used. According to it:

1. Start weights values of all layers' neurons V(t=0) and W(t=0) are put random values.
2. Input pattern X^α is put to net, as a result, an output pattern is generated $y \neq Y^\alpha$. In this case, neurons consistently from layer to layer function according to the following formulas [27]:

Hidden layer:

$$x_i = \sum_i W_{ij} X_i^\alpha, \tag{3}$$

$$y_i = f(x_i), \tag{4}$$

Output layer:

$$x_k = \sum_j V_{jk} y_j, \tag{5}$$

$$y_k = f(x_k), \tag{6}$$

Where, f(x) – layer transition function.

3. Net quadratic error functional for a given input pattern has next form [27-29]:

$$E = 1/2(\sum_k (y_k - Y_k^\alpha)^2), \tag{7}$$

Where, y_k - real net output,

Y_k^α - desired value.

This functional must be minimized. The classical gradient optimization method consists in iterative refinement of the argument according to formula [28]:

$$V_{jk}(t+1) = V_{jk}(t) - h * \frac{\partial E}{\partial V_{jk}}, \tag{8}$$

Where, h – learning speed.

4. In this step, the weights of the hidden layer are adjusted (W) [27-29]:

$$W_{ij}(t+1) = W_{ij}(t) - h * \frac{\partial E}{\partial W_{ij}}, \tag{9}$$

If there are several hidden layers in neural net, back propagation procedure is applied sequentially for each of them beginning with the layer preceding the output, and then the previous one and so on up to the layer following the input.

5. Steps 2-4 are repeated for all training vectors. Training ends when a small total error or the maximum number of iterations is reached.

3. RESULTS AND DISCUSSION

In the case of voice information recognition voice commands samples in the form of discrete signal values sequences (2304) are put to perceptron input. Such signals examples are represented on Figure 1.

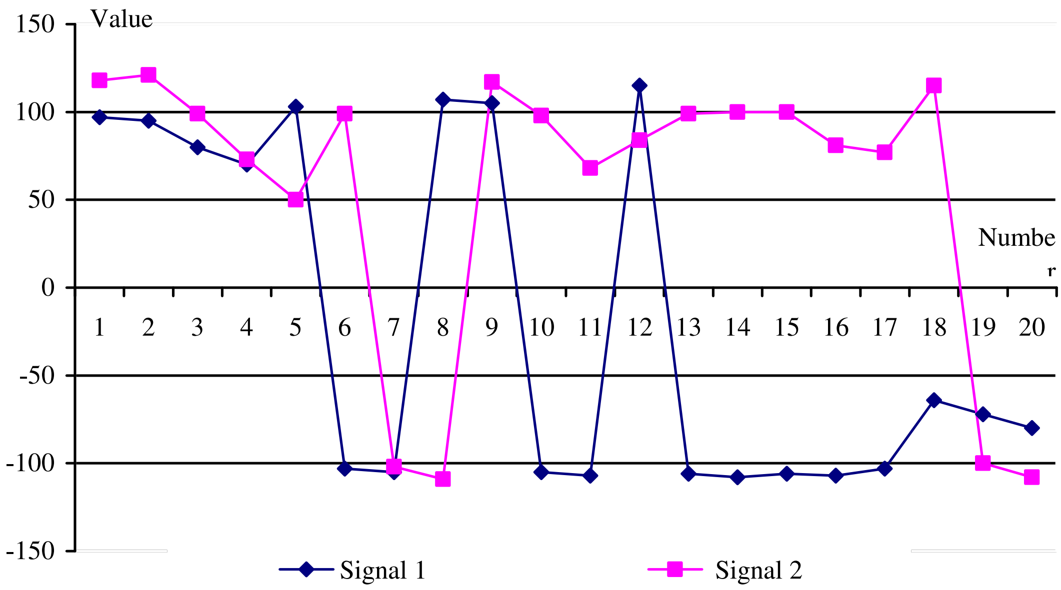

Figure 1 Input signals fragments graphics

Counters number determines net input layer dimension. Proportional to the signal length, perceptron hidden layer dimension is set (24, 48, 256). Output layer dimension is determined by desired result. If we use net transmission function layers in the form of a hyperbolic tangent at the net output we can use values range [-1, 1]. If as a hidden layer sigmoid function values range [0, 1] is set. Training is considered to be completed when error is low than 0.05.

During experiments net was trained for separate phonemes, words and syllables that form voice commands. In particular, we researched training for next words «Forward» (Signal 1), «Left» (Signal 2) and «Stop» (Signal 3). Training steps number is 30. Using hidden layer with hyperbolic tangent function training curve for Signal 1 should converge to -1, for Signal 2 – to 0, Signal 3 – to 1.

Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, Figure 7 and Figure 8 represent training results.

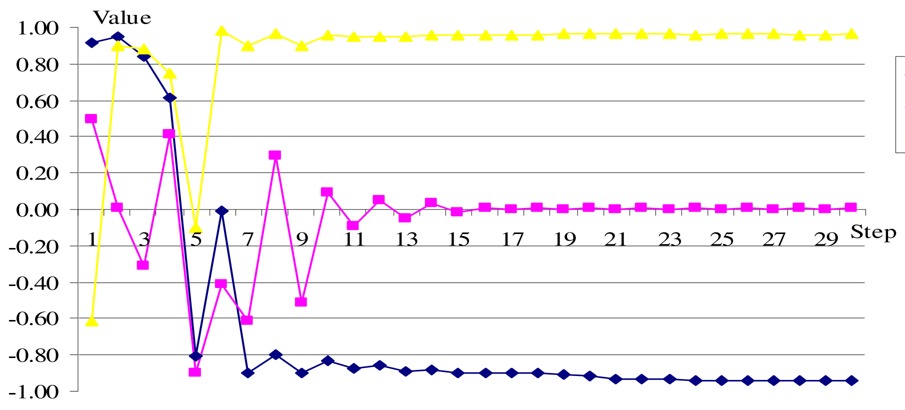

Figure 2 Perceptron training with hidden layer with hyperbolic tangent function (hidden layer dimension – 48x1)

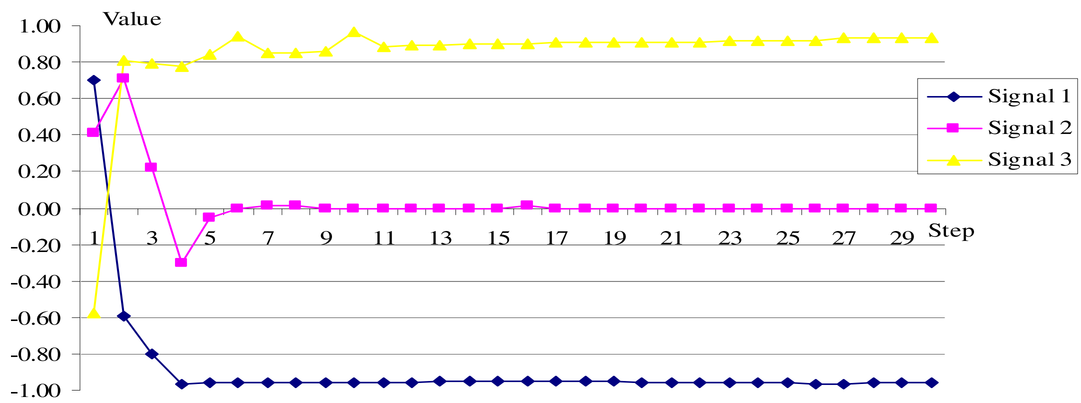

Figure 3 Perceptron training with hidden layer with hyperbolic tangent function (hidden layer dimension – 24x1)

According to net properties research we have to note using hidden layer with hyperbolic tangent function with the capacity 256x1, training is not performed. When reducing hidden layer dimension training becomes faster and better. The best results are achieved with its dimensions 24x1.

Using hidden layer with sigmoid function and linearized sigmoid function training curve for Signal 1 should converge to 0, for Signal 2 – to 0.5, Signal 3 – to 1.

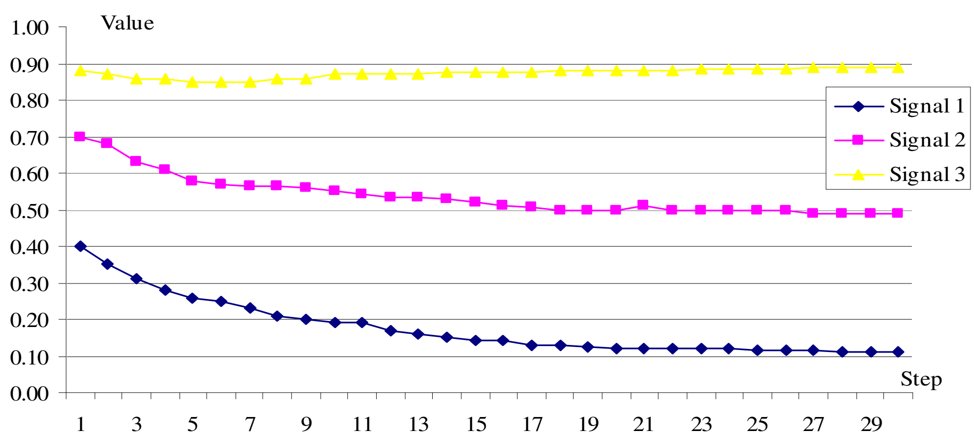

Figure 4 Perceptron training with hidden layer with sigmoid function (hidden layer dimension – 48x1)

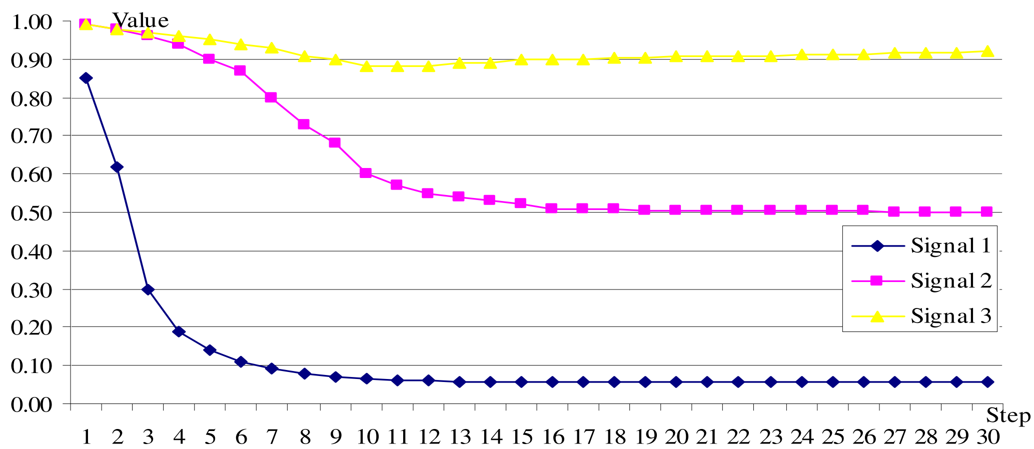

Figure 5 Perceptron training with hidden layer with sigmoid function (hidden layer dimension – 256x1)

Using hidden layer with sigmoid function with capacity 24x1 training process runs very slowly, increasing hidden layer dimension increases training speed and quality. The best results are achieved with dimensions 256x1.

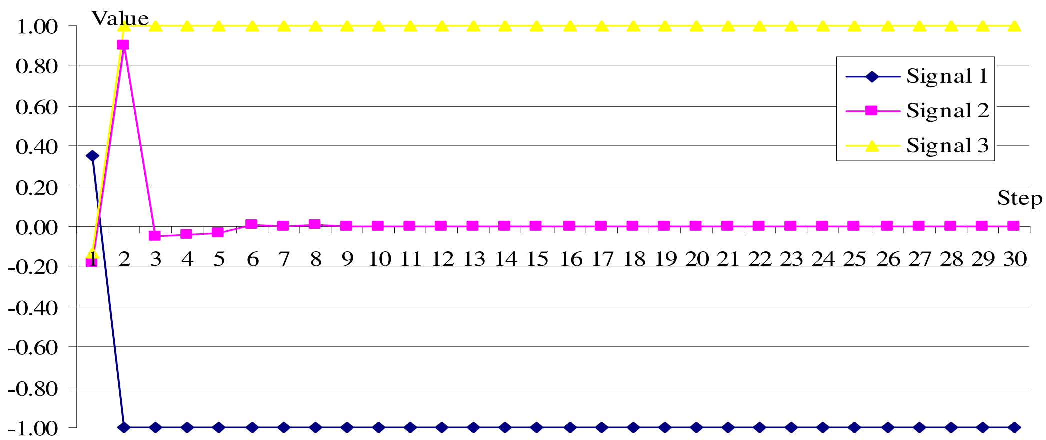

Figure 6 Perceptron training with hidden layer with linearized hyperbolic tangent function (hidden layer dimension – 48x1)

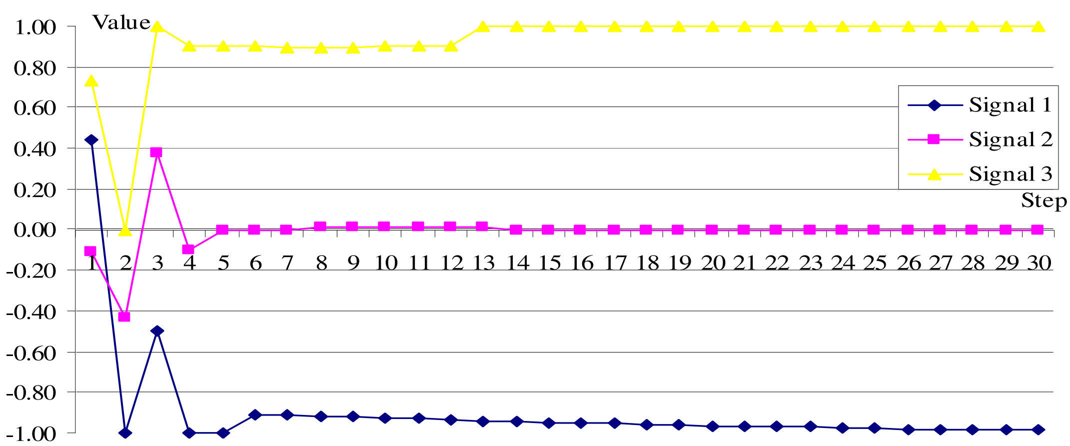

Figure 7 Perceptron Training with Hidden Layer with Linearized Hyperbolic Tangent Function (hidden layer dimension – 24x1)

Using hidden layer with hyperbolic linearized tangent function with the capacity 256x1, training is not performed. The best results are achieved with its dimensions 48x1. When the distance from this value increases training quality decreases. And with an increase of up to 256x1 the results get worse.

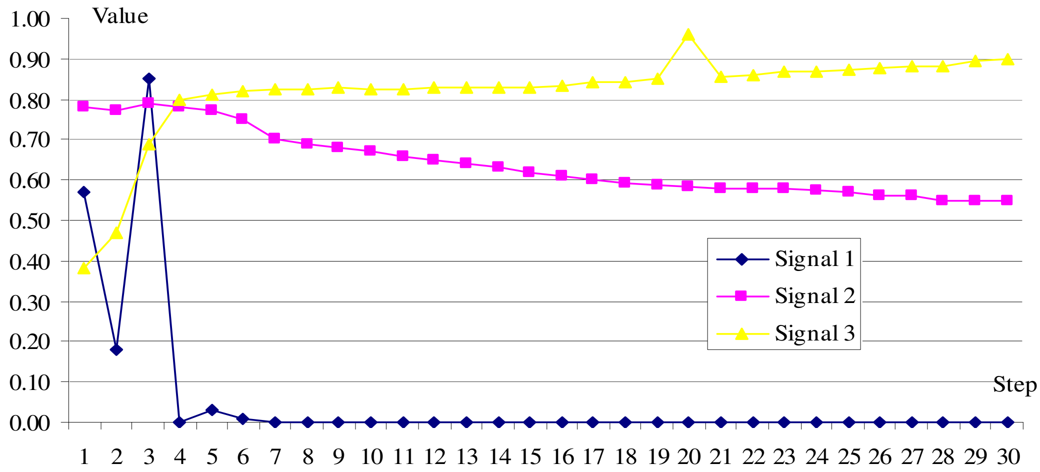

Figure 8 Perceptron Training with Hidden Layer with Linearized Sigmoid Function (hidden layer dimension – 48x1)

Using hidden layer with linearized sigmoid function the best results are achieved with its dimensions 48x1, which is equal to the square root of input layer dimension. When changing the size there is next tendency the greater the deviation from the value of 48x1, the worse is the learning process.

4. CONCLUSION

There are three robot programming methods: programming in training mode, programming in a robot programming language and analytical programming. The use of a voice command for robot control commands will greatly simplify the control process and simplify its programming.

Robot voice control using allows to achieve next advantages:

- worker fatigue decreases,
- commands input speed and flexibility are increased,
- hands are free to perform other functions,
- more saturated, rich content in response to the emerging situation is transmitted,
- the monotony of work decreases, since the operator can use his own auditory organ to check the correctness of the given teams, thereby more actively engaging in the work process [1],
- it becomes possible to implement non-contact management of various systems,
- it is possible to control complex complexes in hazardous conditions for man [2].

In the article software for voice information analyzing using an artificial neural net development is considered. As a base neural net multi-layer perceptron was selected. It was realized by C++ language tools in order to integrate it to robot control system.

Testing results showed the basic possibility of training the neural net for separate words-commands and their recognition.

The scientific value of the paper is to study the use of artificial neural nets to solve the problems of robot's voice control. The practical value of the work is to apply its results when developing an intellectual robot control system.

According to research results we can make next conclusion: for voice commands analysis tasks solution it is expedient to use multi-layer perceptron with linearized functions (the best result was achieved using model with hidden layer with linearized hyperbolic tangent function).

REFERENCES

[1] Andersen, R. S., Bogh, S., Moeslund, T. B. and Madsen, O. Intuitive task programming of stud welding robots for ship construction. In Industrial Technology (ICIT), 2015 IEEE International Conference on (pp. 3302-3307).

[2] Mattern H, Bruckmann T, Spengler A and Konig M, Simulation of automated construction using wire robots. In Winter Simulation Conference (WSC), 2016, pp. 3302-3313

[3] Kuzomin, O., Ahmad, M. A., Kots, H., Lyashenko V. and Tkachenko, M. Preventing of technogenic risks in the functioning of an industrial enterprise. International Journal of Civil Engineering and Technology, 7(3), 2016, pp. 262–270.

[4] Bruckmann, T., Mattern, H., Spengler, A., Reichert, C., Malkwitz, A. and König, M. Automated Construction of Masonry Buildings using Cable-Driven Parallel Robots. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, 33, 2016, pp. 1-10.

[5] Ardiny, H., Witwicki, S. and Mondada, F. Construction automation with autonomous mobile robots: A review. In Robotics and Mechatronics (ICROM), 2015 3rd RSI International Conference on (pp. 418-424).

[6] Saidi, K. S., Bock, T. and Georgoulas, C. Robotics in construction. In Springer handbook of robotics, 2016, pp. 1493-1520.

[7] Maksymova, S., Matarneh, R. and Lyashenko, V. V. Software for Voice Control Robot: Example of Implementation. Open Access Library Journal, 4(8), 2017, pp. 1–12.

[8] Maksymova, S., Matarneh, R., Lyashenko, V. V. and Belova, N. V. Voice Control for an Industrial Robot as a Combination of Various Robotic Assembly Process Models. Journal of Computer and Communications, 5(11), 2017, pp. 1–15.

[9] Abdullahi, Z. H., Muhammad, N. A., Kazaure, J. S. and FA, A. Mobile Robot Voice Recognition in Control Movements. International Journal of Computer Science and Electronics Engineering (IJCSEE), 3(1), 2015, pp. 11-16.

[10] Poncela, A. and Gallardo-Estrella, L. Command-based voice teleoperation of a mobile robot via a human-robot interface. Robotica, 33(1), 2015, pp. 1-18.

[11] Mavridis, N. A review of verbal and non-verbal human–robot interactive communication. Robotics and Autonomous Systems, 63, 2015, pp. 22-35.

[12] Maurtua, I., Fernandez, I., Kildal, J., Susperregi, L., Tellaeche, A. and Ibarguren, A. Enhancing safe human-robot collaboration through natural multimodal communication. In Emerging Technologies and Factory Automation (ETFA), 2016 IEEE 21st International Conference on (pp. 1-8).

[13] Manish, N. and Reddy, B. J. Review of Voice Control Robot Applications. Int. J. Adv. Eng, 1(9), 2015, pp. 671-674.

[14] Annapurna, G. S. and Mamatha, B. Transmission by an Embedded System with Enhancements in Voice Processing Technologies. International Journal of Engineering Research and Applications, 4(3), 2014, pp. 381-388.

[15] Phan, T. The Materiality of the Digital and the Gendered Voice of Siri. Transformations, 29, 2017, pp. 23-33.

[16] Kawai, S., Uehara, M., Okawa, S. and Fukushima, M. Autonomous mobile robot to improve sound environment for speech conversation. The Journal of the Acoustical Society of America, 140(4), 2016, pp. 3178-3178.

[17] Sawada, H. A Talking Robot and the Expressive Speech Communication with Human. International Journal of Affective Engineering, 14(2), pp. 2015, 95-102.

[18] Meszaros, E. L., Chandarana, M., Trujillo, A. and Allen, B. D. Compensating for Limitations in Speech-Based Natural Language Processing with Multimodal Interfaces in UAV Operation. In International Conference on Applied Human Factors and Ergonomics, 2017, (pp. 183-194).

[19] Wang, L., Yoshida, Y., Kawakami, Y. and Nakagawa, S. Relative phase information for detecting human speech and spoofed speech. In Sixteenth Annual Conference of the International Speech Communication Association. 2015, pp. 2092-2096.

[20] Patil, S. P. and Gowdy, J. N. Use of baseband phase structure to improve the performance of current speech enhancement algorithms. Speech communication, 67, 2015, pp. 78-91.

[21] Siniscalchi, S. M., Svendsen, T. and Lee, C. H. An artificial neural network approach to automatic speech processing. Neurocomputing, 140, 2014, pp. 326-338.

[22] Shahamiri, S. R. and Salim, S. S. B. Real-time frequency-based noise-robust Automatic Speech Recognition using Multi-Nets Artificial Neural Networks: A multi-views multi-learners approach. Neurocomputing, 129, 2014, pp. 199-207.

[23] Kozik, V. I., Nezhevenko, E. S. and Feoktistov, A. S. Studying the method of adaptive prediction of forest fire evolution on the basis of recurrent neural networks. Optoelectronics, Instrumentation and Data Processing, 50(4), 2014, 395-401.

[24] Chaphalkar, N. B., Iyer, K. C. and Patil, S. K. Prediction of outcome of construction dispute claims using multilayer perceptron neural network model. International Journal of Project Management, 33(8), 2015, 1827-1835.

[25] Blanco, A., Pino-Mejías, R., Lara, J. and Rayo, S. Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. Expert Systems with Applications, 40(1), 2013, 356-364.

[26] Tang, J., Deng, C. and Huang, G. B. Extreme learning machine for multilayer perceptron. IEEE transactions on neural networks and learning systems, 27(4), 2016, 809-821.

[27] Baldi, P. and Sadowski, P. A theory of local learning, the learning channel, and the optimality of backpropagation. Neural Networks, 83, 2016, 51-74.

[28] Mirjalili, S., Mirjalili, S. M. and Lewis, A. Let a biogeography-based optimizer train your multi-layer perceptron. Information Sciences, 269, 2014, 188-209.

[29] Zamanlooy, B. and Mirhassani, M. Efficient VLSI implementation of neural networks with hyperbolic tangent activation function. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 22(1), 2014, 39-48.

[30] J. Krishnaraj, K. Sangeetha, M.V. Babu Tanneru, VVS Harnadh Prasad and M. Vishnu Vardhan. A Mecanum Wheel Based Robot Platform for Warehouse Automation. International Journal of Mechanical Engineering and Technology, 8(7), 2017, pp. 181–189.

[31] K. S. Suresh, K. S. Ravichandran, S. Ananthakrishnan, and S. Venugopal, Adaptive strategy to improve the efficiency of robot path planning using population based Algorithm, International Journal of Mechanical Engineering and Technology 8(8), 2017, pp. 1441–1448.