Automated Support System Designing for People With Limited Communication

[1]A. Govintharaj, [2]R. Bhuvaneshwaran, [3]M. Gowtham, [4]Manikandan, [5]Dr .S. Ezhilarasan
1,2,3,4 B.Tech Students, Department of EEE, Prist University, Vallam, Thanjavur, Tamilnadu

Abstract : The article describes designing, modeling, and testing of the automated gesture recognition system, with access and control from wearable mobile devices. It will help people with disabilities to communicate with people who do not understand gesture language. For example, during a painful attack on Board the aircraft to ask doctor help. The developed system consists of of subsystems that perform tasks of authorization, recording, storage, and conversion of video, sending data to the server and recognition. The neural network with error back propagation is used for recognition. The topology of the used neural network is perception. This is a prototype system that works under light conditions, and in the future will be refined.

IndexTerms - cloud computing, neural network, portable devices, gesture recognition, mobile automated system

I. INTRODUCTION

Nowadays, technologies related to signal generation and processing systems in the field of communication are widespread. They are used for different purposes, and now for translation of gesture language. People with limited communication activity are helped by cellular portable devices in such trivial tasks as going to the clinic, shop or contact with people on Board the plane or on the train. For this purpose, Intel Edison[1] is developed, which translates from gesture language using the method of reference vectors. But to use it you need a special glove, which greatly limits the use of this system. There are also many projects such as Zardoz[2], Team[3], ViSiCast[4]. But all of them can only translate the text into gesture language, and we are faced with the opposite task.

On this subject written a large number of articles, for example, Ultrasonic hand gesture recognition for mobile devices(describes a novel system for ultrasonic gesture recognition targeted at handheld devices, such as smartphones)[5], Gesture recognition based on BP neural network improved by chaotic genetic algorithm (About new method that combines the chaos algorithm with the genetic algorithm)[7], Vision-based Hand Recognition Based on ToF Depth Camera (To provide a low-cost gesture recognition method for wearable devices, this thesis used merely the Time-of-Flight depth camera to achieve a lightweight gesture recognition method)[8], Hand gesture recognition using deep convolutional neural networks (Hand gesture recognition is the process of recognizing meaningful expressions of form and motion by a human involving only the hands)[9], Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition (In this work, authors address human activity and hand gesture recognition problems using 3D data sequences obtained from full-body and hand skeletons, respectively)[10] Life is very unpredictable and deaf and dumb people can get into an emergency in transport, on the street away from home. How to explain the situation to a stranger who does not know the gesture language? He does not know the gesture language, and a person with disabilities used to talk "hands" and can briefly, naturally for him to explain the situation with the help of gesture language. Therefore, it needs to have a program already trained on the transmitter device that will instantly decode what the user "says".

Scientific research at the department ISUA MTUCI is conducted in the areas related to high-performance computing for image recognition and intelligent systems based on machine learning [11-15]. One of the directions is the development of an application for a mobile device that provides a translation of gestures.

The goal is to develop a mobile automated system that allows people with limited communication in an emergency, in case of a threat to health or life to immediately inform emergency response services about what is happening or to explain to others with the help of their mobile device. The article describes the design, modeling, and testing of an automated system to support people with limited communication skills. The system integrates a set of subsystems, which are responsible for recognition of gesture language, the recording, and processing of video output of translation in the Russian-language text or voice message. The application installed on the devices of the sender and the recipient of the message carries the greatest load in the working mode. User - the sender must select the mode with the app, either press the SOS button. After that, you will start recording a video in which you need to send a gesture language message to the camera of your device. The application must provide a translation, display the result of the transfer on the sender's device, and send the message to the recipient's device.

II. ARCHITECTURE DESIGNING OF THE AUTOMATED SYSTEM OF PEOPLE SUPPORT WITH LIMITED COMMUNICATION

The authors analyzed the interaction of people with limited communication with the outside world and existing systems that help them communicate more comfortably in the social environment, as a result, the following requirements for supporting people with limited communication:

- Recognize and translate user gestures based on a neural network for to text or voice output;
- The speed of translation should be within 1-2 seconds;
- Ability to collect additional data for testing and learning the neural network recognition;
- The sender of the message has the opportunity to receive and reproduce the translation result on his device;
- A special mode of operation for emergencies (SOS button);
- Convenience and ease of use.

Based on the above requirements, the architecture of the support system for people with limited communication was designed (Fig. 1). In the physical decomposition of the architecture of the system, two main components are identified: the cloud information processing system (server or server group) and the user device. Each component contains a set of programs that ensure the functioning of the AS. The components are decomposed into subsystems that are responsible for the operation logic of the support system for people with limited communication.

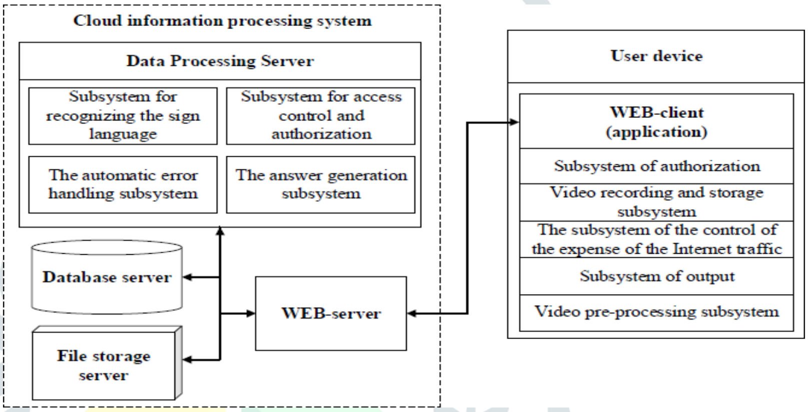

Fig 1 Architecture of support system for people with limited communication

Cloud information processing system contains a group of subsystems:

Data Processing Server:

- Subsystem for recognition based on a neural network;
- Subsystem for storing users' data and system operation data;
- Subsystem for registration and authorization of users;
- The subsystem for collecting, analyzing and testing the errors of the recognition algorithm.
  1. Database server;
  2. File storage server;
  3. Web server.

Table 4.1: Assignment of Subsystems For Support System of People With Limited Communication

Subsystem name	Subsystem functions
Data Processing Server	
Access control and authorization system	Performs registration, authorization and accounting of users in the system and theiraccess

Gesture language recognition system	Carries out recognition of the gesture language using the machine learning algorithm
Automatic error handling System	Performs a request and processing of source video materials from the user's device when an error is detected in recognition
Response Forming System	Generates and sends a message to a Recipient
User application (WEB-client)	
Authorization system	Authorize user in the system
Video recording and Storage	Record video using the camera of the user's device and store the recording until the translation process is completed
Internet traffic control System	Manages the volume of outgoing and incoming volume of Internet traffic
Output system	Carries out the output of the received translation in text format (on the device screen) or voice

A user device (smartphone, tablet, PC or other) contains an installed application (web client) that contains subsystems: authorization subsystem, video recording and storage subsystem, video pre-processing subsystem, subsystem for monitoring Internet traffic and subsystem for obtaining result of translation and output. Each subsystem provides the functionality listed in Table I.

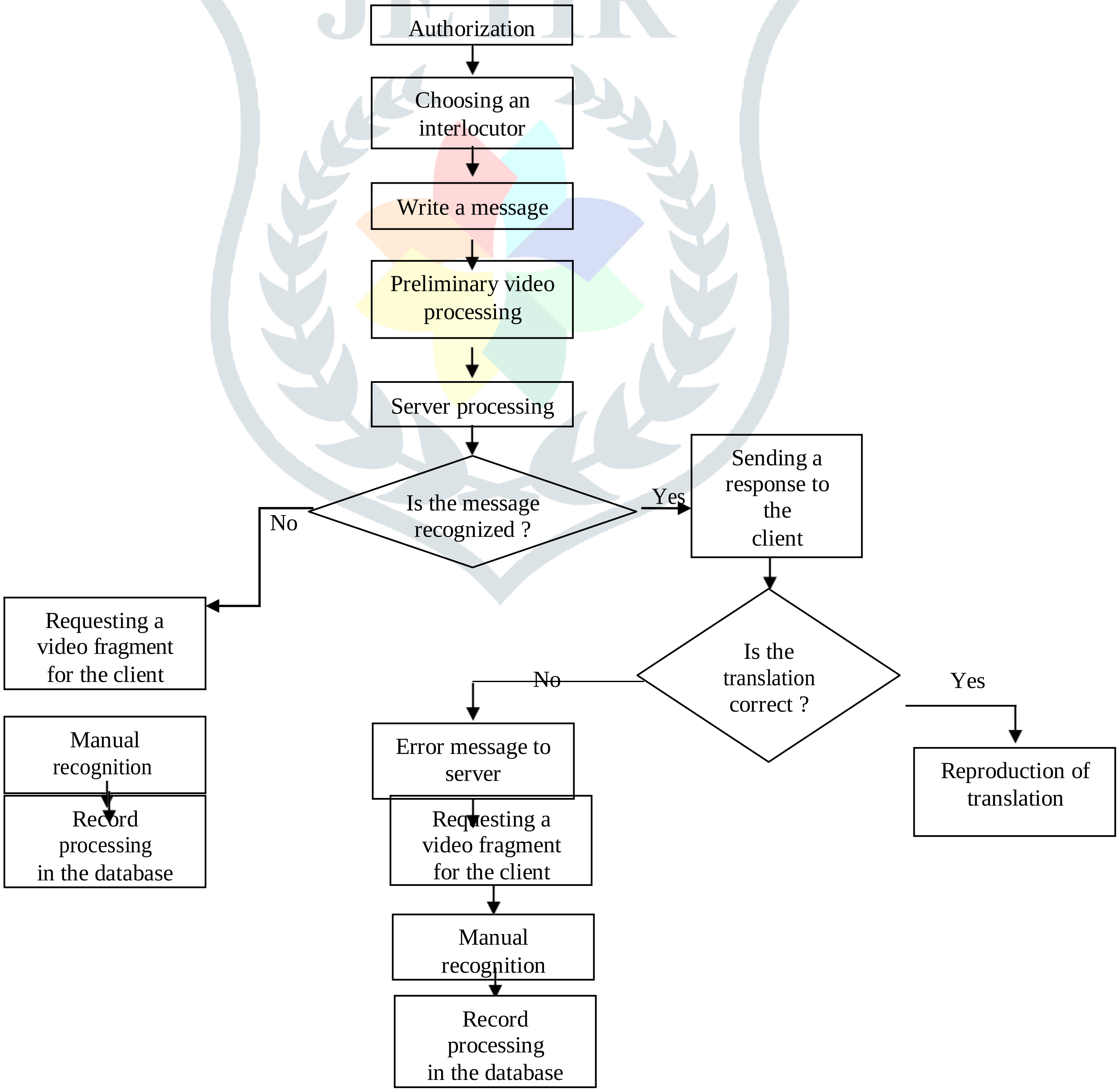

Fig 2 Script of the work of the support system of people with limited communication

Script work of the AS support people with limited communication includes the following sequence of actions. The user launches the application on his course, in case of a threat to life or health, an available SOS script (he will directly connect the user with the emergency response services by transferring geolocation data). If the user wants to send or reproduce his message, he selects the appropriate menu item. Starting the recording of messages, it shows gestures to the device's camera. in real time, the application records and makes a preview video then sends the data to the processing server, which processes and sends the translated message to the recipient selected, and a translation is displayed in text form so that he can understand the correctness of the recognition. The block diagram of the operation logic is shown in Fig. 2.

II. DESIGNING A USER APPLICATION

A. Subsystem for recording, storing and transmitting a video stream

There are two ways to transfer the video stream to the server: streaming (online broadcast) and recording in memory and video transmission. However, in order to minimize the use of the volume of Internet traffic, none of these ways of transferring the video stream is suitable. In the first way, there is a dependence on the quality of the Internet connection, video distortion is possible, due to losses during streaming data transmission. The second method of transmission uses a large amount of Internet traffic, depending on the speed and quality of the connection - it is possible to increase the transmission time, which is not suitable for real-time systems. To reduce the use of Internet traffic resources, you must enter the video pre-processing subsystem in the user Application. The subsystem, based on a neural network, must pre-process a section of the video: determine the position of the face and hands in the frame, create a map of points, record a map of points in an array of output data. Before the end of the recognition and output of the message, the video fragment will be stored in the temporary memory of the user device. In the event of a recognition error, the processing server will request the original video fragment from the user application for subsequent recognition and error detection.

B. Response output system

Based on the requirements for the AS, the user application must output the result of the translation in two versions: text and speech. The text output will be displayed on the device screen in a special field. To implement the voice output of the message, a voice packet with audio files of words and phrases is provided. The result of the translation will contain the sequence of launching the voice packet files, for playback of the user message. If the voice packet does not contain a specific set of sound files that match the response, a ready-made audio track will be sent from the data processing services that will be played on the user's device. The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

III. DESIGNING THE CLOUD SYSTEM

A. Subsystem for recognizing gestures

The main subsystem of the data processing server is the gesture recognition subsystem. Its development is based on a neural network of a recurrent type. These neural networks allow you to remember the previous values, thereby adding a single solution from the sequence of actions. In the support system for people with limited communication, we need to recognize gestures based on the movement of the hands and the head of a person, this type of neural network fits most correctly. The training set of a neural network is formed from processed video files located on the file storage server, which shows one or another gesture. At the first stage of the design and implementation of the AS, a training set for the battle gesture language of the Russian alphabet should be prepared. In the future, it is planned to retrain the algorithm for recognizing the words and phrases of the gesture language.

B. The automatic error handling subsystem

The recognition system may not recognize messages correctly. This is due to the fact that the video processing subsystem of the user application may incorrectly recognize the position or motion in the frame. In this connection on the server of the automated system, the subsystem of automatic error recognition is necessary. It will process the original video fragment received from the client according to the following algorithm: the original fragment is divided into frames, each frame is translated into a vector and written into the matrix of input data, the matrix of input data is calculated using the recursive neural network of the subsystem;

subsystems there was an error, or it was not (accidental or involuntary movement of the user), the results are entered in the error register and if necessary, the application is written in the specific subsystem.

C. Designing a data exchange system

Between the user application and the data server, need to ensure a secure data transfer process. In doing so, he must use a small amount of information transmitted. In the architecture of the user application, the subsystem of preliminary processing of video files is included. With her help, we will avoid the transfer of a large amount of information and the load on Internet traffic will appear only if the processing server recognition subsystem cannot correctly recognize the received data. Then a fragment of the video file will be transferred from the user application. Data from the user application must be transmitted in JSON format. This format is a text format for data exchange. It is used in different programming languages. It is convenient both for transmission and for use in server subsystems. Data from the data processing services will be transmitted depending on the type of response generated. In the case of text transmission, or an array of numbers of audio files to be launched, the JSON language will be used. If the server application will generate an audio file, the file will be transferred to the user application.

IV. TOOLS FOR IMPLEMENTING SOFTWARE

The user application will be installed on various devices, such as computers, tablets, and smartphones, you need to consider that they all work on different operating systems. To implement the gesture language recognition software, need to use the software and programming languages: Android Studio, Java, Xcode, Objective-C, Swift, C #, Apache, MySQL, MatLab, and Octave.

V. HARDWARE

For the stable functioning of the AS, it is necessary to determine the minimum technical characteristics that will be met by all the hardware used by the system. Minimum data server requirements: processor - Intel, 4- core, RAM - 4 GB, OS - Linux Ubuntu, Linux Debian, hard disk - 2 pieces, 1000 GB, in a RAID array, Internet connection - 15 Mbps. Minimum requirements for a smartphone running Android or iOS: processor - at least 1 Hz, RAM - 512 MB, OS -Android version 6.0, IOS version 9.0, internal memory – at least 100 MB of free space. The above requirements are tentative and are chosen based on the minimum requirements of programs and applications that solve similar problems. The exact characteristics of the hardware can be determined based on the statistics of the applications and the consumed computing resources of the devices.

VI. SIMULATION AND TESTING OF THE RECOGNITION SYSTEM PROTOTYPE

To make a decision on the choice of the neural network topology for the gesture recognition module, the authors conducted a test simulation of a number of ductile characters using the perceptron model in OCTAVE. The authors created a dataset consisting of 10,000 photographs depicting Russian gestures. It was divided into 20 classes, with 500 images for each class. To simplify used static gestures against a plain background. Fig. 3 shows how the computer sees the image as a rectangular matrix, each element of which corresponds to.

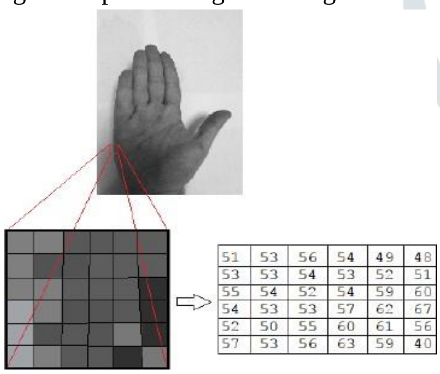

Fig 3a Representation of a fragment of the gray image in matrix form

For vectorization, transfer color image to black and white and reducing the size of using the program Octave and built-in package "image" to process the image. The image is loaded by the "imread" function, a three-dimensional matrix of size is obtained 4032*3024*3 pixels. Each "layer" represents the brightness of one of the RGB colors. The original images are converted to black and white format by function "rgb2gray", which reduces the size of the matrix by three times. To reduce the simulation time, the authors decided to reduce the size of photos by 60 * 80 pixels using the "imresize" function (Fig.4.)

Fig 3b Circuit kit for image processing

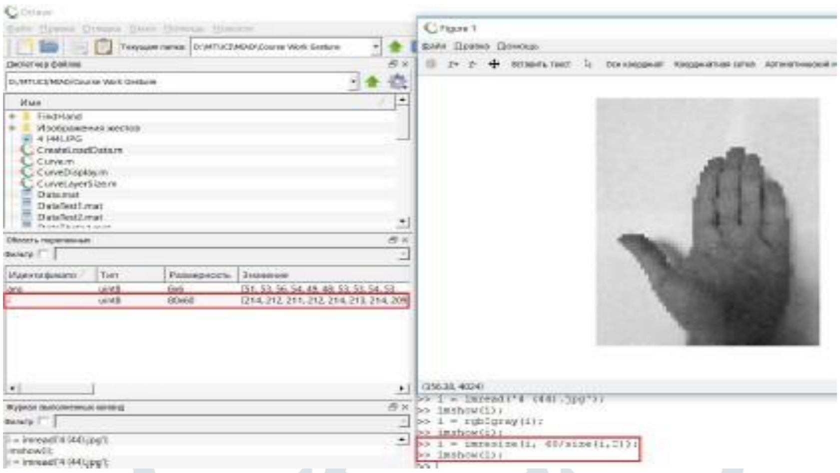

Fig 4 The process of reducing the image (the matrix and function are highlighted with red rectangles)

The resulting matrix is converted to a vector of dimension 4800. All vectors corresponding to a set of 10,000 photos from the dataset are written as rows in the multidimensional variable X of size 4800(row length)*10,000(number of columns). Save the resulting matrix in the file "Data.mat". The matrix X is the input to the neural network. We add another column to the input matrix, with the class numbers to which the images belong. Fig. 5 shows part of the training set.The topology of the neural network– a multilayer perceptron with back propagation of errors was chosen for modeling. Which contains an input layer containing an image matrix, a hidden layer consisting of 25 neurons, and an output layer (Fig. 6).

Neural network can suffer from "retraining", then it loses the ability to generalize. Because of this, the program is not able to recognize new material that is not included in the training set. To prevent this, the training data was divided into 2 parts: 70% of randomly selected data is used for training, and the remaining 30% for testing. For training we ask the error ε=10-3. The learning process is to minimize the cost function calculated by the formula:

$$J(\theta)=\frac{1}{2*m}*\sum_{i=1}^{m}(h_{\theta}(x^{(i)})-y^{(i)})^2 \qquad (1)$$

Where m - the number of training examples, x - the input feature vectors, y - the output variables, ($x^{(i)}, y^{(i)}$) - the i - th training set, θ - the parameters (weights) of the network.

Learning is reduced to optimizing the weights of the neural network, which are initialized randomly before learning. The learning process stops if the difference between the cost function values in the previous iteration and the current iteration is less than the specified error. The resulting sets of weights are recorded in variables "Theta1" and "Theta2" and stored in file: DataTheme1.mat" and "DataTheta2.mat" for later use when classifying new images. To test the neural network, take the selected 30% of the data. Let's try to recognize the gestures shown on them and calculate the accuracy of the classification according (2), where N - the number of correctly defined classes divided by
the size of The Xtest matrix, which displays the total number of test examples

$$A=\frac{N}{size(Xtest,1)}*100\% \qquad (2)$$

The accuracy of the classification was 96.6%. The trained and tested neural network was used to recognize multiple words based on sets of photos that were not related to the working dataset. (Fig. 7 presents some of the words)

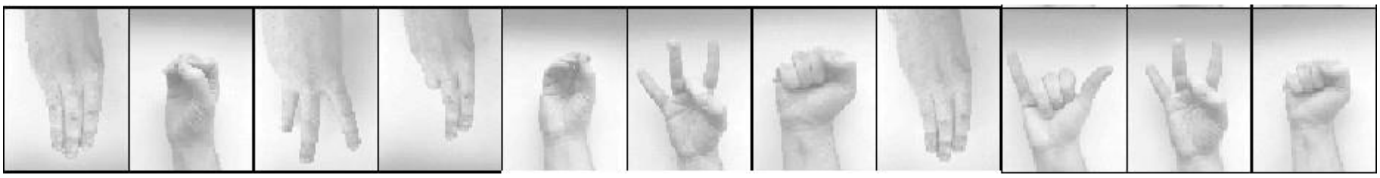

Fig 7 Photos for word recognition in developed Neural network

Each set of photos according to the scheme described above" loaded "into the variable/matrix Test, and for classification used the function "predict", the input of which was supplied to the matrix Test and matrix weights. At the output of the neural network we

Source word	The recognized word with the probability
Помогите Темнеет Врач	т(33%) о(96%) м(81%) о(81%) г(94%) и(83%) т(81%) е(94%) т(83%) е(86%) м(91%) н(78%) е(53%) е(96%) п(61%) в(95%) р(81%) а(74%) ч(57%)
Мне	м(46%) н(86%) е(92%)
Спасите	с(96%) п(81%) а(87%) с(96%) и(78%) т(85%) е(87%)
Нужен	н(91%) у(87%) ч(47%) е(92%) н(92%)
Путешествовать	т(45%) у(93%) п(63%) е(94%) ш(53%) е(94%) с(96%) т(57%) в(95%) и(85%) е(91%)
Жар	ж(66%) а(78%) р(97%)
Температура	т(89%) е(74%) м(83%) п(44%) е(91%) р(85%) а(90%) т(81%) у(92%) р(97%) а(89%)
Живот	ж(63%) и(87%) в(87%) о(96%) т(43%)

obtain the probability that the image belongs to this class. Select the class for which the probability is maximal. Fig.8 shows that the national Assembly incorrectly classifies three letters that makes some errors. Let us calculate the recognition error on a given array of examples using (3).Only 64 of the symbol, neural network made a mistake in 4 letters. The error is 0.0625 or 6.25%. Based on the results obtained, the following conclusions can be drawn(Fig.8)/

Fig 8 The results of testing the work of the national assembly

A lot of simplifications were used to implement the neural network prototype, which led to a decrease in the recognition quality, but for educational purposes, the model is quite suitable as a prototype that demonstrates the principle of gesture language recognition technique. In reality, neural networks with more complex topology should be used, for example, convolutional networks; syntactical analysis should be carried out to detect errors in recognized words, in case the network fails; high – performance computing methods should be used; large training sets should be created.

VII . CONCLUSION

The article describes the architecture design of an automated system to support people with limited communication for gesture recognition, with access and control from wearable mobile devices. The developed system consists of a variety of subsystems that perform tasks of authorization, recording, storage and conversion of video, sending data to the server and recognition Decomposition of the system into physical components and logical subsystems is made. The neural network with error back propagation is used for recognition. The topology of the used neural network is the perceptron. A training set of gestures was created and modeling of recognition subsystem based on neural network in Octave was made.

VIII . REFERENCES

[1] “Support-vector-machine-implementation-for-sign-languagerecognition-on-intel-edison”,2016 https://habrahabr.ru/company/intel/blog/306948/

[2] Zardoz (Cross modal comprehension in ZARDOZ an English to signlanguage translation system http://dl.acm.org/citation.cfm?id=1641450)

[3] Team (A Machine Translation System from English to American Gesture language http://dl.acm.org/citation.cfm?id=749243)

[4] ViSiCAST (VisiCAST Milestone: Final Report http://www.visicast.co.uk/members/milestones/milestones_list.htm)

[5] Saad M. et al. Ultrasonic hand gesture recognition for mobile devices //Journal on Multimodal User Interfaces. – 2017.– С.19.

[6] Chakraborty B. K. et al. Review of constraints on vision-based gesture recognition for human–computer interaction //IET Computer Vision. – 2017. – Т. 12. – №. 1. – С. 3-15.