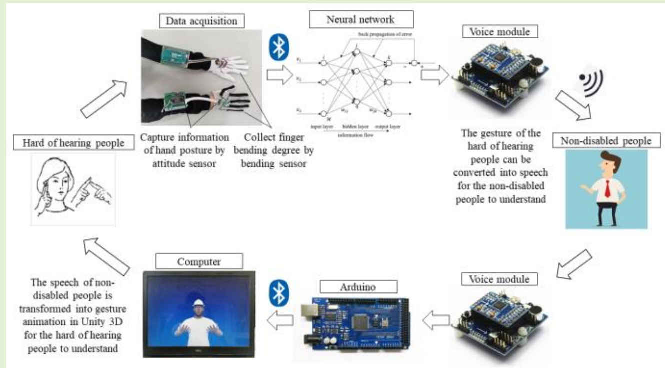
Design of Intelligent Human-Computer Interaction System for Hard of Hearing and Non-disabled People

Qiang Fu, Jiajun Fu, Songyuan Zhang, Xun Li, Jian Guo, and Shuxiang Guo, Fellow, IEEE

***Abstract*—Since the hard of hearing cannot communicate effectively with the non-disabled, which may cause various inconveniences. As an essential member of a harmonious society, it is particularly urgent to solve their communication problems with non-disabled people. Effective communication between the hard of hearing and the non-disabled has become possible with the continuous development of artificial intelligence. In this paper, an intelligent human-computer interaction system is designed to solve communication inconvenience between the hard of hearing and the non-disabled. This system combines artificial intelligence with wearable devices and classifies gestures with BP neural network, effectively solving the communication problem between the hard of hearing and the non-disabled.**



***Index Terms*—human-computer interaction, neural networks, gesture recognition, data gloves**

I. INTRODUCTION

DUE to the inability to communicate effectively with the non-disabled, the hard of hearing will face various daily travel and life difficulties. As a member of a harmonious society, the hard of hearing is an indispensable and vital part of a socialist harmonious society. Other non-disabled people need to effectively solve the communication difficulties between the hard of hearing and the non-disabled to help the hard of hearing to integrate into society positively and optimistically. At present, the effects of auxiliary devices used by the hard of hearing to communicate with the non-disabled are not ideal. For example, the electronic artificial throat lacks precise control of the time and pitch of voice. The use effect of visual-based sign language translation equipment is easily affected by the surrounding environment [1-5].

Therefore, it is of great practical value to study a portable, efficient and accurate human-computer interaction device for the communication between the hard of hearing and the non-disabled [6]. With the development of the Internet of Things and the Mobile Internet, wearable devices have completed the transition from large-scale to small-scale, which has attracted extensive attention from society [7]. At present, China has a large group of the hard of hearing. With the development of the cause of the disabled in our country, there is an increasing demand for the hard of hearing to participate in society [8]. However, corresponding sign language interpreters are quite scarce [9-15], which causes many difficulties for hard of hearing people in learning, living, working and other areas where they need to communicate [16]. According to data from the Sixth National Census and the Second National Survey of Disabled Persons, the total number of persons with disabilities in China has reached more than 85 million, accounting for approximately 6% of the total population of China [17]. Among them are at least 20.57 million hard of hearing people [18]. However, the sign language industry in China is developing slowly. Only a few teachers' colleges and universities provide sign language courses in particular education majors [19]. Sign language education, particularly social training, is weak. In various industries such as culture, medical care, social security, sports, commerce, and social services, the lack of sign language interpreters makes it hard for the deaf to communicate

This work was supported in part by Natural Science Foundation of Tianjin (18JCYBJC43200) and Tianjin Key Laboratory for Control Theory and Application in Complicated Systems (TJKL-CTACS-201903) and Innovative Cooperation Project of Tianjin Scientific and Technological Support (18PTZWHZ00090).

Corresponding authors: *Songyuan Zhang*.

Q. Fu and J. Guo are with Tianjin Key Laboratory for Control Theory & Application in Complicated Systems and Tianjin International Joint Research and Development Center, Tianjin University of Technology, Tianjin 300380, China (e-mail: fuqiang6369@hotmail.com; gj15102231710@163.com).

J. Fu and X. Li are with the Intelligent Robot Laboratory, Tianjin University of Technology, Tianjin 300380, China (e-mail: 779835867@qq.com; lix1o1@outlook.com).

S. Zhang is with State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150001, China (e-mail: zhangsy@hit.edu.cn).

S. Guo is with the Department of Intelligent Mechanical Systems Engineering, Kagawa University, Takamatsu 761-0396, Japan (e-mail: guo@eng.kagawa-u.ac.jp).

effectively with the outside world. For example, the hard of hearing always do not want to go to the hospital because of poor communication with doctors and it is considerably challenging to ask the police for directions, hard of hearing people have many difficulties in daily life. Therefore, developing an intelligent human-computer interaction system between the hard of hearing and the non-disabled will have a self-evident impact on the everyday communication between the hard of hearing and the non-disabled [20].

In this paper, we proposed an intelligent human-computer interaction system. In a variety of complex environments, it still maintains a high recognition rate. The gestures of the two hands of the hard of hearing were collected, and the data of the forearm were also collected to make the collected data more comprehensive and reflect the gestures of the hard of hearing more comprehensively. In addition, voice recognition is added and the recognition results are presented in the form of Unity 3D animation, which makes it easier for the hard of hearing to understand, and this system completes the two-way communication between the hard of hearing and the non-disabled.

II. Human-Computer Interaction Systems

The human-computer interaction system designed in this paper is mainly composed of data gloves, Arduino Mega2560 microcontroller, BP neural network, LD3320A voice module, wireless serial port module and the human model in Unity 3D. The overall human-computer interaction system for the hard of hearing and the non-disabled is shown in Fig. 1.

The data glove in Fig. 1 is used to collect posture data of the hand and forearm, and there are five bending sensors installed in the five finger positions to detect the bending degree of the fingers. The bending sensor is based on the resistance carbon element. The bend sensor achieves a significant form factor on a thin, flexible substrate as a variable printed resistance. When the substrate is bent, the sensor generates a resistance that is related to the bending radius. The smaller the output radius, the higher the resistance value. Attitude sensors were used to detect the posture of the palm and forearm. The bending sensor is used to detect the bending degree of 5 fingers, and the attitude sensor is used to detect the motion state of the upper limbs [21].

The calculation formula for the voltage of the bending sensor is given as Eq. (1).

$$V_{out} = V_{in}\left(\frac{R_1}{R_1 + R_2}\right) \quad (1)$$

The controller used to collect data is Arduino mega2560. The data formula of the analog voltage output in Arduino is given as Eq. (2).

$$A_0 = \frac{R_1}{R_1 + R_2} * 1023 \quad (2)$$

A_0 is the analog voltage output by the bending sensor at the analog port of the controller. When the bending angle of the bending sensor becomes more prominent, the resistance value of the bending sensor becomes smaller and the analog voltage distributed to the bending sensor decreases, A_0 becomes smaller. The data detected by each sensor is received through the Arduino mega2560 control chip, and the corresponding AD

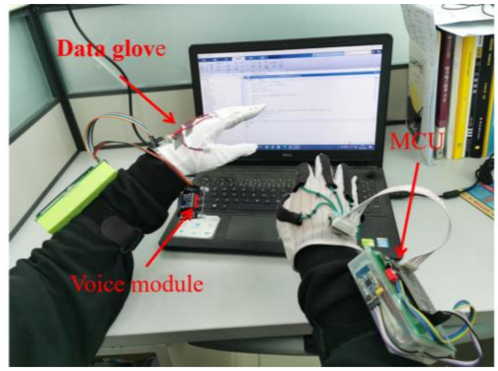


Fig. 1. The overall human-computer interaction system for the hard of hearing and the non-disabled.

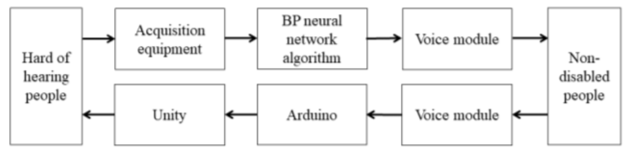


Fig. 2. Block diagram of human-computer interaction system for hard of hearing and non-disabled.

conversion is performed, and the voltage data of the finger and wrist collected by the ADC is printed out with the serial port.

Fig. 2 shows a block diagram of a human-computer interaction system for the hard of hearing and the non-disabled [22]. When the hard of hearing want to communicate with the non-disabled, the hard of hearing wear a data glove and makes gestures [23]. Using the bending sensor and the attitude sensor of the back of the hand and forearm to collect the bending signal of the finger of the hard of hearing and the 3-axis Euler angle data of the back of the hand and forearm, sending the collected data to the controller module. The collected data is wirelessly transmitted to MATLAB through the serial port. The gesture data is normalized and preprocessed to obtain the characteristic data. The purpose of normalization is to speed up training the network [24]. The purpose of changing a number to a decimal between (0,1) is to facilitate data processing. It is more convenient and faster to map the data to the range of 0 to 1 for processing [25]. Then putting the data into the trained BP neural network as the input of the neural network to get the output feature vector [26]. The obtained feature vector is wirelessly transmitted to the Arduino microcontroller through the serial port. After receiving the instruction, the microcontroller controls the LD3320A voice module to emit a sound corresponding to the gesture and converting the gesture information of the hard of hearing into the voice information that non-disabled person can understand. Accomplish the purpose of the hard of hearing want to communicate with the non-disabled [27].

When the non-disabled person wants to communicate with the hard of hearing person, the hard of hearing will press the button switch, then a non-disabled person can speak into the LD3320A voice module. The LD3320A voice module recognizes the sound. When the recognition is successful, the voice module sends a hexadecimal command to the Arduino microcontroller, and the Arduino microcontroller sends the 22 human body angle data corresponding to the hexadecimal command to Unity. The first five digits of the first 11 digits are the degree of bending of the right-hand finger, and the last six digits are the x, y, and z-axis spatial angles of the right hand and the right forearm, and the first five digits of the last 11 digits are the degree of bending of the left-hand finger, and the last six digits are the x, y, and z-axis spatial angles of the left hand and the left forearm [28]. These data control the characters in Unity 3D to make gesture animations corresponding to the voice. For example, translating the voice of learning this word into sign language actions will turn the voice information of the non-disabled person into a gesture animation that can be understood by the hard of hearing person, and complete the communication between the non-disabled person and the hard of hearing person [29].

III. Gesture Recognition And Voice Recognition

A. BP neural network

In the gesture recognition algorithm, the BP neural network algorithm is selected. Because neural net-works are characterized by self-learning and self-adaptive capabilities. The BP neural network can identify at a fast speed and store and process information in parallel, which is very suitable for the sign language recognition of the hard of hearing. Moreover, the neural network has a strong anti-interference ability and strong fault tolerance. Therefore, in this paper, the BP neural network technology is applied to the human-computer interaction system of the hard of hearing and the non-disabled [30]. Fig. 3 shows the BP neural network structure diagram.

Suppose the input data of BP neural network is n vectors, represented by vector X, the vector X is given as Eq. (3).

$$X = \left[x_0, x_1, \cdots, x_n\right] \tag{3}$$

The network generates m input data, represented by a vector Y, the vector Y is given as Eq. (4).

$$Y = \left(y_1, y_2, \cdots, y_m\right)^{\mathrm{T}} \tag{4}$$

Then the network will correspond to n input nodes and m output nodes. In this way, the BP neural network can be regarded as a nonlinear mapping from the n-dimensional input space to the m-dimensional output space.

BP neural network is divided into two parts in learning: forward propagation and back propagation. During forward propagation, the input information is transmitted from the input to the output layer after being processed by the hidden layer. The state of each layer of neurons only affects the state of the next layer of neurons. If the desired output cannot be obtained in the output layer, it will be transferred to back propagation, the error of the output layer node will be propagated back to the input to distribute to each connection node. Thus, the reference error of each connection point can be calculated, and the corresponding adjustment can be made according to the weight of each connection, so that the network can achieve the output which is suitable for the requirements, and the mapping from X to Y of the mode can be realized. The training process of the BP network is a process of constantly adjusting the weights. The training ends when the input error reaches the expected accuracy or reaches the preset number of learning and training times.

B. Feedforward calculation of BP neural network

In the learning stage of training the network, N training samples are set, and it is assumed that the input and output

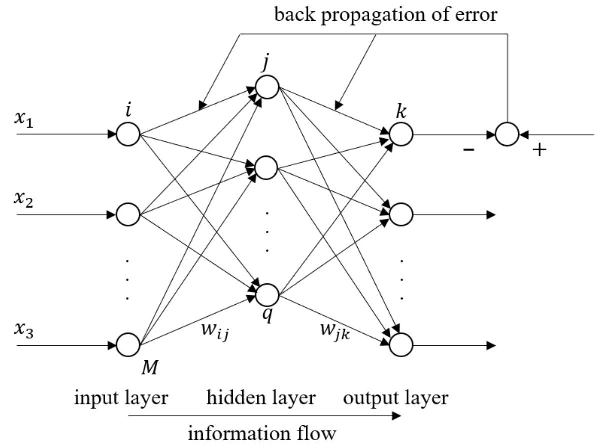


Fig. 3. BP neural network structure diagram.

mode and the X_p network $\{d_{pk}\}$ in one of the fixed samples are used for training. For ease of writing, the notation of sample P is temporarily omitted from the formula.

The input of the jth node of the hidden layer is given as Eq. (5).

$$\mathrm{net}_{pj} = \mathrm{net}_j = \sum_{i=1}^{M} w_{ij} o_i \tag{5}$$

The output of the jth node is given as Eq. (6).

$$O_j = f\left(\mathrm{net}_j\right) \tag{6}$$

The excitation function is given as Eq. (7).

$$f\left(\mathrm{net}_j\right) = \frac{1}{1 + \mathrm{e}^{-\frac{\left(\mathrm{net}_j - \theta_j\right)}{\theta_0}}} \tag{7}$$

where θ_j represents bias or threshold, and the positive θ_j is used to shift the excitation function to the right along the horizontal axis. The function of θ_0 is to adjust the shape of the δ function. The smaller θ_0 makes the δ function close to the step function, and the larger θ_0 makes the δ function flatter.

Take the derivative of Eq. (8) to obtain

$$f'\left(\mathrm{net}_j\right) = f\left(\mathrm{net}_j\right)\left[1 - f\left(\mathrm{net}_j\right)\right] \tag{8}$$

The output O_j of the jth node will be propagated forward to the kth node through the weighting coefficient W_{jk}, and the total input of the kth node of the output layer is given as Eq. (9).

$$\mathrm{net}_k = \sum_{j=1}^{q} w_{jk} o_j \tag{9}$$

The actual network output of the kth node in the output layer is given as Eq. (10).

$$O_k = f\left(\mathrm{net}_k\right) \tag{10}$$

The topological structure of the BP neural network model used in this system includes the input, hidden, and output layers. The number of nodes in the input layer and output layer has been determined in experiments. The most important thing is the number of nodes in the hidden layer. It needs to be estimated by calculation. Existing studies have found that the number of neurons in the hidden layer is related to the problems we need to solve, the degree of complexity, the characteristics of data samples and other factors. The number of hidden layers

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3107949, IEEE Sensors Journal

TABLE I
GESTURE CORRESPONDS TO A TRAINING SET SAMPLE OF A SET OF DATA

Gestures	Thumb	Index	Middle	Ring	Pinky	X1	Y1	Z1	X2	Y2	Z2
Digital 1	113	159	99	123	83	93.98	33.29	-23.09	-78.89	-37.67	-43.84
Digital 2	120	176	167	122	91	93.64	39.03	88.14	-87.50	-44.81	-50.21
Digital 3	121	102	162	162	117	115.74	16.62	-117.06	-74.75	-37.31	-74.73
Digital 4	112	173	162	164	130	115.23	5.60	-58.52	-90.68	-31.32	-46.92
Digital 5	164	158	165	158	135	93.47	22.63	-76.48	-91.86	-33.96	-40.62
Digital 6	152	111	110	130	110	120.76	30.44	-86.96	-96.32	-38.98	-41.71
Digital 7	120	142	124	136	129	99.05	0.68	-65.84	-83.56	-27.10	-60.18
Family	173	177	174	175	172	43.29	-11.05	7.93	-137.25	43.08	22.01
People	154	146	130	171	154	37.08	-11.76	20.36	-141.34	49.45	-3.53

TABLE II
GESTURES CORRESPONDING TO EACH FEATURE VECTOR

Feature vectors	Gestures
1 0 0 0 0 0 0 0 0	Digital 1
0 1 0 0 0 0 0 0 0	Digital 2
0 0 1 0 0 0 0 0 0	Digital 3
0 0 0 1 0 0 0 0 0	Digital 4
0 0 0 0 1 0 0 0 0	Digital 5
0 0 0 0 0 1 0 0 0	Digital 6
0 0 0 0 0 0 1 0 0	Digital 7
0 0 0 0 0 0 0 1 0	Family
0 0 0 0 0 0 0 0 1	People

TABLE III
THE RESULT OF GESTURE RECOGNITION

Gestures	A	B	C	D
Digital 1	98.0%	94.1%	93.1%	93.1%
Digital 2	97.1%	95.1%	94.1%	94.1%
Digital 3	96.1%	97.1%	96.1%	95.1%
Digital 4	98.0%	96.1%	95.1%	94.1%
Digital 5	97.1%	94.1%	95.1%	95.1%
Digital 6	95.1%	96.1%	96.1%	94.1%
Digital 7	96.1%	97.1%	96.1%	93.1%
Family	98.0%	97.1%	95.1%	94.1%
People	92.2%	93.1%	92.2%	93.1%

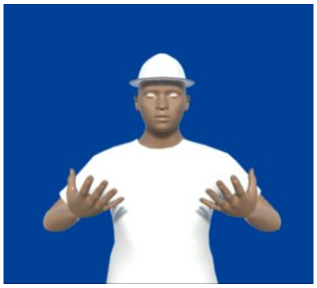


Fig. 4. Human body model in Unity 3D.

is determined as shown in Eq. (11-13).

$$m = \sqrt{n+1} + \alpha \tag{11}$$

$$m = \log_2 n \tag{12}$$

$$m = \sqrt{nl} \tag{13}$$

In the above formula, m represents the number of hidden layer nodes, l represents the number of output layer nodes, n represents the number of input layer nodes, and α represents a constant between 1-10.

The number of nodes in the hidden layer should be less than N-1(N is the number of training samples); otherwise, the systematic error of the network model is independent of the characteristics of the training samples and tends to 0. This means that the established network model has not generalization ability and use-value. Similarly, the number of nodes in the input layer must also be less than N-1. Through these three formulas, the maximum and minimum values are calculated, and the prediction performance of each network is compared step by step. The corresponding number of nodes with the best performance is selected as the number of neurons in the hidden layer, and the number of neurons in the hidden layer is determined to be 2.

The training process of the BP neural network includes three aspects: determining the input matrix, expected output and network layer parameters. First, the training sample set is read into the network, that is, the value of each gesture sensor, and then the expected output is read into the network. Finally, the BP neural network is trained according to the features of the training sample to create a neural network of the sample set. Table 1 is a set of data for each gesture of the training set sample.

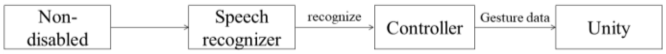


Fig. 5. Voice recognition flowchart.

The first 5 data in table 1 are the five analog voltage data of the bending sensor installed on the glove finger. The last three are the data of the 3 Euler angles of the attitude sensor of the wrist joint, and the last 3 data are the Euler angle data of the elbow joint. A total of 11 bits of data can collect the gesture information of the whole hand more comprehensively.

Correct recognition of gestures is critical. Table 2 shows the result of the feature vector output by the BP neural network corresponding to 9 common gestures.

C. The design of voice recognition

The voice recognition module used in this article is LD3320A, which combines the functions of voice recognition and voice output simultaneously, avoiding the disadvantages of using two modules and reducing the size of the system. As shown in Fig. 4, a human body 3D model is established in the human body modeling software. After the voice recognition module LD3320A recognizes the voice of a healthy person, the human body model will demonstrate the human body animation in Unity 3D for the hard of hearing to understand.

Fig. 5 shows a voice recognition flowchart; when the non-disabled person wants to communicate with the hard of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3107949, IEEE Sensors Journal

TABLE IV
VOICE RECOGNITION CORRESPONDING TO THE DATABASE OF THE HUMAN BODY MODEL

Voice	Data vector of human body model																				
Learning	8	12	4	7	20	15	-30	4	130	0	-10	8	12	4	7	18	15	30	-4	130	0	10
You	90	12	45	45	45	-30	30	-20	30	-180	-205	8	12	4	7	18	15	30	-4	50	-180	230
One	100	12	30	30	30	-30	30	60	35	-18	-205	8	12	4	7	20	15	30	-4	50	-180	230
Two	100	12	12	30	30	-30	30	60	35	-18	-205	8	12	4	7	20	15	30	-4	50	-180	230
Three	100	30	12	12	12	-30	30	60	35	-18	-205	8	12	4	7	20	15	30	-4	50	-180	230
Four	100	12	12	12	12	-30	30	60	35	-18	-205	8	12	4	7	20	15	30	-4	50	-180	230
I	45	11	46	42	19	66	-74	-140	76	-113	-130	8	12	4	7	18	15	-83	-4	43	0	0
Be	10	12	4	77	78	15	83	4	-14	-77	-161	8	12	4	7	18	15	-86	-4	42	0	0
Deaf	8	12	4	58	59	15	83	4	16	56	93	8	12	4	7	18	15	-8	-4	43	0	0
Home	8	12	4	7	18	15	-83	-4	27	162	180	8	12	4	7	19	15	83	4	23	-180	-180

hearing person, the non-disabled person will make a sound to the LD3320A voice module. The LD3320A voice module recognizes the sound. When the recognition is successful, the voice module sends a hexadecimal command to the Arduino microcontroller, and the Arduino microcontroller sends the 22 human body angle data corresponding to the hexadecimal command to Unity. Table 4 shows the voice recognition corresponding to the database of the human body model. The first five digits of the first 11 digits are the degree of bending of the right-hand finger, and the last six digits are the x, y, and z-axis spatial angles of the back of the right hand and the right forearm, and the first five digits of the last 11 digits are the degree of bending of the left-hand finger, and the last six digits are the x, y, and z-axis spatial angles of the back of the left hand and the left forearm. These data control the characters in Unity 3D to make gesture animations corresponding to the voice. For example, translating the voice of learning this word into sign language actions will turn the voice information of the non-disabled person into a gesture animation that can be understood by the hard of hearing person, and complete the communication between the non-disabled person and the hard of hearing person.

D. Mathematical Model of Human Upper Limb in Unity

Suppose the length of the forearm is d_1, and the length of the palm is d_2. The coordinates of the end of the forearm in the left-hand coordinate system are (x_1, y_1, z_1), The coordinate of the palm end S in the left-hand coordinate system is (x_2, y_2, z_2), The coordinate system is set at the elbow joint as shown in Fig. 6. Setting a coordinate system $O_1X_gY_gZ_g$, $O_2X_gY_gZ_g$ at the elbow joint and wrist joint respectively. By obtaining the spatial posture data γ_1, θ_1, ψ_1, γ_2, θ_2, ψ_2, of each limb relative to the left-hand coordinate system, the movement position of each limb at each time can be found.

After projecting the end L of the forearm, the coordinates (x_1, y_1, z_1) of the end L are given as Eq. (14).

$$\begin{cases} x_1 = d_1 \sin\theta_1 \\ y_1 = d_1 \cos\theta_1 \cos\psi_1 \\ z_1 = d_1 \cos\theta_1 \sin\psi_1 \end{cases} \tag{14}$$

The projection (x_2', y_2', z_2') of the end of the forearm S in the coordinate system $O_2X_gY_gZ_g$ is given as Eq. (15).

$$\begin{cases} x_2' = d_2 \sin\theta_2 \\ y_2' = d_2 \cos\theta_2 \cos\psi \\ z_2' = d_2 \cos\theta_2 \sin\psi_2 \end{cases} \tag{15}$$

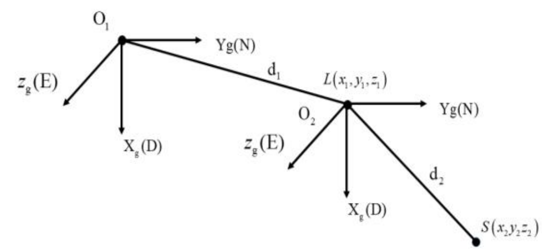


Fig. 6. The position of the upper limbs of the human body in the coordinate system.

The forearm end S coordinates (x_2, y_2, z_2) in the geographic coordinate system are given as Eq. (16).

$$\begin{cases} x_2 = x_1 + x_2' \\ y_2 = y_1 + y_2' \\ z_2 = z_1 + z_2' \end{cases} \tag{16}$$

Namely

$$\begin{cases} x_2 = d_1 \sin\theta_1 + d_2 \sin\theta_2 \\ y_2 = d_1 \cos\theta_1 \cos\psi_1 + d_2 \cos\theta_2 \cos\psi_2 \\ z_2 = d_1 \cos\theta_1 \sin\psi_1 + d_2 \cos\theta_2 \sin\psi_2 \end{cases} \tag{17}$$

The angle between two vectors in space is given as Eq. (18).

$$\cos\delta = \frac{x_1x_2 + y_1y_2 + z_1z_2}{d_1d_2} \tag{18}$$

The angle between the palm and the forearm $\cos\delta_{1-2}$ is given as Eq. (19).

$$\cos\delta_{1-2} = 180^\circ - \arccos\left(\frac{x_1x_2' + y_1y_2' + z_1z_2'}{d_1d_2}\right) \tag{19}$$

Through these formulas, the spatial position of the forearm L and the palm S of the human body model in Unity 3D can be obtained, and the relative angle between the forearm and the palm can be calculated to control the movement of the human body model in Unity 3D.

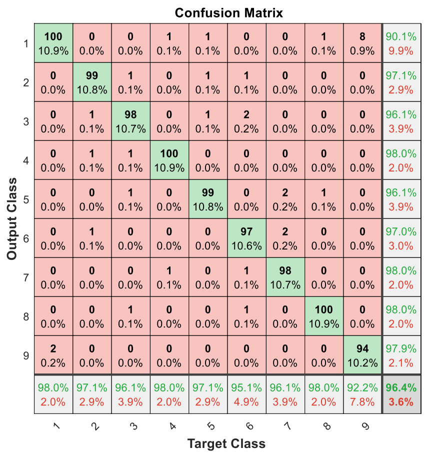
IV. EXPERIMENTS AND RESULTS

The experiment results for the hard of hearing communicating with the non-disabled are reflected in the pictures of MATLAB. After the BP neural network is trained with the data gloves, the data is transmitted to MATLAB in real-time, which can realize the correct identification from 9 kinds of gestures. MATLAB will pop up the picture corresponding to the gesture, and the voice module will emit the corresponding voice when the gesture is recognized. Fig. 7

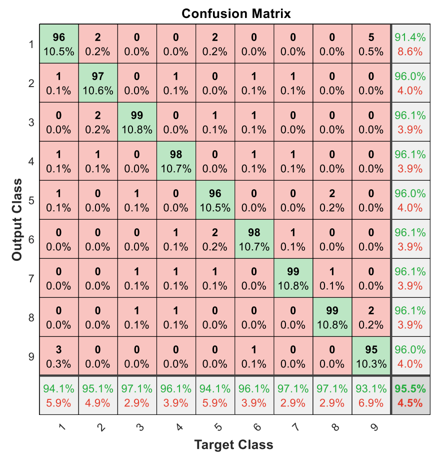
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3107949, IEEE Sensors Journal

(a). Identification number 1 (b). Identification number 2 (c). Identification gesture family (d). Identification gesture people

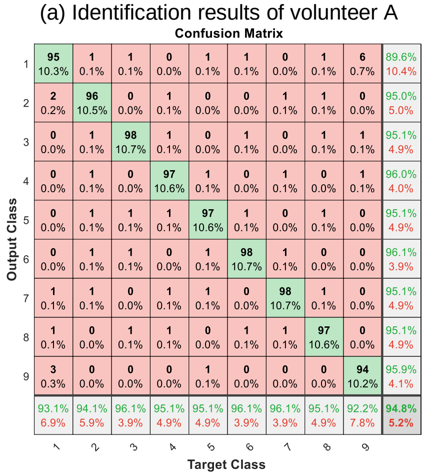
Fig. 7. Recognition results.



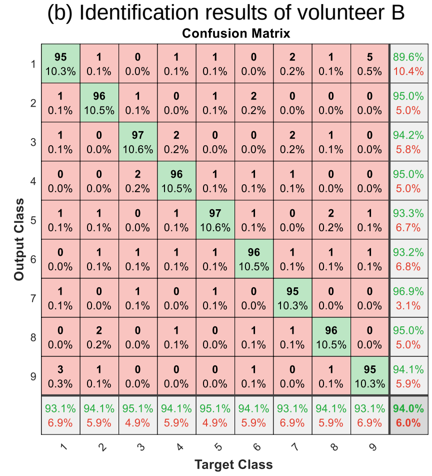
(a) Identification results of volunteer A



(b) Identification results of volunteer B



(c) Identification results of volunteer C



(d) Identification results of volunteer D

Fig. 8. The confusion matrix of the performance.

shows the correct recognition of the numbers 1 to 2, people and family gestures.

In the motion recognition experiment, four healthy volunteers (aged 22-26 years old, half male and female, two left-handed and two right-handed) were selected for the experiment, and each motion was repeated 20 times for each volunteer. In each motion, five samples were selected for collection at the sampling rate of 9600Hz. There were 102 samples of each gesture and nine kinds of gestures from 1 to 9. Each volunteer collected a total of 918 sets of data. Table 3 lists the results of the recognition accuracy rate. Among them, A, B, C and D respectively represent four volunteers.

In this system, after different volunteers put on the data glove, the range of the bending sensor and the attitude sensor is different. If the raw data collected by the data glove is used directly, the experimental results may be affected. In this paper, the maximum-minimum standardization is used to normalize the data, so that the indicators are in the same order of magnitude, convenient for comprehensive comparison and

TABLE V
THE RESULT OF VOICE RECOGNITION

Voice	A	B	C	D
Learning	93.3%	90%	86.7%	90%
You	90%	86.7%	90%	86.7%
One	93.3%	96.7%	90%	90%
Two	86.7%	90%	93.3%	93.3%
Three	90%	93.3%	90%	86.7%
Four	86.7%	96.7%	96.7%	83.3%
I	90%	90%	86.7%	86.7%
Be	96.7%	93.3%	90%	90%
Deaf	90%	86.7%	93.3%	93.3%
Home	93.3%	90%	86.7%	90%
Average correct rate	91%	91.3%	90.3%	89%

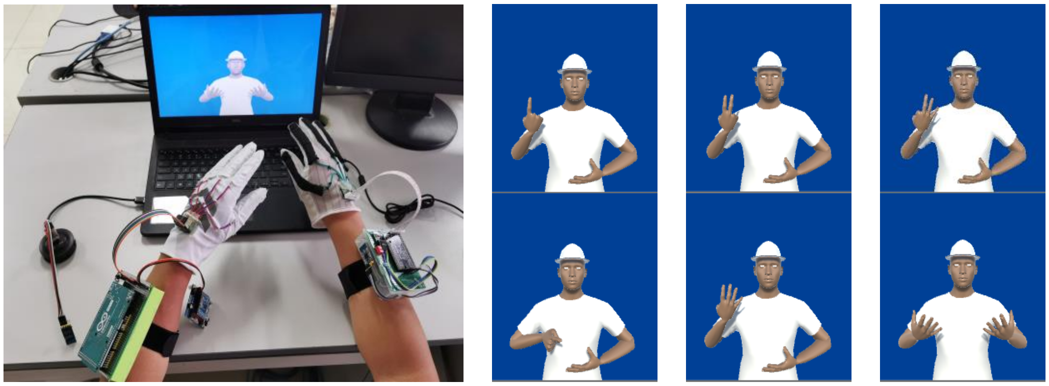


Fig. 9. Voice recognition results.

evaluation. Maximum-minimum standardization is a linear transformation of the original data. Let min A and max A be the minimum and maximum values of attribute A, respectively, and map an original value x of A to the interval [0, 1] through maximum-minimum standardization. The value x' is given as Eq. (20).

$$x^{'} = \frac{x - \min A}{\max A - \min A} \tag{20}$$

Fig. 8 shows the confusion matrix of the BP neural network. The rows correspond to the predicted classes, and the columns correspond to the real classes. The diagonal cells correspond to correct classification results. The cells outside the diagonal correspond to misclassified results. The right-most column of the confusion matrix graph represents the percentage of all examples predicted for each class that falls into the correct and false categories. These indicators are referred to as precision. The precision function is given as Eq. (21).

$$\text{Precision } = \frac{TP}{TP + FP} \tag{21}$$

where, TP is true positive that means the gestures sample was correctly identified. FP is false positive and means the gestures sample is wrongly classified.

The rows at the bottom of the confusion matrix graph show the percentage of all examples belonging to each category, correctly and incorrectly classified. These metrics are commonly referred to as recall. Recall is the percentage of the true positives over the number of true positives and the number of false negatives. Recall function is given as Eq. (22).

$$\text{Recall } = \frac{TP}{TP + FN} \tag{22}$$

where, FN is the false negative that means the gestures sample does not belong to the current category. The cell in the lower right corner of the confusion matrix diagram shows the overall accuracy, which represents the accuracy of the overall judgment of the classification model.

The result of the experiment that the non-disabled communicate with the hard of hearing is realized in Unity 3D. The non-disabled say the six words the number 1, number 2, number 3, number 4, study and you, and convert these sounds into human animations to show them to the hard of hearing, so that the hard of hearing can understand. As shown in Fig. 9. The voice recognition rate is shown in Table 5. The overall recognition rate of voice recognition was 90.4%.

V. Conclusions

In this paper, an intelligent human-computer interaction system for the hard of hearing and non-disabled is proposed. When the hard of hearing wants to communicate with the non-disabled, the system collects the gesture data of the hard of hearing in real time through the data glove, and BP neural network was used for classification, and finally broadcasts the result of the classification to the non-disabled by voice, so that the non-disabled can understand. The nine kinds of gestures are recognized, and the overall recognition rate is 95.18%.

When a non-disabled person wants to communicate with a hard of hearing person, the LD3320A voice module recognizes the sound of the non-disabled person's voice, and uses the human body model in Unity 3D to express the gestures represented by the voice, so that the hard of hearing person can understand. The human-computer interaction system between the hard of hearing and the non-disabled completed in this paper effectively solves the communication problem between the hard of hearing and the non-disabled. The above experimental results also prove the possibility and real-time nature of this system.

References

[1] Y. Peng, J. Wang, K. Pang, W. Liu, J. Meng and B. Li, "A Physiology-Based Flexible Strap Sensor for Gesture Recognition by Sensing Tendon Deformation," *IEEE Sensors J.*, vol. 21, no. 7, pp. 9449-9456, 1 Aprill1, 2021.

[2] K. Liu, C. Chen, R. Jafari and N. Kehtarnavaz, "Fusion of Inertial and Depth Sensor Data for Robust Hand Gesture Recognition," *IEEE Sensors J.*, vol. 14, no. 6, pp. 1898-1903, June 2014.

[3] R. Xie and J. Cao, "Accelerometer-Based Hand Gesture Recognition by Neural Network and Similarity Matching," *IEEE Sensors J*, vol. 16, no. 11, pp. 4537-4545, June1, 2016.

[4] Y. Zhang et al., "Static and Dynamic Human Arm/Hand Gesture Capturing and Recognition via Multiinformation Fusion of Flexible Strain Sensors," *IEEE Sensors J.*, vol. 20, no. 12, pp. 6450-6459, 15 June15, 2020.

[5] B. G. Lee and S. M. Lee, "Smart Wearable Hand Device for Sign Language Interpretation System With Sensors Fusion," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1224-1232.

[6] G. Yuan, X. Liu, Q. Yan, S. Qiao, Z. Wang and L. Yuan, "Hand Gesture Recognition Using Deep Feature Fusion Network Based on Wearable Sensors," *IEEE Sensors J.*, vol. 21, no. 1, pp. 539-547, 1 Jan.1, 2021.

[7] U. von Agris, J. Zieren, U. Canzler, B. Bauer, and K.-F. Kraiss, "Recent developments in visual sign language recognition," *Universal Access Inf. Soc.*, vol. 6, no. 4, pp. 323–362, Feb. 2008.

[8] S. G. M. Almeida, F. G. Guimarães, and J. A. Ramírez, "Feature extraction in Brazilian sign language recognition based on phonological

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3107949, IEEE Sensors Journal

structure and using RGB-D sensors," *Expert Syst. Appl*., vol. 41, no. 16, pp. 7259–7271, 2014.

[9] Z. Lu, X. Chen, Q. Li, X. Zhang, and P. Zhou, "A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices," *IEEE Trans. Human-Mach. Syst*., vol. 44, no. 2, pp. 293–299, Apr. 2014.

[10] W. C. Stokoe, Jr., "Sign language structure: An outline of the visual communication systems of the American deaf," *J. Deaf Stud. Deaf Edu*., vol. 10, no. 1, pp. 3–37, 1960.

[11] N. Sriram and M. Nithiyanandham, "A hand gesture recognition based communication system for silent speakers," presented at the *Int. Conf. Hum. Comput. Interact*., Chennai, India, Aug. 2013.

[12] S. P. More and A. Sattar, "Hand gesture recognition system using image processing," presented at the *Int. Conf. Elect., Electron*., Opt. Techn., Chennai, India, Mar. 2016.

[13] S. Shin and W. Sung, "Dynamic hand gesture recognition for wearable devices with low complexity recurrent neural networks," in *Proc. IEEE Int. Symp*. Circuits Syst. (ISCAS), Montreal, QC, Canada, May 2016, pp. 2274–2277.

[14] J. C. Núñez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Vélez, "Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition," *Pattern Recognit*., vol. 76, pp. 80–94, Apr. 2018.

[15] Y. Li, N. Yang, L. Li, L. Liu, and Y. Yang, "Finger gesture recognition using a smartwatch with integrated motion sensors," *Web Intell*., vol. 16, no. 2, pp. 123–129, Jun. 2018.

[16] Q. Bu, G. Yang, X. Ming, T. Zhang, J. Feng, and J. Zhang, "Deep transfer learning for gesture recognition with WiFi signals," *Pers. Ubiquitous Comput*., pp. 1–12, Jan. 2020.

[17] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Trans. Syst*., Man, Cybern. A, Syst., Humans, vol. 41, no. 6, pp. 1064–1076, Nov. 2011.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput*. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[19] J. Singha, A. Roy, and R. H. Laskar, "Dynamic hand gesture recognition using vision-based approach for human–computer interaction," Neural Comput. Appl., vol. 29, no. 4, pp. 1129–1141, Feb. 2018.

[20] R. Naoum, H. H. Owaied, and S. Joudeh, "Development of a new Arabic sign language recognition using K-nearest neighbor algorithm," *J. Emerg. Trend. Comput. Infor. Sci*., vol. 3, no. 8, pp. 1173–1181, Aug. 2012.

[21] W. Jingqiu and Z. Ting, "An ARM-based embedded gesture recognition system using a data glove," presented at the 26th Chinese Control Decision Conf., Changsha, China, May/Jun. 2014.

[22] A. Z. Shukor, M. F. Miskon, M. H. Jamaluddin, F. A. Ibrahim, M. F. Asyraf, and M. B. Bahar, "A new data glove approach for Malaysian sign language detection," *Procedia Comput. Sci*., vol. 76, no. 1, pp. 60–67, Dec. 2015.

[23] S. V. Matiwade and M. R. Dixit, "Electronic support system for deaf and dumb to interpret sign language of communication," *Int. J. Innov. Res. Sci., Eng. Technol*., vol. 5, no. 5, pp. 8683–8689, May 2016.

[24] S. Goyal, I. Sharma, and S. Sharma, "Sign language recognition system for deaf and dumb people," *Int. J. Eng. Res. Technol*., vol. 2, no. 4, pp. 382–387, Apr. 2013.

[25] Y. Zhu, S. Jiang, and P. B. Shull, "Wrist-worn hand gesture recognition based on barometric pressure sensing," in Proc. IEEE 15th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN), Las Vegas, NV, USA, Mar. 2018, pp. 181–184.

[26] L. Lamberti and F. Camastra, "Real-time hand gesture recognition using a color glove," presented at the Int. Conf. Image Anal. Process., Ravenna, Italy, Sep. 2011.

[27] S. P. Dawane and H. G. A. Sayyed, "Hand gesture recognition for deaf and dumb people using GSM module," *Int. J. Sci. Res*., vol. 6, no. 5, pp. 2226–2230, May 2017.

[28] D. Lu, Y . Yu, and H. Liu, "Gesture recognition using data glove: An extreme learning machine method," in *Proc. IEEE Int. Conf. Robot*. Biomimetics, Qingdao, China, Dec. 2016.

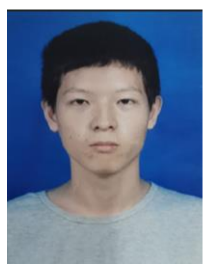
[29] A. Prakash, N. Sharma, and S. Sharma, "Novel force myography sensor to measure muscle contractions for controlling hand prostheses," Instrum. Sci. Technol., vol. 48, pp. 43–62, Jan. 2019.

[30] J. Wu, Z. Tian, L. Sun, L. Estevez, and R. Jafari, "Real-time American sign language recognition using wrist-worn motion and surface EMG sensors," presented at the IEEE 12th Int. Conf. Wearable Implant. Body Sensor Netw., Cambridge, MA, USA, Jun. 2015.

Qiang Fu received the Ph.D. degree in intelligent mechanical systems engineering from Kagawa University, Takamatsu, Japan, in 2017. He is currently an associate professor in Tianjin University of Technology, Tianjin, China. His current research is on biomedical robots, such as the development of multi-module magnetic actuated microrobotic systems for biomedical applications.

Jiajun Fu is currently working toward the M.S degree in College of Electrical and Electronic Engineering from Tianjin University of Technology, Tianjin, China. His research interests include the design of the Human-computer interaction system.

Songyuan Zhang received the Ph.D. degree in intelligent mechanical systems engineering from Kagawa University, Takamatsu, Japan, in 2016. He is currently an associate professor in State Key Laboratory of Robotics and System, Harbin Institute of Technology, China. His research interests include bio-inspired robotics, human robot interaction and medical robots.



Xun Li is currently working toward the M.S degree in College of Electrical and Electronic Engineering from Tianjin University of Technology, Tianjin, China. His research interests include the design of the variable stiffness robot joint.

Jian Guo received the Ph.D. degrees in intelligent machine system from Kagawa University, Japan, in 2012. He is currently a full professor in Tianjin University of Technology, Tianjin, China. He has published about forty refereed journal and conference papers in recent three years. His current research is on biomedical robots, such as wireless microrobots in pipe and robotic catheter systems for biomedical applications.

Shuxiang Guo (SM'03，F'21) received the Ph.D. degree in mechanoinformatics and systems from Nagoya University, Japan, in 1995. He is currently a full professor at the Faculty of Engineering and Design, Kagawa University, Japan. He received the Chang Jiang Professor-ship Award from the Ministry of Education of China, in 2005. Dr. Guo is Editor in chief for the International Journal of Mechatronics and Automation.