Hindawi
Advances in Multimedia
Volume 2022, Article ID 6141295, 6 pages
https://doi.org/10.1155/2022/6141295

Research Article

Research and Implementation of Intelligent Home Pension System Based on Speech and Semantic Recognition

Guokun Xie,[1] **Sen Hao**,[2] **Peipei Zhang**,[1] **and Ningning Wang**[1]

[1]*Xi'an Traffic Engineering Institute, Xi'an 710000, China*
[2]*Xi'an Huawei Technology Co., Ltd., Xi'an 710000, China*

Correspondence should be addressed to Guokun Xie; 201801010418@stu.hnie.edu.cn

Received 23 December 2021; Revised 18 February 2022; Accepted 28 February 2022; Published 30 March 2022

Academic Editor: Qiangyi Li

In order to study the current intelligent home pension system control technology, the new voice interaction technology is applied to the existing intelligent home pension system, and a multicontrol entry intelligent home pension system method is proposed. In this method, data transmission, instruction uploading, and receiving are completed by designing the communication interface of home appliance terminal to build the wireless intelligent home communication subsystem, and the voice module of IfI is adopted. The intelligent cloud is selected as the development cloud platform, and the related hardware is selected to realize the remote data communication. It has been proved that the key technology of speech recognition has been developed rapidly, the speech recognition rate has been improved (up to 97%) and the low-power speech wake up technology had breakthrough, the use of voice interaction is gradually expanding to intelligent hardware and robots, and voice interaction is undoubtedly the mainstream intelligent home pension system interaction mode after keyboard, mouse, and touch screen and also the main entrance of the future intelligent home pension system.

1. Introduction

With the progress of human society and the rapid development of science, people began to pursue intelligent and convenient home environment, so the intelligent home pension system has been rapidly rising. Intelligent home pension system refers to the integrated control and management of life-related facilities and equipment mainly based on people's living environment by using sensor network, automatic control technology, wireless communication technology, etc., so as to improve the safety, convenience, comfort, and artistic quality of home and, at the same time, achieve an environmentally friendly and energy-saving living environment [1]. In recent years, new breakthroughs have been made in key technologies such as far-field speech recognition and cloud semantic recognition. The gradually mature application environment has ignited the enthusiasm of giants to develop intelligent home pension system again. Voice is the most natural and convenient human-computer interaction method in parallel with touch control, gesture, eye movement, brainwave monitoring, and other human-computer interaction methods. It is simple and easy to use [2]. In her 2016 Internet trends report, MaryMeeker, the "queen of the Internet," argued that, for the past 75 years, from the very beginning of Tactile 1.0 to the current touch screen of tactile 3.0 to the most effective form of voice interaction, the basic human-computer interaction mode is constantly innovated every ten years [3]. Voice interaction frees hands and eyes and can be used at home, in the car or anywhere, expanding the concept of "always-on." The combination of voice technology and intelligent home pension system appliances gives birth to smart speaker. Smart speaker in addition to Bluetooth speaker music playing basic skills home device control, more importantly, as a new generation of intelligent home pension system device center, with intelligent equipment scene management function. Smart speakers realize the control of TV, curtain, lighting, air conditioning, and other home appliances based on voice interaction, get rid of the restriction of having to rely on mobile phones, remote controls, and other terminals,

and can conveniently manage intelligent home pension system appliances at any time [4]. Especially, for the current generation of elderly users, the cost of learning to operate a mobile app (App) or a home control panel is very high, and by sending control instructions to the intelligent speaker in the way of natural voice, the device can be controlled simply and humanized. Therefore, voice, as an important way of human-computer interaction in the era of Internet of Things, is favored by all parties and is the best way to realize intelligent home pension system at the present stage.

Researchers apply dynamic programming algorithms to speech recognition, Tiede et al. proposed the dynamic time regularization algorithm, which was a hot topic in speech recognition research; at that time, it is mainly a structured calculation of speech signal combining time and distance and has been widely used in speech recognition studies with small vocabularies [5]. Later, Zhang proposed the most famous hidden Markov model theory in the field of speech recognition, which opened up a new research direction for speech recognition research, especially in recognition algorithm [6].

By analyzing the principle and key technology of ASR speech recognition, it is pointed out that the current speech recognition accuracy has reached the practical condition, has been applied to intelligent hardware products such as smart speaker, theoretically understood the speech interaction technology, analyzed the key speech technologies involved in speech interaction, including speech wake up, speech recognition, far-field speech pick up, and speech synthesis technology, and explained the importance of speech data preprocessing; from the perspective of application, the influence of microphone array on speech recognition accuracy is analyzed, and the current data processing methods of microphone array are studied.

2. Speech Interaction Theory

Human-machine interaction (HMI) is a study of the Interaction between a system and a user. At present, in the intelligent home pension system, the control methods used are buttons, gesture control, mobile terminals, and smart speakers, among which the voice control method based on smart speakers is the latest and most natural interactive way. The interaction process is shown in Figure 1, including five processes, speech recognition, natural language understanding, conversation management, natural language generation, and text-to-speech.

2.1. Speech Recognition. Automatic speech recognition (ASR) takes speech as the research object; speech signals are converted into computer-readable input through speech signal processing and pattern recognition so that the machine can automatically recognize and understand the spoken language [7]. Speech recognition is a wide range of interdisciplinary, belongs to the field of signal processing science, and acoustics, linguistics, phonetics, information theory, and pattern recognition and other disciplines have a very close relationship. Speech recognition technology is approaching us step by step, filling every link of our life, and gradually becoming the key technology in computer information processing technology; the application of voice technology has become a competitive new high-tech industry. Speech recognition system is essentially a pattern recognition system, including feature extraction, feature matching, and basic units of model base. The principle of speech recognition is shown in Figure 2.

The user's voice signal is collected by the microphone, and the digital signal is obtained after the preprocessing such as noise reduction and filtering. Then, the acoustic features are extracted and the pattern matching is carried out, and the recognition result is obtained after the processing. At present, the main recognition methods include dynamic time warping (DTW) algorithm, traditional hidden Markov model (HMM) algorithm, vector quantization (VECq) algorithm, VQ, support vector machine (SVM), and deep neural network (DNN). At present, speech recognition technology is mainly realized through DNN, which can achieve up to 97% recognition rate in quiet environment.

2.2. Voice Wake up. C keyword spotting refers to the transition from a dormant or working state of not recognizing external sounds to an interactive state of accepting and recognizing external sounds, using a specific wakeword. The length of the wake up word is usually 4–6 syllables, such as "Xiao Ai Classmate" of Mi smart speaker, "Ding Dong Ding Dong" of Ding Dong speaker, and "Tmall Genie" of Tmall Genie X1 [8]. Speech wake up is the first step of speech interaction. Only by switching the speaker from other state to speech recognition state can the following speech interaction be carried out.

Voice wake is a local real-time work, the wake module is a small speech recognition engine, and wake up can be thought of as a low-resource keyword retrieval task; therefore, the corresponding acoustic model and language model are small and occupy low system resources. There are two commonly used methods of voice awakening: the HMM-based keyword/filler method and the neural network-based method [9]. Neural network-based methods include three types. The first type is acoustic model modeling using neural network HMM. The second is template matching with neural network, which uses neural network as feature extractor. The third type is based on the end-to-end solution, input speech, and output for each wake up probability, a model solution.

The HMM-based keyword/Filler scheme consists of three parts: decoding network, keyword modeling, and filler modeling. The decoder network contains the keyword and filler paths, modeling the decoder for the keyword and including any sound and noise outside the keyword in the filler.

The evaluation index of voice wake up includes four aspects: wake up rate, false alarm rate, response time, and power consumption.

(1) Wake up rate: it is also called recall rate, which means the proportion of the correct number of wake up times to the number of wake up words inputted.

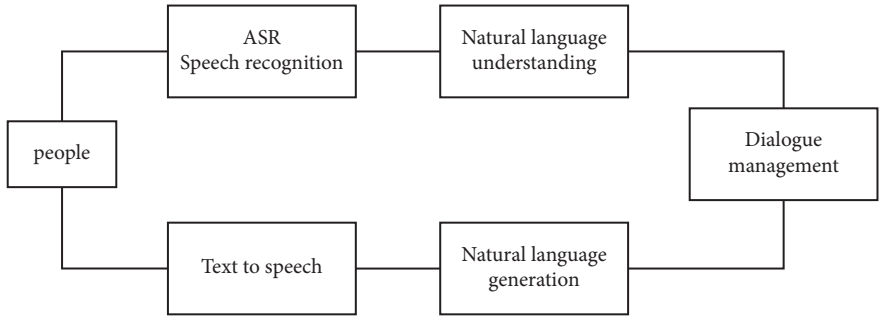

Figure 1: Voice interaction flow.

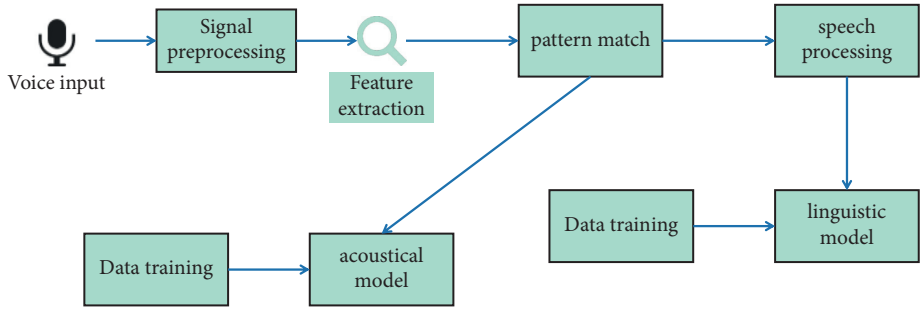

Figure 2: Principle of speech recognition.

(2) False alarm rate: it is the probability that the device is awakened when the user does not interact.

(3) Response time: it refers to the time interval between the user saying the wake word and the device giving feedback.

(4) Power consumption: it refers to waking up the power consumption of the system. The wake-up module works for a long time and should meet the low power consumption.

2.3. Voice Intelligent Home Pension System Architecture. The design of the system architecture needs to first clarify the requirements, choose the appropriate technology for planning, and use the complete tools to achieve a complete system. Intelligent home pension system is an intelligent system that integrates many technologies. It not only includes the transmission protocol of communication, wireless networking technology of terminal, and automatic control technology of equipment but also includes the terminal control software for managing equipment. The intelligent home pension system that combines voice interaction needs to follow certain design principles when adding smart speaker terminals and speech recognition interfaces. From the perspective of structure, the system consists of four parts: user terminal, cloud platform, access gateway, and home appliances.

2.3.1. User Terminal. The control mode of the intelligent home pension system or the entrance of intelligent home pension system should be multiple rather than single, so in terms of interaction, the current single mobile phone terminal control mode cannot meet the increasingly complex use scenarios and a variety of user groups. From the user's point of view, the interaction mode combining voice and mobile touch can adapt to more scenarios and user groups. Voice interaction does not require a series of operation processes, such as unlocking the phone and opening the app, so device management can be conducted more quickly and naturally. The mobile phone terminal application control mode can realize the remote control of home appliances more accurately and reliably [10]. Therefore, the control mode of voice + mobile phone is adopted in the system design.

2.3.2. Cloud Platform. The cloud platform is the core component of the system and is structurally connected to the user terminal and access gateway. In terms of functions, the cloud platform has computing, network, and storage capabilities, coordinates the interaction of control points between user terminals and intelligent home pension system gateways, and provides service API interfaces for developers and users. For the gateway, the cloud platform provides the cloud communication function to receive and process the device status sent by the intelligent home pension system gateway and stores various home environment data. For users, they can respond to requests sent by terminal app or smart speaker to realize remote control and information viewing.

2.3.3. Access Gateway. Gateway is the heart of intelligent home pension system and is not only the home appliance equipment communication Internet bridge but also home appliance equipment control hub. Gateway adds devices and

provides powerful networking functions, such as the current ZigBee WIFI gateway, a ZigBee module is set in the gateway, which acts as a coordinator to build the ZigBee network and receive network access requests from terminal devices [11].

2.3.4. Home Appliances. Terminal appliances are not only the basic part of the system but also the user's control object. Wired or wireless communication module is adopted in the home appliance controller to enable the traditional equipment to have the ability of data transmission; it can facilitate the construction of local area network for home appliances, provide a convenient entry for remote control of home appliances, and also provide conditions for linkage control of home appliances.

3. Experimental Analysis

The hardware environment used in this experiment is Intel Core(TM) i3 processor, 3.20 GHz dominant frequency, and 3G memory. Operating system is Ubuntu9.L 0. Semi-supervised algorithm tool is SVMIin. Since the toolkit is written in C++, a compilation tool is required. The compilation tool of choice is g++ 4.6.3. Command words ranging from 4 to 13 in length were selected for experimental data, including 100 positive sample samples, 100 negative sample, and 800 unlabeled samples.

SVMIin is a software package for support vector machines, which can deal with large sample and multifeature datasets. SVMIin includes standard SVM algorithm and direct push SVM based on semisupervised learning. The supervised learning algorithms supported by SVMIin include linear regularized least squares classification algorithm and improved Newton linear support vector machine. Semisupervised learning algorithms include direct push support vector machine and deterministic annealing linear semisupervised support vector machine. Since there are many voice signals in the home and most of these signals are unmarked samples, therefore, direct push support vector machine can be used to train and classify these signals, so as to get accurate recognition results. The direct push support vector machine in SVMIin is fast and simple in operation and has a good effect on the processing of a large number of multifeature unmarked samples; this method is more suitable for the recognition of post-voice control commands in intelligent home pension systems.

The version selected is SVMIin V1.0. After decompression, there are four files, namely, make-file, ssl.h, ssl. CPP, and svmlin.cpp; copy the folder to an ubuntu system with the g++ compiler and compile the make-file with the make command to get the executable svmlin. This executable can be used for training and testing datasets.

The dataset needs to be processed before training and testing. SVMIin requires that the training and testing samples be in LIBSVM format, i.e., <feature>:<value> ...<feature>:<value>; in addition, labeled category files of training samples and test samples are required. In the files, 1 means positive example, -1 means negative example, and 0 means unlabeled.

3.1. Basic Process of Experiment

(1) Select and sort out the commonly used voice phrase commands in the home, and select 100 of them as positive example training samples and 100 as positive example test samples; some characteristic nonphrase commands are selected, 100 of which are selected as negative example training samples and 100 as negative example test samples; the unlabeled sample was a random selection of phrases ranging in length from 4 to 13, and there were 800 of them. The positive example sample content is turn on the living room light, turn off the air conditioner, turn on the background light, and other command phrases, and the negative example sample content is the living room light is broken; fix the air conditioner and the bedroom is too dirty.

(2) Use the word segmentation program to perform word segmentation on the above training samples and test samples to get the sample format as LIBSVM format and mark the phrases in the training samples and test samples accordingly.

(3) Use SVMlin to train the training samples to get the classification model, and save the weight vector of the training samples in training. _examples. weights; the classification accuracy of voice command phrases was obtained by testing the test samples with test commands, and the results were saved in test_ examples.outputs.

3.2. Result Analysis. There are ten groups of experiments in this experiment. Each group has the same operation, but different data selection and algorithm selection; the first five groups of experiments selected the algorithm of direct push support vector machine; the last five groups of algorithms are the direct push support vector machine with deterministic annealing technology. Firstly, the experimental model is trained, and 1000 samples are trained to get the post-voice control command recognition model, and then, the test is carried out according to the recognition model. The test data of the five groups of experimental tests were taken from 200 test samples, and 20 of the 100 positive samples were randomly selected as positive samples in each group of experiment, 20 of the 100 negative samples were randomly selected as the negative sample test. Table 1 shows the experimental results of algorithm 1, and Table 2 shows the experimental results of algorithm 2.

According to the experimental results in Tables 1 and 2, combined with the experimental comparison diagram, it can be found that the accuracy of the direct push support vector machine algorithm is about 62.4%, while the accuracy of the direct push support vector machine with deterministic annealing technology is 67.5%, it can be seen from Figure 3 that the recognition effect of the

Table 1: Experimental results of algorithm 1.

	Correct number	Wrong number	Correct (%)
The first set of experimental results	26	14	65
The second set of results	28	12	70
The third set of results	23	17	57.5
The fourth set of experimental results	22	18	55
The fifth group of results	26	14	65

Table 2: Experimental results of algorithm 2.

	Correct number	Wrong number	Correct (%)
The first set of experimental results	28	12	70
The second set of results	28	13	70
The third set of results	26	14	64
The fourth set of experimental results	25	15	62.5
The fifth group of results	28	12	70

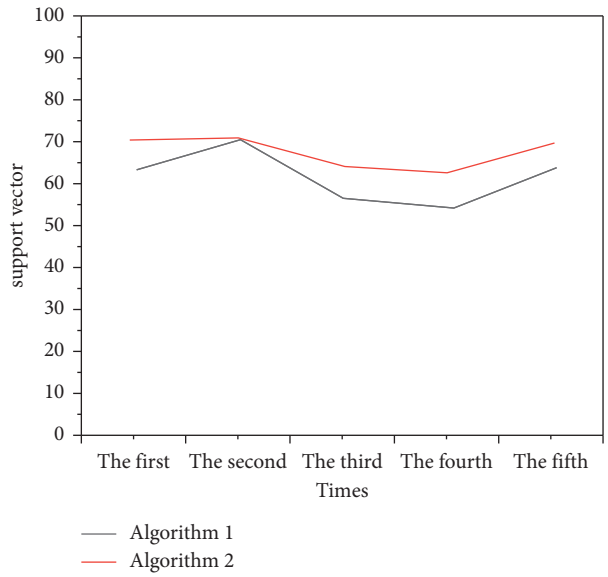

Figure 3: Experimental comparison.

direct push support vector machine algorithm with deterministic annealing technology is slightly better than that of the direct push support vector machine; however, the accuracy _ is not particularly significant, which may be related to the number of training samples and test samples and the choice of phrase commands. Through this experiment, the purpose of the design experiment is basically achieved, the semisupervised learning algorithm is used for post-speech semantic recognition of command phrases, and the accuracy is about 65%; however, due to the limited number of training samples, there may be some deficiencies in the model training so that the recognition model recognition command phrase effect did not reach the ideal state. In addition, the identification of the positive example command phrase and the negative example command phrase may not be particularly obvious, resulting in the general effect of the experiment.

4. Conclusions

This study analyzes the current development of domestic intelligent home pension system field, from the perspective of application to explain the current intelligent home in human-computer interaction deficiencies and the current application of speech recognition technology in the intelligent home pension system, the speech recognition, front-end signal processing, and communication key technologies involved in the system are studied and analyzed theoretically. Through the test of system functions on the mobile end and AIUI evaluation board, the functions of adding devices, controlling devices, and querying device status can be realized on the mobile end; the control of all functions of home appliances can be realized through the Android control panel; the voice control of rice cooker and scene-based control can be realized through the test of the smart speaker. The system realizes three control modes of mobile phone + voice + central control panel. Through the test of voice awakening, the performance of the system's voice interaction is verified, which basically satisfies the function of controlling household appliances by the smart speaker in the home environment. The system needs to be updated constantly, and the technology is also developing constantly. It is believed that, with the further development of artificial intelligence, the intelligent home will be more perfect and intelligent.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] X. Yang, T. Gao, T. Li, Y. Liu, and Y. Wang, "Research and implementation of mine risk area semantic retrieval system based on ontology," *International Journal of Advanced Pervasive and Ubiquitous Computing*, vol. 8, no. 3, pp. 37–86, 2016.

[2] B.-H. Yu, H. Wang, X.-P. Dong, and X.-Z. Zhang, "Design and implementation of a semantic gateway based on ssn ontology," *Procedia Computer Science*, vol. 183, no. 6, pp. 432–439, 2021.

[3] J. Cai and Y. Liu, "Research on English pronunciation training based on intelligent speech recognition," *International Journal of Speech Technology*, vol. 21, no. 3, pp. 633–640, 2018.

[4] H.-G. Lee, M. Kim, and H. Kim, "Implementation of a high-performance answer snippet retrieval system based on multiple ranking models for biomedical documents," *Advanced Science Letters*, vol. 23, no. 10, pp. 9486–9490, 2017.

[5] D. Tiede, A. Baraldi, M. Sudmanns, M. Belgiu, and S. Lang, "Architecture and prototypical implementation of a semantic querying system for big earth observation image bases," *European Journal of Remote Sensing*, vol. 50, no. 1, pp. 452–463, 2017.

[6] J. Zhang, "Design and implementation of intelligent tourism system based on the android platform," *Agro Food Industry Hi-Tech*, vol. 28, no. 1, pp. 2525–2529, 2017.

[7] J. Liu, W. Chang, J. Li, and J. Wang, "Design and implementation of human-computer interaction intelligent system based on speech control," *Computer-Aided Design and Applications*, vol. 17, no. S2, pp. 22–34, 2020.

[8] H. Li, "Research and implementation of art animation system based on video stream," *Revista de la Facultad de Ingenieria*, vol. 32, no. 6, pp. 153–161, 2017.

[9] Z. Zhifeng and H. Susu, "Design and implementation of search engine based on big data," *Agro Food Industry Hi-Tech*, vol. 28, no. 1, pp. 1355–1359, 2017.

[10] Y. Yu, L. Gu, P. Yin, and G. Zhao, "Research and implementation of ontology modeling and retrieval technology of mbd model," *Beijing Hangkong Hangtian Daxue Xuebao/ Journal of Beijing University of Aeronautics and Astronautics*, vol. 43, no. 2, pp. 260–269, 2017.

[11] R. Ghazal, A. K. Malik, N. Qadeer, B. Raza, and H. Alquhayz, "Intelligent role-based access control model and framework using semantic business roles in multi-domain environments," *IEEE Access*, vol. 8, no. 99, 2020.