Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 198 (2022) 505–511

Procedia Computer Science

www.elsevier.com/locate/procedia

International Workshop on Internet of Smart Things (IST 2021)
November 1-4, 2021, Leuven, Belgium

Research and Design of Intelligent Speech Equipment in Smart English Language Lab Based on Internet of Things Technology

Xinxia Cheng*, Yabin Fan

Shijiazhuang University, No. 288 Zhufeng Street, Shijiazhuang 050035, China

Abstract

With the wide popularization of Internet of Things technology, the design and implementation of intelligent speech equipment has attracted more and more researchers' attention. Speech recognition is one of the core technologies to control intelligent mechanical equipment. In this paper, the English speech recognition function is realized by the Hidden Markov model, which strengthens the processing ability of speech signal, and the intelligent speech equipment is applied to the smart language lab. The test results show that the designed system has comparatively accurate recognition ability to English speech and relatively strong control ability to intelligent device. Users are liberated from the traditional interaction mode and can complete the operation of intelligent language lab by non-contact way of voice signal in a non-fixed place without interrupting the user's current behavior.

Keywords: English speech recognition; Hidden Markov model（HMM); Internet of Things (IoT); Intelligent language lab; speech equipment

1. Introduction

Internet of Things technology is a comprehensive application technology integrating wireless sensor technology, computer technology, communication technology and embedded technology. In 2008, IBM proposed the three elements of "Smart Planet" : Internet of Things, interconnection and intellectualization. The Internet of Things and the Internet began to integrate comprehensively, which became a milestone event in the development of the Internet of Things. With the commercialization of 5G technology in 2019, the Internet of Things will definitely get greater

* Corresponding author. Tel.: +1-333-137-8790; Tel.: +1-338-321-5790
E-mail address: 36321898@qq.com; fansjzxy@qq.com

10.1016/j.procs.2021.12.277

development, especially in the fields of smart home, smart transportation and smart classroom.

Smart classroom uses Internet, Automatic Control, Multi-media, Sensors and other technologies to integrate teaching-related hardware facilities through the classroom building and provide a safe, convenient, comfortable and energy-saving learning environment. The Internet of things technology provides the functions such as electric control, audio and video equipment control etc., thus achieving internal and external network communication and remote control, but also being convenient for teachers to optimize teaching methods and to improve the teaching quality.

Apple's "ACOT", Stanford University's iroom in America are the representative of the research on smart classroom in foreign countries. The smart classroom of Mc Gill University in Canada uses an integrated control panel to realize the control of electrical equipment in the classroom [1].

The most typical representative of smart classroom research in China is the "Smart classroom" of Shanghai Jiao Tong University, which integrates teaching, attendance, environmental testing and remote video monitoring, and is the benchmark of smart classroom in domestic universities [2].

Intelligent devices are widely used. As one of the key technologies of intelligent control, the ability of speech recognition control directly determines the degree of intelligence of the device [3]. However, there are restrictions on the operation location and interaction mode for users. This paper proposes and designs an embedded English speech recognition control system, which uses Hidden Markov model to recognize English speech and selects WTV180 chip to process speech signal. The system has a high degree of English speech recognition and strong control ability. It can use speech signals to complete the operation of some intelligent devices in the intelligent language lab in a non-contact manner and in an unfixed place without interrupting the user's current behavior.

2. Structure and composition of intelligent language lab

The hardware construction of smart language lab is mainly the circuit design of ZigBee wireless communication node. The control system is generally composed of four parts: gateway, functional submodule, communication network inside the classroom and external communication of the control system.

Firstly, the data exchange between the coordinator, monitor and relay is realized by ZigBee protocol; then the gateway realizes the data transmission through the connection of serial communication line with coordinator; finally, the data is preprocessed and displayed in the control system on the gateway. Gateway is the core unit of the control system of smart language lab. The overall system diagram is shown in Fig. 1:

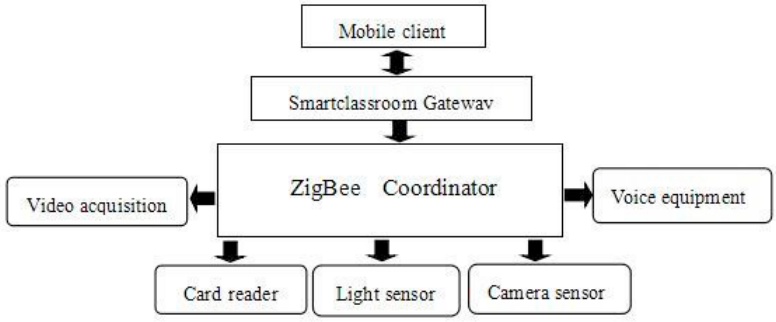

Fig. 1. Smart classroom overall system block diagram

As the core part of the whole system, gateway control system mainly completes the remote control and management of the classroom, to achieve the control of the same LAN Intranet electrical equipment, and at the same time to collect classroom environmental parameters and switch modes if necessary.

For energy saving and security, it is planned to adopt low power consumption and high performance embedded chip as the gateway development platform. Data are collected and monitored by the light sensor, temperature and humidity sensor, human body infrared sensor in the ZigBee network connected with the gateway, so as to lay the foundation for automatic control equipment. The embedded gateway system is designed with four functional modules: registration and login, environmental monitoring, electrical control and automatic condition control.

2.1 Communication part

With the rapid development of WSN, ZigBee technology has been applied in many fields such as smart home, smart transportation, and smart agriculture. Its advantages are as follows: low power consumption, self-networking, safe and reliable connection of a maximum of 255 nodes to the network at the same time. CC2530 of TI, which will be used in this paper, is the current mainstream ZigBee development board.

A ZigBee wireless sensor network generally consists of a coordinator, router and end-device, and can be AD hoc networking if the ZigBee protocol stack is satisfied.

Speech recognition control the SPK microphone collects the voice and sends the signal to the WTV180 chip, in which the speech signal is converted. The output signal is sent to the STM32 embedded system using the DATA line, which controls the intelligent device according to the command of the speech signal.

2.2 Hardware of the system

ZigBee data acquisition terminal uses CC2530F256 as the chip of application development. Its main characteristics are: high anti-interference ability and high sensitivity, low power consumption, small size, easy to package, high programmable output power. In addition, it can support CSMA/CA transmission mechanism and channel sharing, visual signal strength and connection quality, and TI official IAR development technology [4].

For gateway hardware selection, this paper adopts Tiny210 designed by Guangzhou Friendly Arm Company as the embedded development platform. In the development board, there is an SD card adapter interface. In the smart classroom management system, the SD card can be used to write system image files such as Uboot and kernel for the embedded gateway, thus forming the final embedded gateway control system.

For hardware components of the speech recognition system, the core processor is STM32 processor with Cortex-M3 as the core. This chip adopts a tail-chaining interrupt structure, processing speed of the data very high. At the same time, its integrated thumb-2 instruction set greatly improves the instruction execution efficiency and the operating performance of the chip, and is widely used in the industrial control industry.In this paper, for the design of intelligent equipment speech control in electrical module, WTV180 chip is selected for processing the English speech signal. Its technical advantages are mainly reflected in the following points. First, under the condition that the speech signal frequency is 6KHZ, the maximum recognition length can reach 340s, and the number of internal types is high and suitable for a variety of environments; second, the chip integrates DAG and optimizes sound quality algorithm and PSG speech synthesizer to achieve high sound quality [5]. The hardware structure of the embedded English speech recognition control system is shown in Fig. 2.

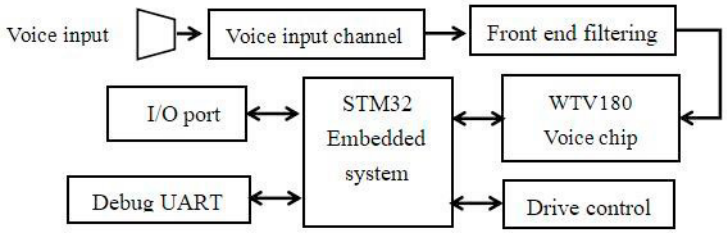

Fig. 2. Speech recognition hardware structure diagram

2.3 Software of the system

The control platform of smart classroom system uses VMware Workstation. Its powerful features make it easy for users to run other operating systems on their current system without having to add new hardware and install dual or multiple systems, which greatly reduces the cost of development of time and money.

The embedded system is basically carried out under Linux system. Thus, Qt under Ubuntu system is used to develop the graphical interface program of gateway control system of smart classroom, and the development environment of Qt 4.7.0 or higher version is built. CortexTM-A8 gateway is mainly used for programming applications of smart classroom. Users can operate the smart classroom module through the touch screen.

The software part of the speech recognition system is based on the modular program for the training and vocabulary storage of intelligent mechanical equipment. During the training, each single control instruction is

repeated 10 times; after the training is completed, the input interface function is called to request the speech recognition module of the system to start recognition, and then the output interface function is called to submit the recognition results. Spoken language corpora are databases containing audio and transcribed files. In the field of speech technology, spoken language corpora are widely used to create acoustic models. In this paper, Libri Speech corpus is used as the speech model data of speech recognition system.

3. Methods of speech recognition system

3.1 The development of speech recognition system

In the 21st century, with the popularity of electronic products, embedded speech processing technology has developed rapidly. However, there are still many difficulties to be solved. One is noise environment: due to many noise sources, the recognition rate of the system is obviously decreased in the actual environment. Another one is continuous speech. English continuous speech has high degree of continuity and serious coarticulation, but the recognition rate decreases under continuous speech condition. At present, most commercial software needs tedious "training", so that the speech recognition system can be targeted to the user's speech recognition. However, once the system or user is replaced, it needs to be "trained" again, which brings a lot of inconvenience to the user.

3.2 Mathematical model of speech recognition algorithm

The mathematical model corresponding to the speech recognition algorithm consists of three parts: radiation, excitation and vocal channel. The concrete structure is shown in Fig. 3.

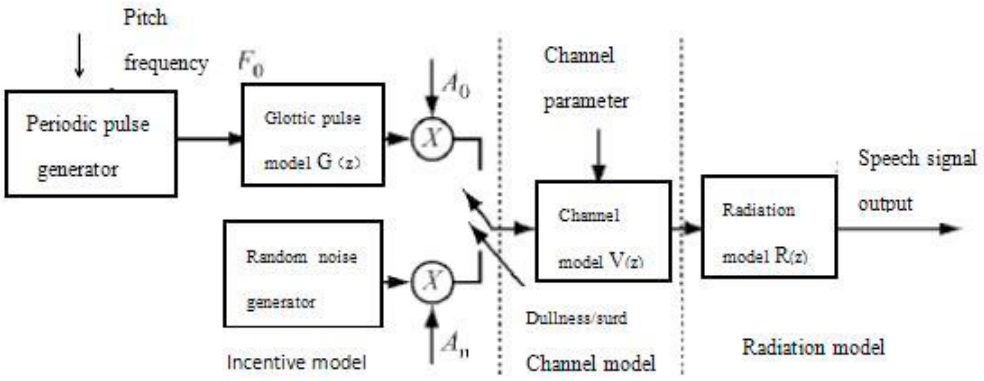

Fig. 3. The mathematical model structure of speech recognition algorithm

G (z) of the excitation part is the glottic pulse signal; V (z) of the sound channel is the transmission function of the sound channel; R (z) of radiation component is radiation resistance; A is the tuning coefficient, used to adjust the energy or amplitude of the function.The core function of this model is to determine the output function H (z) required for speech signal conversion processing, as shown in Equation (1).

$$H(z) = A \cdot G(z)V(z)R(z) \tag{1}$$

3.3 Feature Extraction

The accuracy of the system's speech recognition results depends on the algorithm's ability to accurately extract speech features. In this paper, MFCC method is selected for frequency domain feature extraction of speech signal. The extraction process is shown in Fig. 4.

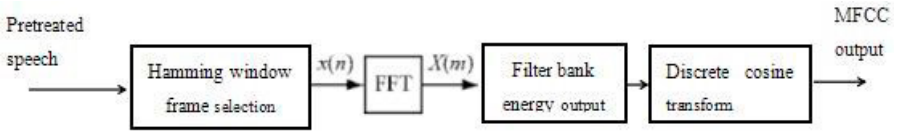

Fig. 4. Frequency language feature extraction process based on MFCC method

3.4 HMM speech recognition algorithm

Among the existing speech recognition algorithms, HMM algorithm is widely recognized for its excellent ability of predicting process state [6]. Therefore, this paper adopts this algorithm to realize the speech recognition function of the system.

Suppose S_1, S_2... , S_n is a set of state quantities in the model, and the unique model state corresponding to the time node n is x (n). When n = 0, the starting point probability vector π is expressed in Equation (2).

$$\pi_i = P\{x_0 = S_i\}(i = 1, 2, \cdots, n) \tag{2}$$

The state value corresponding to each subsequent time node is only related to the state value x_{n-1} at the previous time point; thus the transition probability matrix A = $\{a_{i,j}\}$ can be obtained, which can be expressed as Equation (3).

$$a_{i,j} = P\{x_n = S_j \mid x_{n-1} = S_i\} \tag{3}$$

Except for the starting point, the states of all remaining time points are hidden, so only the random monitoring vector Q_n corresponding to a single time point R_q can be calculated, and its relationship with X_n (P_n) is shown in Equation (4).

$$P_{x_n} = S_i\{Q_n\} = P\{Q_n \mid S_i\} \tag{4}$$

The HMM algorithm is used to obtain Qn when extracting speech signal features. In this paper, Q_n is fitted based on mixed Gaussian distribution, and Equation (5) is obtained.

$$P\{x_n = S_i, Q_n = Y\} = \sum_{m=1}^{M} C_m N(\mu_m, \sigma_m^2) \tag{5}$$

Where, m is the fitting order value; Y is the state value of Q_n; C_m is the weighting coefficient.

The first step of HMM algorithm should be to create English phonetic lexicon. A word list containing V English words is given, and the model coefficient λ_V is assigned to each word, then equation (6) is given.

$$\lambda_v = (A_v, B_v, \pi_v) \tag{6}$$

In speech recognition, at first, the monitoring vector Q = $\{Q_1, Q_2 \ldots Q_T\}$ is extracted, and then the partition probability P (Q | λ) of the mode system λ_V is calculated. Where, V∈[1, V], and finally, the word V^* with the maximum likelihood probability is obtained, which is the corresponding value of the speech recognition result, as shown in Equation (7).

$$v^* = \arg\max_{1 \le v \le V} \{P(Q \mid \lambda)\} \tag{7}$$

4. System testing

In order to verify the actual performance of the English speech recognition system designed in this paper, according to the function of the system, the experiment is for English speech recognition. The test process of speech recognition function is as following: 5 English direction instructions (run, left, right, backwards, stop) are selected and repeated 20 times each to verify the accuracy of English speech recognition. The test results are shown in Fig. 5.

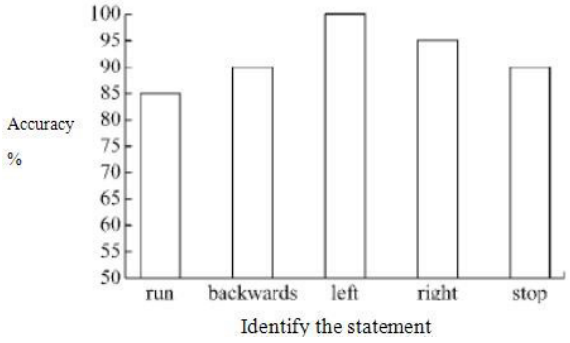

Fig. 5. Speech recognition function test results

As can be seen from Fig. 5, the system's recognition accuracy of each English instruction is higher than 85%, and most of them reach over 90%. It can be seen that the system designed in this paper has high recognition accuracy.

Intelligent mechanical equipment is a set of intelligent audio and video acquisition equipment, using a fixed base and a rotating plate and the base is provided with a servo motor. The output shaft of the servo motor is coaxially fixed with a driving gear, and a lifting shaft that can slide vertically is set in the rotating shaft. The upper end of the lifting shaft is fixed with a rotating plate. Audio recorders are fixed on both sides of the display screen embedded in the front of the rotating plate, and infrared cameras are arranged on the upper edge. The cameras can carry out the circumferential rotation and height rise and fall. The left side of the camera is equipped with a sound sensor, which receives external voice commands. The control and adjustment are simple and convenient, and the process is smooth to achieve higher audio and video recording quality.

The control function of intelligent mechanical equipment is tested based on the consistency between the path of intelligent camera and the instruction. The test results are shown in Fig. 6.

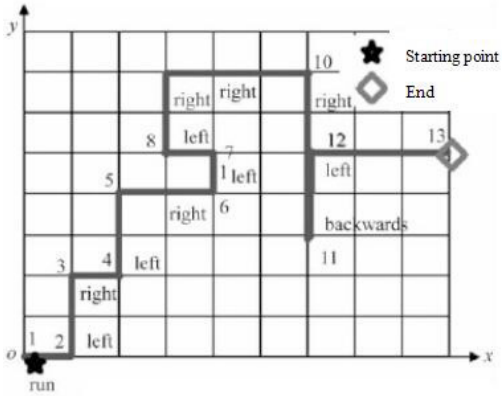

Fig. 6. System control function test results

5. Conclusion

Speech recognition control is an important automation technology and has become one of the core technologies of intelligent equipment research and development. Combined with Internet of Things and artificial intelligence technology, an English language lab of intelligent speech control for English learners is built, which is convenient to control and operate various equipment. Teachers and students can adapt to a variety of learning styles, including remote teaching. The speech recognition control system proposed and designed based on Internet of Things in this paper realizes the recognition function of English speech based on HMM algorithm, and uses WTV180 chip to process English speech signal, which greatly improves the speech recognition and processing ability of intelligent devices, and has certain reference value for the research of artificial intelligence related technologies.

The embedded gateway control system completed in this paper is still rudimentary, and its functions are not perfect. It needs to be further improved. In terms of technology implementation, more advanced, low-power, long-range wireless networking technology, such as LoRa, can be applied to realize smart campus.

References

[1] Sun Ming-li, Bao Jian, Zhang Shuo. (2010) "1-Wire Bus Technology and Realization of Temperature Measurement of DS18B20." *Journal of Atmospheric and Environmental Optics*, 5(4):322-326.

[2] Lei, Zhang. (2017) "Exploration and Thinking on the Construction of Smart Classroom under the Background of 'Internet +'." *Network Security Technology & Application* (6):118-131.

[3] Lijun, Deng and Tao, Wang. (2020) "Research on English speech Automatic recognition System Based on Threshold." *Microcomputer Applications* 36 (38): 48-50.

[4] Jing, Ma. (2017) "Design and Implementation of Smart Home System Based on ZigBee Wireless Network." *Technology Innovation Application* 188 (4):40-41.

[5] Wei, Liang. (2020) "Development and Application Research of English Phonetic Alphabet Assisted Learning Platform Based on Speech Recognition Technology." *Computing Technology and Automation* 39 (2): 155-159'

[6] Deyan, Fan. (2016) "Design of ZigBee Intelligent Environment Monitoring System Based on CC2530." Dissertation of Shenzhen University.