

# Research on the method of home remote-control based on embedded speech recognition

Chaoqun Chen<sup>1,a</sup> and Ping Qian<sup>2,b</sup>

<sup>1</sup>School of Mechanical Engineering, Shanghai Institute of Technology, Shanghai 201418,China;

<sup>2</sup>School of Engineering Innovation, Shanghai Institute of Technology, Shanghai 201418,China.

<sup>a</sup>273552201@qq.com, <sup>b</sup>qping@sit.edu.cn.

**Keywords:** embedded system; smart home; speech recognition technology

**Abstract.** In order to control the current smart home products and improve the convenience of the use of special people, this paper designs a smart home control method for speech recognition to control smart home applications. The method is based on the method of voice sensor to receive signals to be pretreatment, then in the embedded system deal with the problem of evaluation, change signals into digit signal and learn the problem of system sampling. At the end of the paper, the comparison with the actual speech signal is carried out to verify the validity and reliability of the design method.

## 1. Introduction

With the continuous development of computer technology, people's demand for intelligent products have increased. People have put forward higher requirement for communication equipment, household appliances and security equipment, control management and information exchange in the family <sup>[1]</sup>. The traditional home remote control relies on manual triggering or wearing gloves to control the perception of home equipment. For some elderly or disabled people, the operation is relatively complex. Based on the speech recognition technology, this paper studies a new method of home remote control. It is convenient to use to direct voice control by remote control object, for the elderly and people with physical disabilities to provide great convenience.

## 2. Embedded Speech Recognition Technology

Embedded speech recognition system is implemented by using software or hardware on board level or chip level. Compared with speech recognition system on PC, the embedded speech recognition system is limited, but its active power consumption is low, small size, flexible installation, low cost and high reliability.

Speech recognition technology is to get the speech signal through the speech acquisition device, and then get the parameters model which can reflect the characteristics of the speech signal. The parameters model is transferred to the pattern matching module device. The reference [2] from the signal processing point of view, any speech recognition system can express to figure 1. The main process is that speech recognition signal feature is extracted and pattern is matched.

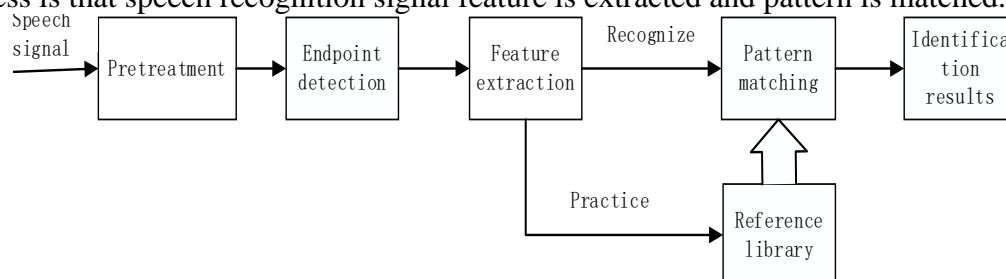


Fig.1 Basic principle of speech recognition

The speech recognition process mainly includes the basic structure: (1) the speech signal is pretreatment by sub frame, add window, pre emphasis; (2) the endpoint detection filter out noise in the speech input, then we get real effective starting point and end point by the endpoint detection; (3)

the feature extraction is to obtain the input speech signal. (4) In the training phase, the user input speech signal is processed by the speech recognition system to get the vector parameters from reference library; (5) recognition is the highest similar degree of the characteristics of the parameters as a result, and output the highest degree of the model as a result of the recognition.

### 3. Design of Embedded Speech System

#### 3.1 Hardware Design of Speech Recognition System

The hardware of embedded speech recognition system includes: power management circuit, MCU circuit, voice recognition circuit, wireless communication circuit, LCD touch display circuit, JTAG debug circuit and control circuit, as shown in Figure 2.

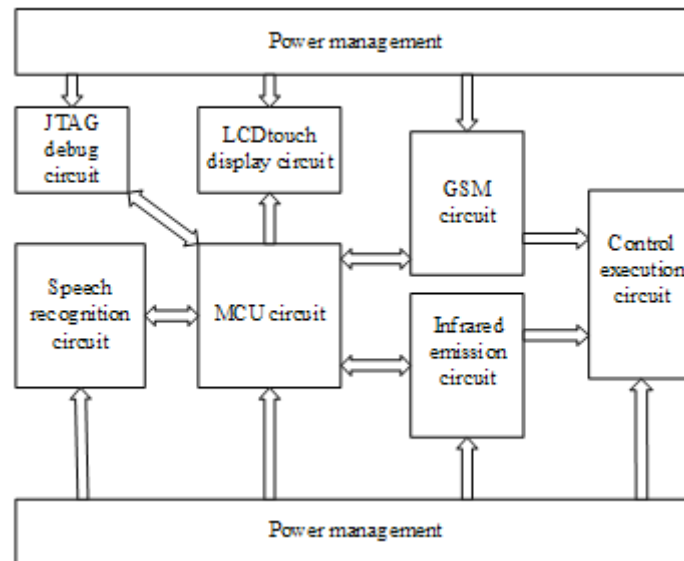


Fig.2 Block diagram of hardware system

The speech recognition module is composed of a speech signal acquisition device MIC, a speech processing chip, and a voice output device SPK. The speech signal acquisition device transmits the collected speech information to the speech processing chip, and the main controller judges the meaning of the speech signal according to the data obtained by the speech chip processing and executes the corresponding control actions.

The external input voltage of the power supply circuit is 5V, the input voltage is required for the whole system. The debug circuit is used to test the initial system board debugging, verify the correctness of the hardware circuit, including serial port circuit. The speech recognition module is the core of the whole system, receive instructions from the microcontroller, processing and result output. The department includes AD or DA converter, RAM, microphone interface, sound output interface, etc. Home device control circuit is a part of the implementation of voice commands, receive instructions from micro controller, the implementation of the switch operation of home devices, wireless data transmission module in the whole system also has an important position, it will be transmitted to the control circuit, and the state feedback to the main controller. LCD touch screen is a system of human-computer interaction interface, can control the system, display current system state.

#### 3.2 Software Design of Speech Recognition System

Speech recognition system using SPCE061A chip of the company named Ling Yang, in the IDE environment to design a speech recognition system.

The embedded speech recognition and control system is as follows:

- 1) System of hardware and software initialization, create a module task and assign priority to the system to enter the main program, start work.
- 2) Waiting for the user to voice or press button to start the acquisition of control commands the voice information

- 3) When the system is collected to the speech information, the speech recognition is carried out in the interrupt processing program. The recognition of the acquired speech signal is successful. If there is no recognition of the success or the error of the process, then the speech signal is collected.
- 4) If the success of the speech recognition, the recognition results are matched to determine whether the identification of the list has been added to the command.
- 5) If the command matches successfully, the control command is transmitted to the execution terminal through the wireless communication network, and then the control action of the corresponding home device is controlled.

## 4. Speech Feature Parameters and Speech Recognition Algorithm

### 4.1 Speech Feature Parameters

In speech recognition system, it is very important to detect the initial position of the speech signal and extract the feature parameters. Feature parameters are extracted from the feature parameters, so the feature parameters are guaranteed by high recognition speed and accuracy, which is usually characterized by the following three features: accurate response and static feature and dynamic feature of voice signal.

### 4.2 Speech Recognition Algorithm

Speech recognition algorithms usually have dynamic time warping method (DTW)、Hidden Markov Model(HMM)and Artificial Neural Network (ANN),ect.DTWapply for local optimization problems, simple and convenient<sup>[3]</sup>; HMM is based on the probability model of process statistics, is an ideal speech recognition algorithm<sup>[4]</sup>;ANN is an adaptive nonlinear system, the identification ability is good<sup>[5]</sup>.

#### 4.2.1 Assessment Problem

Reference [6] according to the description of HMM, the Back method is used to simplify the calculation and the probability as

$$P(O|\lambda) = \sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), 1 \leq t \leq T + 1 \quad (1)$$

Which N represents the state of the model, T is the length of the state sequence,  $O_t$  is the t output symbol,  $a_{ij}$  is the state transfer probability distribution,  $\alpha$ ,  $\beta$  is respective as back and front method and  $i, j$  is an integer variable.

#### 4.2.2 Decoding Problem

The classical method of optimal state sequence search is Viterbi method, Definition  $\delta_t(i)$  as time t along a path  $q_1, q_2, \dots, q_t$ , and  $q_t = \theta_i$  product  $O = O_1, O_2, \dots, O_T$  maximum probability, get

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1, q_2, \dots, q_t, q_t = \theta_i, O_1, O_2, \dots, O_t | \lambda) \quad (2)$$

so, the process of seeking the best state sequence is

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N$$

Initialization:

$$\varphi_1(i) = 0, \quad 1 \leq i \leq N$$

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T, 1 \leq j \leq N$$

Recursive:

$$\varphi_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T, 1 \leq j \leq N$$

$$P = \max_{1 \leq i \leq N} [\delta_T(i)]$$

Result:

$$q_T = \arg \max_{1 \leq i \leq N} [\delta_T(i)]$$

State sequence:  $q_i = \varphi_{t+1}(q_{t+1}), \quad t = T-1, T-2, \dots, 1$

#### 4.2.2 Learning Problems

Learning problem is usually adjusted model  $\lambda = \{A, B, \pi\}$ , make the probability of the output of the observation sequence  $P(O|\lambda)$  be maximum, use classical methods for using parameter revaluation Baum-Welch. Define  $\xi_i(i, j)$  for a given model  $\lambda$  and training sequence  $O$  at time  $t$ , Markov sequence at  $\theta_i$ , when the time on  $t+1$ , the probability of  $\theta_j$  is

$$\xi_i(i, j) = P(O, q_t = \theta_i, q_{t+1} = \theta_j | \lambda) \quad (3)$$

And we can deduction

$$\xi_i(i, j) = [a_i(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)] / P(O | \lambda) \quad (4)$$

so, Markov sequence is at time  $t$  probability of  $\theta_i$  state

$$\xi_i(i, j) = P(O, q_t = \theta_i | \lambda) = \sum_{j=1}^N \xi_i(i, j) = a_i(i) \beta_i(i) / P(O | \lambda) \quad (5)$$

$\sum_{t=1}^{T-1} \xi_i(i)$  express  $\theta_i$  expected value of the number of transfers, but  $\sum_{t=1}^{T-1} \xi_i(i, j)$  express expected value from  $\theta_i$  to  $\theta_j$  times. Deduction revaluation formula of Baum-Welch algorithm is

$$\begin{aligned} \bar{\pi}_i &= \xi_i(i) \\ \bar{a}_{ij} &= \sum_{t=1}^{T-1} \xi_i(i, j) / \sum_{t=1}^{T-1} \xi_i(i) \\ \bar{b}_{jk} &= \sum_{t=1, O_t=v_k}^{T-1} \xi_i(j) / \sum_{t=1}^{T-1} \xi_i(j) \end{aligned} \quad (6)$$

HMM of solving process at parameter  $\lambda = (\pi, A, B)$  is: Known initial model  $\lambda = (\pi, A, B)$  and observation sequence  $O$ , a new set of parameters is obtained by the above - mentioned revaluation formula.  $\bar{\lambda} = (\bar{\pi}, \bar{A}, \bar{B})$ , we can prove  $P(O | \bar{\lambda}) > P(O | \lambda)$ , Show that the formula  $\bar{\lambda}$  is calculated higher accuracy than  $\lambda$  formula. Repeat the calculation process, and gradually improve the parameters until  $P(O | \bar{\lambda})$  convergence, then  $\bar{\lambda}$  is the optimization model.

#### 5. System Testing and Analysis

Experiments in accordance with the design of the system as well as the speech feature recognition algorithm to write programs, using J-Link to download the program to download the program to debug, the system on the power, the LCD screen display and other information, enter the system after the main interface into the voice control system. First of all, we carry out the pronunciation learning and control the speech broadcast, carry on the speech recognition training, and save in the reference model database; secondly, carry on the speech recognition training: the speech signal is the digit 0~9、open、close、+ (plus)、- (minus), in quiet and noisy environment from the 1m, 2m, 3m in the test 20 times, voice broadcast recognition accuracy results are shown in table 1.

From table 1, we can see that the speech recognition results are good, and the distance has little effect on the environment. Overall in the home environment can meet the user's direct voice control by remote control object.

Table 1 Test data of speech recognition rate

identification signal	1m		2m		3m	
	quiet	noisy	quiet	noisy	quiet	noisy
0	95%	85%	90%	75%	90%	60%
1	95%	85%	90%	70%	85%	60%
2	95%	80%	85%	75%	85%	65%
3	90%	85%	85%	75%	85%	65%
4	90%	85%	90%	75%	85%	70%
5	95%	85%	90%	75%	90%	65%
6	95%	80%	90%	80%	85%	60%
7	95%	85%	90%	75%	85%	65%
8	95%	85%	85%	80%	85%	60%
9	95%	80%	90%	75%	85%	65%
open	90%	80%	90%	80%	85%	65%
close	90%	85%	90%	75%	90%	60%
+	95%	85%	90%	80%	90%	60%
-	95%	85%	90%	70%	90%	60%

## 6. Summary

In this paper, the embedded speech recognition method is characterized by good human-computer communication, and it can be easily and conveniently. After the system has been trained to form a reference model, the correct rate of speech recognition is tested in a quiet and noisy environment, and the recognition of the remote control of the home is met.

## References

- [1]Zhu Min-ling,Li Ning.State of Art and Trend of Smart Home in China. TV Technology.Vol.39(2015)No.04,p.82-85
- [2] Zhao Jun Feng,Zhu Ye Ping. Embedded Speech Recognition Based on Multiclass Support Vector Machine. Key EngineeringMaterials. Vol.1104(2011) No. 467, p.1905-1910
- [3]LI Dan, MING Yong, Lu Han-yu.Research of Speech Recognition for Detection of Program Behavior Based on CDHMM. Journal of Wuhan University of Technology. Vol. 34(2012) No. 9, p. 144-148
- [4]Pang Z,Tu S,Su D,et al. Discriminative training of GMM-HMM acoustic model by RPCL learning. Front Electr Electron Eng China. Vol. 6(2011) No. 2, p. 283-290
- [5] Yu D, Deng L, Seide F, The deep tensor neural network with applications to large vocabulary speech recognition.IEEE Trans Audio Speech Lang Process. Vol. 21(2013) No. 2, p.388-396
- [6]LIU Ke. Practical Markov decision process. Tsinghua University press, 2004, p. 110-122.