The Realization of a Speech-control Unit Used for Intelligent Household Appliance

Jing-Hong YAO, Jia-Wei ZHOU

Zhejiang Wanli University, Ningbo, Zhejiang, 315100, China

yjh@zwu.edu.cn

Keywords: Speech-control Unit, Isolated Word Speech Recognition, Hidden Markov Model (HMM), Viterbi Algorithm, Intelligent Household Appliance.

Abstract. This paper introduces a speech-control unit to control the switch of household appliance. The isolated word speech recognition system is designed by using the discrete Hidden Markov model (HMM) to train and recognize the model. According to the specific target of the speech-control unit, we introduce the two methods to realize the speech-control unit, and discuss some specific problems during its realization in C language. And in KEIL uVision5 software development environment, simulate and download the program to realize the intelligent household appliance system.

Introduction

Speech recognition is a cross subject. In the recent twenty years, the speech recognition technology has made remarkable progress, and began to move from the laboratory to the market. It is expected that speech recognition technology will enter the industry, home appliances, communications, automotive electronics, medical, family services, consumer electronics and other fields in the next 10 years. Many experts believe that speech recognition technology is one of the ten important technologies in the field of information technology in the field of information technology. Speech recognition technology involves in the field of signal processing, pattern recognition, probability theory and information theory, sound mechanism and auditory mechanism, artificial intelligence, etc.

To communicate with the machine, let the machine understand what you said, this is what people have long dreamed of. Speech recognition technology is to change voice signal into a corresponding text or command by machine identifying and understanding. Speech recognition technology mainly includes feature extraction, pattern matching and model training technology.

At present, there are more and more types of household appliances. It will obviously bring us more convenience if there is a voice switch to control them. The research of speech recognition, especially on isolated word has reached a considerable height. The development of such a speech module has become possible. In fact such a speech module is an isolated word speech recognition system [1].

This paper, first, introduces the design principle of the isolated word speech recognition system. Then it introduces two design schemes according to specific objective of the speech module. And the specific problems encountered in the process of the speech module are...
discussed. At last, in KEIL uVision5 software development environment, simulate and download the program to realize the intelligent household appliance system.

The Design of Isolated Word Speech Recognition System

The design of isolated word speech recognition system has a variety of more mature methods. In this paper, we use the discrete Hidden Markov model (HMM) to train and recognize the model [2]. The basic design steps are as follows.

Training Program

End-point Detection and Feature Extraction

For isolated word speech recognition system, end-point detection is very important. This paper uses short-time average amplitude and short-time zero crossing rate to detect the end-point. Adding window frames to the speech training data, the frame length is 160(8kHz sampling frequency and 20ms data). The window is a Hamming window. Calculate the average amplitude and zero crossing rate of each frame. Their threshold is 1/20 of the maximum average amplitude and 1/15 of the maximum zero rate. The voice frame will be sentenced to silent and removed if its value is lower than the two values. LPC cepstral coefficients of each frame will be calculated. The feature extraction is the calculation of LPC cepstral coefficients. That is using Durbin recursive algorithm to compute the LPC coefficient and cepstral coefficients is obtained from PLC coefficient.

Codebook Design

Firstly, the initial codebook of the LPC coefficients of all above mentioned are determined by the splitting method. The LBG algorithm is used to obtain the optimal codebook based on Euclidean distance. The procedure for determining the initial codebook by the splitting method is:

The first step: calculate the core of the space of LPC cepstral coefficient as the first initial code word.

The second step: perturbation quantity \( \varepsilon \) divides the initial codeword into two. The two new code words stand for two new space. \( \varepsilon \) is chosen between 0.01 and 0.05.

The third step: calculate the Euclidean distance between each frame's LPC cepstral coefficients and the two new code words. Determine the frame's vector to belong to which one of the two new spaces.

The forth step: recalculate the core in each space.

The fifth step: repeat the second step, each core is divided into two, so there are 4 cores of 4 spaces(code word).

Continue to repeat as above, all PLC cepstral coefficients are divided into more regions(cell). Generally, it is divided into 256 cells, each core is the average value of LPC cepstral coefficient. That is initial codebook.

LBG algorithm is a recursive algorithm, which is optimization based on the initial codebook. Its steps are:

Calculation the Euclidean distance between all LPC cepstral coefficients and initial codebook sized M. According to the average method, calculate the new core of each cell. Recalculate the distance between LPC cepstral coefficients and these new core. According to
nearest neighbor rule these coefficients are divided into M cell again. And so on, until convergence.

**Calculate the HMM of Speech-unit**

Design a HMM for each speech unit, the different speech units have different HMM parameters. The HMM parameter has two groups of data: the probability matrix of the HMM state's transition and the probability matrix of each observation symbol (i.e. the each element of the codebook) in each state. If the state number of each voice unit HMM is 6, the code size is 256, then the first probability matrix of each HMM has 36 elements, and the second probability matrix has 6*256=1536 elements. The piecewise K means algorithm [3] is used to obtain the two probability matrices. First to estimate their initial values. Because the speech structure of the HMM can be thought of two transition. That is a state observation symbols or transition to the state or transition to next state. So observation sequence of symbols of each speech unit training sample (i.e. the training data based on codebook is designed into the vector quantization) is divided into same N segments (N is the number of HMM States). Each segment corresponds to a HMM state, and each observation vector is signed the state number. An initial set of HMM parameters are estimated. Specific formula for:

The probability distribution of the observed symbol for the first j state is estimated to be:

$$b_j(k) = \frac{\text{the number of the Kth codeword in the state j}}{\text{total number of the observed vector in the state j}}$$  \hspace{1cm} (1)

If the duration of the i state is $\tau$, the estimated valuation of transition probability of i state is:

$$a_{ii} = \frac{\tau - 1}{\tau}$$  \hspace{1cm} (2)

The valuation of the transition probability of the i state to the next state $j=i+1$:

$$a_{ij} = 1 - a_{ii}$$  \hspace{1cm} (3)

Other state transition probabilities are zero.

According to this set of parameters, the corresponding optimal state sequence can be calculated by using the Viterbi algorithm. Then the same method is used to estimate a set of new HMM parameters according to the optimal state sequence. And the new optimal state sequence is calculated by the Viterbi algorithm based on the set of HMM parameters. Thus many times repeated until convergence.

**Recognition Program**

In the recognition process, the first step is the same with the training program. It is also first to extract the feature of recognition data, and calculation the LPC cepstral coefficient. The second step is getting the vector data to form the observation data according to the designed code. The HMM of the each speech unit is then calculated to generate the probability of the observed data, and determine them to belong which voice unit. The Viterbi algorithm is used to calculate the probability of the observation data generated by the HMM.
Two Design Scheme and Detail Issues in Programming

The first scheme treats the commands of “open computer”, “off computer”, “open refrigerator”, “off refrigerator” as recognized voice units. “open computer” is a HMM model, “off computer” is another HMM model, i.e. The advantage of this scheme is simple. The lack is too large sound similarity. It will cause error. The second scheme is separate the “open”, “off”, “computer”, and “refrigerator” and so on. The advantage of this method is that the sound similarity is relatively small, which is beneficial to the identification. But this will require the person to have a pause when he command the appliances in the “on”, or “off” in order that the recognition program can identify the boundary point between two voice units.

Due to the speaker in the commands may have an almost imperceptible pause. So, no matter what kinds of methods are related to the problem of continuous frames for zero in the voice data stream. We have encountered the problem in programming. This will calculate the error cepstral coefficient with Durbin recursive algorithm.

The solution is to get rid of the frame data. Because it is just to remove the pause of the voice, it does not affect the recognition.

Another problem in the realization of voice recognition system is that the number of code words in some cell cavities may be zero when being used splitting method for initial code word. In this step, the number of code words in the cell must be divided to get the cell type. Zero is divided. The solution is that the cell type cannot be obtained if the number of code words in a cell is zero. Since the number of code words is zero, there is no need to get the cell type.

In addition, because the probabilities of the two probability matrix of HMM are all less than 1. The more probabilities of less than 1 multiply, the result tends to zero and cannot be ruled when being calculated the path probability with Viterbi algorithm. The solution is to multiply the probability by using the probability of the logarithm instead. But when doing so, the zero in two probability matrix should be instead of a small number (such as 0.000001).

Conclusions

After we get speech recognition control commands, we can control the corresponding household appliances. The J-LINK simulator can directly use the KEIL to burn program to the microcontroller. After the compilation, directly download, burn and simulate program.

In this paper, the C language is used to realize a speech module to control the switch of the household appliance. The recognized speech unit is “open computer”, “off computer”, “open refrigerator”, “off refrigerator”, and other simple commands. The recognition rate seems not to be connected with the state number of HMM. When the state number is N=6, N=8, N=10, the recognition rate is almost the same. However, the recognition rate is connected with the sufficient times of training. For the first scheme, when used to identify a specific person and training times of the voice unit is 5. The recognition rate can reach 99%. When used to identify non specific person, such as three persons, and the same training times, the recognition rate is up to 96%.

For second scheme, the recognition rate can reach more than 99%, when used to identify a specific person. And recognition rate can reach 98% when it is used for non specific person, such as 3 persons. The each voice unit training times is 5.
Acknowledgement

This research is financially supported by the National University Student Innovation and Entrepreneurship Training Program (201410876004).

References

