Excerpt：Implement a speaker-Independent voice recognition system which is under low cost.

This work ,an independent voice recognition system ,costs under fifty dollars.

(1) Research motivation

There are many costly electronic devices having voice recognition in our lives ,such as siri and google. However, they can't run at the cheap devices,so I want to devise a 50 NT dollar voice recognition system, which can be executed on the cheap devices.

(2) Research Purpose

1. low cost 2. simplicity 3. no training is needed in advance

(3) Research Devices

1. PC 2. Breadboard 3. Completed products

(4) Research Process

Introduction of voice recognition hardware (The red dotted line is my system)

I Will simply introduce how I recognize the voice. First, when the voice comes into the microphone, the sonic wave will turn into weak signal.

Second ,via the operational amplifier to turn the signal stronger ,and filter the voice which exceeds 4000Hz. (Because people's voice are under 4000Hz)

Finally ,the software in the microprocessor will deal with the voice.

Introduction of software system in the microprocessor

First ,the voice getting into the microprocessor will have an interruption 8000 times per seconds ,and every interruption will do a AD conversion.

The AD conversion is to change the analog signal into the digital signal , and this process is also known as sampling. The speed of sampling and resolution of sampling have a great impact on the cost of hardware. Therefore ,I use 8k Hz sampling Frequency and 12 bits sampling resolution ,is the cheapest choice. Although the sampling resolution is lowered ,the experiment shows that it doesn't have a great impact on recognition rate.

Next ,the speech endpoint detection. Find the real initial point and final point of the voice ,so that we can avoid processing unnecessary voice to reduce the recognition rate. Following are two ways which I devised to solve the problem.

(turn to the poster)
First, devise a buffer for block, and every block has 80 sound signals. I use the average energy of twelve blocks as the standard of recognizing the volume of voice (silence energy).When there are six loud voices appearing (louder than silence energy),it goes back to the first voice and assume it is the initial of the voice to run the voice recognition
Second, it is the state graph that bases on the volume of the voice in the buffer to decide whether should be in what condition. The final point is decided by this state graph.
For example :The first way mentioned above can be the condition from silence to sound, which is A situation

(5) Mel-Frequency cepstrun(MFCC)

After having the initial and final points, I will do MFCC for every buffer which includes 200 sound signals. MFCC is fits the voice recognition because it has considered that human ears vary in experiencing different frequencies,
Finally, every buffer will calculate 29 characteristics and use Hidden Markvol Modle to compare with databases. Counting the similar rate of the two to decide the end of the recognition.

(6) simplify pinyin

If the databases are large,Rom will raise and reduce speed.Therefore,how to simplify the databases is very important at ultra low cost voice recognition system.

Traditional chinese characters have 13060 words. In order to reduce databases,first,use 1300 Bopomofo to replace 13060 words, and then eliminate five tones.Finally,classify the similar sounds.

p.s. 1 .we can't recognize ㄓㄨˊ and ㄓㄨ after eliminating tones.

2.The command are is made up of at least two words, so many words could form the only combination to distinguish different commands.

(7) Cut the sound

A simple pronunciation should be cut so that we can get the characteristics of this pronunciation. According to the former and the latter pronunciation, cut every pronunciation into six pieces. Finally, we have 133 sounds. For example,ㄅㄤ and ㄆㄤ will change into ㄅ++ and ++ㄤ,ㄆ++and ++ㄤ,and then ++ㄤ could use the same database.

After combination, my databases are smaller than 13060 words

(1/100)Generate databases

First, I bought the pure voice data from MAT through my school,MAT160 and MAT400,these pure voice data are collecting different people's pronunciation, they can't be the databases directly.

(8) Combination

This chart is the introduction of my whole system.

Research result

After all the courses ,the recognition rate on PC is 82 and I also make the hardware reduced to the lowest level and let the cost of system under NT50.

Discussion

I could finish ultra low cost voice recognition system which is speaker-independent,have five important factors.

1. Amplitude normalization: usually it is executed by hardware, but I use software to replace it. Here is my experiment data. I adjust the sound's amplitude to test the recognition rate, and finally we can found that the recognition rates are almost the same. Thus, the amplitudes have only a little impact on it. Also, as mentioned before,12bits sampling resolution(the cheapest one) doesn't have any influence.
2. Cut the sound: After simplify pinyin, the databases are 1/100 of 13060 words.
3. Simplify databases：I use the pure voice data from MAT and calculate 29 characteristics by MFCC. Normally,MFCC will have 39 characteristics, but after my experiments, I found that only using 29 characteristics doesn't have a great impact. And my system is used integer operation to replace floating point operation. To sum up, the databases could be 10% smaller than the original ,in addition to the improvement of sound lookup table or Newton's Method to replace some parameter to sound simplification, it would become only 0.1%of the original.
4. Integer operation replace floating point operation
   Integer operation is 20 times faster than floating point operation.
5. Accelerating math operation

We can use Lookup table or Newton's method to replace some parameter to accelerate. For example,MFCC's Hamming Window is using 200 points to correspond sin. We can use Lookup table to calculate the number first, and we could only find the number while the sound is being recognized.