Voice Control System Based On Folding Mechanism

Wang Su-Xu, Cheng Wu-Shan

(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science)
Corresponding Author;Wang Su-Xu

ABSTRACT: Based on the folding mechanism, a control system capable of using voice control is designed, which also includes a WIFI module, which can realize WIFI remote voice control.
KEY WORDS: Voice controlWireless transmission

The biggest problem that a voice interaction system needs to solve is speech recognition, but most existing speech recognition systems either require a lot of computation. In response to the above problems, the system uses ARM Cortex M4 core STMicroelectronics 32-bit embedded microprocessor STM32F407, LD3320 speech recognition chip. And finally the human-computer voice interaction has been achieved.

I.   HARDWARE CIRCUIT DESIGN AND SELECTION SCHEME
1.1 Microcontroller core
The system uses the high-performance 32-bit STM32F407 from STMicroelectronics (ST) as the MCU. The chip has powerful built-in functions and peripheral interfaces to meet the needs of various scenarios. The chip can reach 180MHz, and has built-in high-speed memory: 256KB SRAM and 1024KBFLASH.

STM32F429IGT6 has three 12-bit ADCs, two 12-bit DACs, twelve 16-bit timers, two 32-bit timers, six SPIs, two full-duplex I2Ss, one SAI, and three IICs. , 8 serial ports and other bus or serial communication interface.

The control chip has the advantages of high stability, low power consumption, real-time performance, etc., and can meet the requirements of a speech recognition system based on a folding mechanism.

1.2 Speech recognition module
This system uses LD3320 non-specific human speech recognition chip produced by IC Route. It integrates a unique fast and stable speech recognition processor and optimization algorithm, so it can perform non-specific speech recognition without prior recording and training. According to the official data, the recognition accuracy can reach 95%.

In the LD3320 chip, a high-performance A/D converter, D/A converter, microphone audio signal input interface, audio decoding output interface, etc. have been integrated, and no auxiliary Flash chip, RAM chip and AD chip are required. As long as the user transmits the keyword sentence to the chip in the form of characters, the user can perform the approximation analysis on the recognized speech signal, and output the response result after recognition.

II.   BASIC PRINCIPLES OF SPEECH RECOGNITION
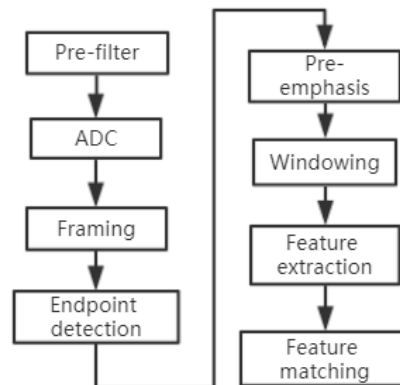### 2.1 Feature extraction and feature model
Firstly, the most important technical part of speech recognition is introduced: feature extraction and feature model establishment.

Feature extraction generally falls into two categories: time domain feature extraction and frequency domain feature extraction. The extraction method selected in this system is frequency domain feature extraction, and linear prediction analysis (LPC) is used to extract features. In LPC, the spectral amplitude of a signal can be obtained by performing a small number of operations on characteristic parameters. Less computational complexity can reduce the digital rate at the time of encoding, which can improve recognition rate, reduce computation time, and improve work efficiency.

When a speech sample is acquired, the module enters the training mode, begins to extract feature and model creation, and stores the model in the speech recognition module. When speech recognition begins, the above process is repeated for the newly entered sample to obtain a new model, and the new feature model is compared with the stored model.

2.2 Basic process

The basic flow of speech recognition is shown in Figure 1. It can be divided into six parts: signal acquisition, preprocessing, endpoint detection, feature signal extraction, model training and feature matching. The signal acquisition can be divided into front-end amplification, pre-filtering and ADC pre-preparation.

Figure 1 Basic flow of speech recognition

After the sound signal is radiated, it generates 6dB/oct attenuation, which has an adverse effect on the feature extraction of the signal. Therefore, after the acquisition is completed, the high frequency compensation is performed, and the pre-emphasis processing is performed. Due to the short-term characteristics of the sound signal, the collected data is segmented into segments of the same length for short-term analysis, and then the segmented signal is windowed.

In a signal with a background sound, the start and end points of the valid information need to be detected. We need to ignore background sound information and extract valid information. At this time, it is necessary to extract the Mel frequency cepstral coefficient (MFCC) of each frame of the speech signal in the effective speech. Multiple samples are acquired for each voice instruction, and feature templates for each instruction are obtained from the samples. The dynamic time warping algorithm is used to calculate the matching distance between the input speech signal and each template, identify the input speech signal, and perform feature matching.

III. CONCLUSION

Human-computer voice interaction is a research hotspot and difficulty in the field of current speech signal processing. This design uses the STM32 microprocessor as a controller to achieve human-machine voice interaction. This design is small in computation and does not require networking. It is suitable for embedded devices.

REFERENCES

[1]. TomprosS,MouratidisN,DraaijerM,etal.Enablingapplicationofenergysavingappliancesofthehomeenvironment[J].IEEENetwork,2009, 23(6):8-16．

[2]. Lee E A,Seshia S A.Introduction to embedded systems:acyberphysical systems approach[M].Lee & Seshia,2011:4-5.

[3].]STM32F407-417Datashee［EB /OL］. https://www.st.com/en/microcontrollers/stm32f407-417.html?querycriteria=productId=LN11

[4]. LD332XDatasheet[EB/OL2010-09-11http://www.icroute.com/doc/LD3320%E6%95%B0%E6%8D%AE%E6%89%8B%E5%86%8C.pdf

[5]. LD332XDatasheet[EB/OL]2011-10-13http://www.icroute.com/doc/LD3320%E5%BC%80%E5%8F%91%E6%89%8B%E5%86%8C.pdf

[6]. Zhao Li. Speech signal processing[M]. Beijing: Mechanical Industry Press,2007

[7]. Du Xidong, Wang Zhiwei, Pan Li, etc. Design of non-specific voice control car based on LD3320[J].Information Technology，2015（1）:53-55.

[8]. XiongZhihui, Wang Renbo. Intelligent learning table lamp design based on STM32[J]. Electronic world, 2016（8）:160-161.