Small Intelligent Home System with Speech Recognition Based On ARM Processor

Hua Jiang [1,2 +], Zihao Chen [1]

[1] School of Information Science and Technology, Nantong University, China

[2] School of Tongke Microelectronics, Nantong University, China

Abstract. The research is to design and implement the speaker independent speech recognition home system, and the final function achieves the recognition of isolated words. Through the module function integration, the speech recognition module LD3320 and the development board OK6410-A are connected. The results are processed, analyzed and judged after the speech module recognition results have been received by the ARM processor. Then the processor calls its own module function to indicate the recognition results. The program is implemented by C language in cross system compiler under Linux system. The downloading program is completed by the debugging function of the uboot. The opening and closing of the LED and the buzzer on the development board is completed by using 4 simple speech commands, including "kai deng", "la ba", "guan deng" and "fou jue". Then the intelligent home environment was simulated by 10 complex speech commands, including "wo shi", "chu fang", "ke ting", "dian shi" and so on. The results shows the small intelligent home system has been achieved.

Keywords: Speech recognition, ARM, Linux, OK6410-A, LD3320

1. Introduction

Embedding and hiding speech recognition system in the ARM platform will make full use of their advantages. Such products will be able to simulate people's consciousness and behavior for achieving more intelligent and humane. When users use an embedded speech recognition system, it can transform the original complicated mechanical operation steps into simple and understandable speech commands. Speech commands give users great convenience, especially in the disabled medical and elderly service industries.

The first speech recognition system in the world was born in Bell Laboratories in the mid-20th century. This system is based on English pronunciation. In China, the research of speech recognition technology began at the Institute of Acoustics of the Chinese Academy of Sciences (IACAS). Nowadays, the research of speech recognition technology in China has developed rapidly in combination with the characteristics of Chinese Pinyin pronunciation. Many institutions have invested in the research of speech recognition technology[1-3].

The next section describes the system architecture and design principle. Section 3 details the software design of the system. The system function testing is provided in section 4. Section 5 presents conclusions.

2. System Architecture And Design Principle

2.1. System design

The purpose of the design is to recognize the isolated words spoken by the speaker-independent. The system is mainly composed of the microphone (MIC), the LD3320 and the OK6410-A, which is shown in Fig. 1. That is, when the MIC collects the vocabulary, it is handed over to the speech module for processing.

+ Corresponding author: Hua Jiang Tel: + 86 13962954626; fax: +86 51385707018.
  E-mail address: ntit9818@163.com.

The module sends the processed results to the ARM processor for analysis and judgment, and the processor uses its peripherals to indicate the judgment results.
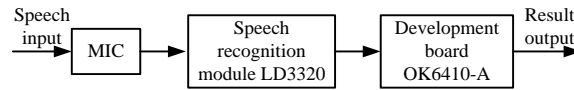

Fig. 1: System architecture

2.2. Hardware components

Hardware is mainly composed of three parts. The first part is the speech recognition module LD3320, which is used to collect and receive speech signals. After receiving, the words stored in the keyword list of the LD3320 are compared and recognized. After the recognition is successful, a binary return value will be issued. The second part is the development board OK6410-A, which completes the processing and judgment of the LD3320 return value. The communication mode of the first two parts is to send and receive information through the UART. The third part is the peripheral circuit of the OK6410-A, which uses the speech result to operate the peripheral circuit after processing and judging. At this time, the result is actually the binary number. In the experiment, the LED lamp and the buzzer are used to show the different control functions of speech signal to them.

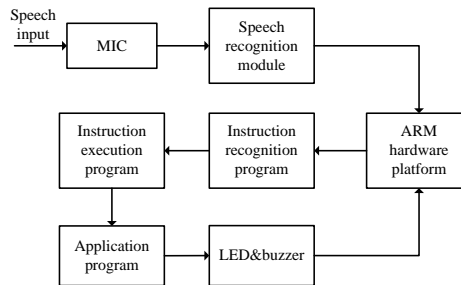The overall design block diagram of the system hardware is shown in Fig. 2.

Fig. 2: System hardware block diagram

The OK6410-A is equipped with a total of four small LED lights. Low level turns the diode on. Because buzzer is opened and closed by the pulse width modulation (PWM). The LCD screen mounted on the development board is 4.3 inches in size (480*272). The LCD screen can be used to display the results of each recognition. The MIC achieves speech pickup functions[4].
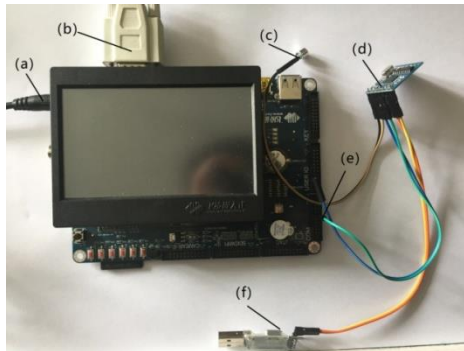

Fig. 3: Physical connection diagram

The hardware connection diagram is shown in Fig. 3. (a) is the power supply for power lines and development boards; (b) is the debugging line of the serial port and the serial port information of the development board can be easily viewed through the serial port tool DNW; (c) is the MIC, which is used to collect speech signals; (d) is the speech recognition module LD3320; (e) is the serial communication line and the two terminals are connected to the TXD terminal of the LD3320 and the RXD terminal of the OK6410-A UART1 respectively; (f) is the tool of the serial port to TTL and it can be used to edit the LD3320 on the PC side.

3. Software Design of the System

The purpose of software design is to control the switch of the LED and the buzzer by four simple speech commands. On this basis, the design simulated an intelligent home scenario experiment and enriched the number of commands. Through ten complex speech commands, the design achieves the control of small intelligent home system.

3.1. Experimental basis

3.1.1. Cross compilation

The development environment and the download environment is based on the cross-compiler environment of the Linux system and the DNW tool download of the bare-machine file under the uboot[4-5]. This design chooses the ARM platform and the Linux system, using the cross-4.2.2-eabi compiler which compiles the binary program image (.bin) and runs on the OK6410-A. The Vmware is used to install the Centos7 system for the construction of the compiler environment.

3.1.2. DNW tool

In the design, the DNW is an indispensable serial port tool when experimenting with the OK6410-A. The DNW can be used to download bin files, burn image files, run images and other functions. The DNW is a serial port small software developed by Samsung. Its status is equivalent to the super terminal under the Windows XP system. On the basis of the DNW, many new functions have been added, such as transferring files between the PC and the development board with the USB.

Serial ports can be set in the configuration of the menu bar such as the baud rate, the serial number, and the starting address. And .bin files can be downloaded through the USB cable in the USBport.

3.1.3. Download in uboot

The BIOS process in the Linux system is called Bootloader, which is a program that must be run before the Linux system is run formally. Most BootLoaders contain two different modes of operation. They are called the boot loading mode and the debugging mode respectively. The download function in the debugging mode is used, and the. bin file needs to be downloaded in the uboot boot. Specific steps of downloading the uboot are shown in Fig. 4.

> *Step1: Make a SD card for one-click downloading the uboot.*
> *Step2: Set up the development board to start from the SD card.*
> *Step3: Power on.*
> *Step4: Connect the serial port of PC with COM0 of OK6410-A by serial port line.*
> *Step5: Open DNW and set up serial port information.*
> *Step6: Turn on the switch and start downloading automatically.*
> *Step7: Put the dial switch back to NANDFLASH to start.*

Fig. 4: Steps of downloading uboot

3.2. Design simple functional program

The software program flow of the speech recognition system is mainly divided into three parts[4-6]. The first part is to transmit the speech collected by the MIC to the LD3320. In the second part, the LD3320 compares the received speech with the words in its keyword list library and determines that a recognition result is passed to the OK6410-A as a return value. In the third part, the OK6410-A judges the return value one by one to control the LED and the buzzer. The design flow chart is shown in Fig. 5.
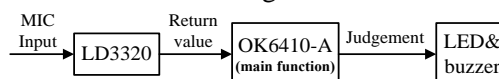

Fig. 5: Design flow chart of software program

3.2.1. Initialization of keyword list

This design uses the serial communication as the communication mode. First, the keyword list of the LD3320 needs to be initialized and edited. The module can be edited online by any serial assistant software.

After connecting the speech module with PC, the serial assistant XCOM V2.0 is used for serial communication[4-6].

According to the manual of the speech module, some commands can be used to edit the keyword list online, such as {d1} for entering the debugging state and {d0} for exiting the debugging state. In the simulation experiment, four simple commands are preseted first. {a0kai deng} and {a0guan deng} indicate that the "turn on" and "turn off" commands are added to the keyword list, respectively. And {c0} represents clearing the original keyword list. {l0} denotes a list of overloaded keywords, which is generally used after all new commands have been added. In the experimental stage, four key words including "kai deng", "la ba", "guan deng" and "fou jue" are written to the LD3320. The list of keywords is shown in Table 1.

Table 1: List of keywords for LED and buzzer

Serial number	Chinese pinyin	Keyword setting	Action
0x00	kai deng	{a0kai deng}	Turn on LED
0x01	la ba	{a0la ba}	Turn on buzzer
0x02	guan deng	{a0guan deng}	Turn off LED
0x03	fou jue	{a0fou jue}	Turn off buzzer

After keywords are added, the LD3320 will correspond each keyword to the hexadecimal number in order. According to the four preset commands, the corresponding return values are 0x00, 0x01, 0x02 and 0x03. These four numbers are the return values of the successful recognition of the speech module.

3.2.2. Programming of UART serial communication

The communication of the return value is carried out by the LD3320 and the UART1 of the OK6410-A, so the UART needs to be programmed. The UART serial communication of the OK6410-A uses many different registers. When each register is assigned different values, it will have different functions. And the communication mode is also divided into the first-in-first-out (FIFO) and the non-FIFO. This design chooses the non-FIFO mode.

3.2.3. Programming of main function

After the UART function is designed, the return value can be received normally through the serial port. After receiving the return value, the different return value will be judged and the corresponding operation will be carried out. And in the process of operation, it constantly detects whether there is a new value entering. And if there is, it enters the operation corresponding to the new value. If not, repeat the current operation.

3.2.4. Programming of LED and buzzer

Programming of the LED are as follows: (1) The LED of the OK6410-A is configured to the GPM port. (2) The GPMCON address is 0x7F008820. (3) The GPMDAT address is 0x7F008824. If one of the last four bits is set to 0, the corresponding LED light is on. The buzzer is programmed according to the GPF15 register function in the S3C6410 user manual[4-5].

3.3. Complex scenario simulation program

In order to enrich the commands, ten complex speech commands which are often used in the home environment are selected and simulation experiments are completed[7]. The LCD screen equipped with the OK6410-A is used to display the results. When a specified speech command is input, the LCD will display the corresponding results of the input speech command in real time through programming, and the results will be displayed in English. The command set is shown in Table 2.

Table 2: Command set of intelligent home system

Serial number	Chinese pinyin	English	Keyword setting	Action	Serial number	Chinese pinyin	English	Keyword setting	Action
0x00	wo shi	bedroom	{a0wo shi}	Turn on bedroom light	0x05	kong tiao	air	{a0kong tiao}	Turn on air conditioner
0x01	chu fang	kitchen	{a0chu fang}	Turn on kitchen light	0x06	tai deng	light	{a0tai deng}	Turn on desk lamp

0x02	ke ting	diningroom	{a0ke ting}	Turn on diningroom light	0x07	sao di	clean	{a0sao di}	Cleaner work
0x03	dian shi	tvison	{a0dian shi}	Turn on TV	0x08	jia shi	humidifier	{a0jia shi}	Turn on humidifier
0x04	bin xiang	fridge	{a0bin xiang}	Power on refrigerator	0x09	di nuan	heater	{a0di nuan}	Turn on floor heating

The main function contains initialization functions for each component. Program judgment is executed by the case statement with 10 branches. Each branch displays different characters. The character display calls the lcd_putc() function. In this character display function, the character library function has been connected. Each English character is displayed through a character library.

4. System Function Testing

The test platform is the PC and the OK6410-A. The main program of speech recognition running on the OK6410-A is the generated binary file with the .bin suffix described above, which is the main part. The PC side uses the DNW to burn programs and view information, which is the auxiliary part.

Firstly, the power cord of the OK6410-A is connected. Then the serial port of the OK6410-A and the USB download line connect is connected to the PC. Finally, the TXD of the LD3320 is connected to the RXD of the OK6410-A UART1. The power supply of the LD3320 is still completed by the USB to the TTL module.

4.1. Simple functional testing

Four simple commands are tested, and the test vocabulary is "kai deng", "la ba", "guan deng" and "fou jue". After connecting the objects, the experimental step is to open the DNW on the PC and power the OK6410-A, then quickly press the space on the PC and enter the uboot.

Input of "dnw 0x50000000" means that data is stored from address 0x50000000. The address value is related to the initial setting of the DNW. Next, the. bin file can be downloaded through the USB on the PC side. The next step is to select the test. bin file to download. After typing "go 0x50000000", it means that the development board should start executing the program from the address 0x50000000. At this point, the speech command that the experiment presupposes in advance, such as "kai deng", will be given to the MIC, which will make the corresponding light-emitting diodes on the development board bright. Then the test of "la ba", "guan deng" and "fou jue" is carried out in turn.

4.2. Scenario simulation experiment

After the testing in Section 4.1, the keyword recognition list in the LD3320 should be re-edited by the XCOM serial assistant on the PC.

The first step is to type {c0} to clear the keywords. Then, ten new commands are set according to Table 2. Also ten speech commands are written in the keyword list in the hexadecimal order from 0x00 to 0x09. The last step is to type {l0} and reload the new commands so that the new ten commands can be loaded into the keyword list.

After the keyword list of the LD3320 is reset, the test of scenario simulation experiment is started. The physical connection is still shown in Fig. 3. After that, the file lcd.bin is downloaded to the OK6410-A to run and test, using the DNW to do the same operation based on Section 4.1. For example, after entering "wo shi", "ke ting", "sao di", "kong tiao" and "di nuan" into the MIC, the test results are shown in Fig. 6. The whole recognition process is real-time. The LCD can display the corresponding results immediately after the input of the MIC.

Fig. 6: Test results of scenario simulation experiment

5. Conclusions

In this paper, a small intelligent home system with speech recognition was proposed and implemented based on the ARM processor with the speech recognition module LD3320 and the development board OK6410-A. Four simple speech commands was designed and downloaded for verification. Then, ten complex speech commands were also proved by the fast and accurate small intelligent home system. After testing the non specific human speech recognition system, the system identified the speaker's speech commands well. On the basis of the speech intelligent home test, it had made a good paving for the realization of the speech cloud intelligent home by using ZigBee and other wireless communication means.

6. Acknowledgements

7. References

[1] Y. Zhang, P. Zhang, and Y. Yan. Long short-term memory with attention and multitask learning for distant speech recognition. Journal of Tsinghua University, 58(3): 249-253.

[2] J. Chaloupka. Digits to Words Converter for Slavic Languages in Systems of Automatic Speech Recognition. Proceedings of the 19th International Conference on Speech and Computer, SPECOM 2017, pp. 312-321.

[3] T. Fu, X. Wu.Multi-scale feature based convolutional neural networks for large vocabulary speech recognition. Proceedings of the 2017 IEEE International Conference on Multimedia and Expo, ICME 2017, pp. 1093-1098.

[4] Forlinx Embedded Tech. Co., Ltd.. OK6410-A Development Board Hardware Manual V2.1. http://www.forlinx.com/14.htm. 2010.

[5] Samsung Electronics Co., Ltd. S3C6410X RISC Microprocessor User's Manual. http://www.waveshare.net/datasheet/ SAMSUNG_PDF/S3C6410.PDF, 2008.

[6] L. Wang. Design of speech recognition system based on LD3320 chip. Proceedings of the 2016 3rd International Conference on Materials Engineering, Manufacturing Technology and Control (ICMEMTC 2016), pp. 180-183.

[7] P. Li, J. P. Li. Embedded intelligent home control system based on Arm-Linux. Proceedings of the 2012 International Conference on Wavelet Active Media Technology and Information Processing, ICWAMTIP 2012, pp. 429-431.