Application Research and Implementation of Voice Control System Based on Android Speech Recognition

Xiangrui Meng1, *, Yan Song2

1Applied Technology College of Dalian Ocean University Liaoning, Dalian 116300
2Dalian Wafangdian Normal School Liaoning, Dalian 116300

*Corresponding author: xiangruimeng@dlou.edu.cn

Abstract. Based on the voice recognition technology of Android platform, combined with WiFi network communication technology and embedded control system, an audio voice control system is realized, which collects voice commands on the mobile phone APP for recognition and controls the operation of audio equipment through WiFi network and embedded control module. The voice recognition adopts the voice recognition SDK provided by IFLYTEK open platform, the WiFi network is provided by WiFi router, and the embedded control module adopts the STC15F204EA controller, which is matched with NL6621-M3WiFi module to receive processing instructions. After experimental test, the audio system can be effectively controlled by voice, which meets the requirements of modern audio equipment intelligence.

Keywords: Speech recognition; IFLYTEK open platform; Android; WiFi network communication, WiFi module; Embedded control module.

1. Introduction

With the development of electronic technology, people's requirements for sound system are no longer limited to using remote control, and they want to control more freely, comfortably and conveniently. At the same time, with the rapid development of smart phones, mobile phones have been popularized in people's lives, and more and more functions are integrated into smart phones [1]. In the family residence, relying on WiFi communication technology, the mobile phone APP is connected with the audio equipment, and the audio equipment is controlled by the mobile phone to construct an efficient and convenient audio control system, which is convenient for people to operate and manage the audio equipment. As a new intelligent information technology, language recognition technology has attracted wide attention and high attention [2], which can make machines understand people's words. People can use voice to control the equipment, which makes man-machine interaction more convenient and direct [3]. Voice control replaces manual control, which can free people's hands and allow people to control the sound system while doing other things. In this paper, combining the speech recognition technology of Android platform, WiFi network communication and embedded technology, an audio voice control system based on Android speech recognition is proposed. The system can control audio equipment by WiFi network and embedded control module on the mobile phone, which can meet the requirements of modern audio equipment intelligence and has a good application prospect.
2. System design

2.1. Overall plan
The whole sound control system consists of mobile phone control terminal, WiFi router, embedded home gateway and sound equipment. The mobile phone develops APP based on Android platform, which has the functions of voice command control and interface command control. Voice command control is based on the voice recognition function provided by IFLYTEK open platform. Embedded home gateway is the embedded control module, which uses STC15F204EA single chip as microcontroller and NL6621 chip integrated with Cortex-M3 as WiFi communication module. WiFi router provides WiFi communication network, connects mobile phone with home gateway, and provides networking between mobile phone and Internet. The audio equipment receives instructions from the embedded home gateway and executes them. The overall structure of the system is shown in Figure 1.

![Figure 1. Overall structure diagram of system](image)

The mobile phone APP sends the received voice to IFLYTEK Voice Cloud Service on the Internet through WiFi router for voice recognition, matches the data returned from the cloud with the local instruction, finds the matching instruction and sends the instruction to the home gateway through WiFi router. After receiving the instruction, the home gateway sends a control signal to the audio equipment, and the audio equipment executes corresponding actions according to the control signal. If speech recognition is performed locally, there is no need to connect to the Internet. If the mobile phone is controlled by interface commands, it does not need to recognize speech, and sends commands directly.

2.2. Speech recognition processing
Speech recognition technology is essentially a process of pattern recognition and matching, and it is a technology that the machine transforms the speech signal into corresponding text or command through the process of recognition and understanding [4]. IFLYTEK open platform provides a variety of speech recognition functions [5], among which speech dictation is based on natural language processing, which converts a speech into text and extracts the text information contained in the speech. Speech dictation includes online engine and offline engine. The online engine needs to use the network, which is slower and generates a certain amount of traffic, but it has a better recognition effect. Offline engines do not need to use the network, and the recognition speed is faster, but they need to use the offline SDK package [6]. The mobile phone voice recognition of this system adopts the mobile voice terminal SDK provided by IFLYTEK open platform, and delivers the locally collected voice to IFLYTEK voice cloud for recognition through the network, or to IFLYTEK local voice recognition SDK package off-line for recognition, and then matches the returned recognition results with the system-related instructions and executes the corresponding instructions. The software workflow is shown in Figure 2.
Figure 2. Workflow of software

It can be seen from Figure 2 that the mobile phone speech recognition supports online engine and offline engine. The online engine must ensure that the mobile phone is connected to the network, so as to send the voice to IFLYTEK Voice Cloud Service, and the cloud server will return the recognition result after recognition. The local engine must pre-install the local speech recognition package before use, and use the recognition package for speech recognition. The mobile voice terminal SDK provides the SpeechUtility class for software initialization, the SpeechRecognize class realizes the voice dictation function, and the RecognizerDialog class realizes the voice dictation UI function. Most of the functions provided by the mobile voice terminal SDK use interfaces and callbacks to return results and status. The speech recognition result is returned in the callback function, and the system processes the recognition result in the callback function [7]. The recognition accuracy of speech recognition is the core index of system function. The speech recognition function provided by IFLYTEK open platform has a high recognition accuracy [8]. However, from many tests, it is found that the recognition rate is relatively high for continuous sentences with contextual semantics, and the recognition rate decreases for some single phrases without contextual connection such as "stop" and "menu three". In order to make the declining recognition rate not affect the matching rate of the system, the system adds a list of similar syllables of instructions. For example, the similar syllable list corresponding to the "Stop" command includes similar syllable phrases such as "Pavilion", "Listen", "Straight" and "Gold". As long as the speech recognition result contains one of the phrases in the similar syllable list of the command, it is considered to match. Test before and after adding similar syllables list, and the test results are shown in Table 1 and Table 2. Among them, the number of correct recognition refers to the number of times that the original instruction phrase can be correctly recognized, and the number of matching times is the number of times that the recognition result matches any phrase in the similar syllable list. From the test results, it can be seen that the correct rate of single instruction speech recognition is low, but the matching range is enlarged after adding the similar syllables list of instructions, which can improve the matching rate of speech recognition of the system.

Table 1. Test results before adding a similar syllable list

| Recognition instruction | Number of tests | Correct recognition times | Recognition accuracy rate /% |
|-------------------------|----------------|--------------------------|-----------------------------|
| Stop                    | 150            | 108                      | 72                          |
| Menu 3                  | 150            | 113                      | 75                          |
### Table 2. Test results after adding a similar syllable list

| Recognition instruction | List of similar syllables                      | Number of tests | Matching times | Matching rate /% |
|-------------------------|------------------------------------------------|-----------------|---------------|------------------|
| Stop                    | Pavilion, listening to gold, standing straight | 150             | 142           | 95               |
| Menu 3                  | In Danshan, but sour, in single, simple three  | 150             | 146           | 98               |

2.3. Embedded home gateway

Embedded home gateway takes STC15F204EA microcontroller as the core, connects with audio equipment, cooperates with NL6621-M3WiFi module, receives the instructions transmitted by WiFi network, and controls the operation of audio equipment. The hardware structure of embedded home gateway is shown in Figure 3.

![Figure 3. Structure diagram for hardware of embedded home gateway](image)

2.4. Android SDK architecture

Android platform is composed of operating system, middleware, user interface and application software. It adopts the architecture of software stack layer, which is mainly divided into three parts. The bottom layer is based on Linux kernel, developed by C language, and only provides basic functions; The middle layer includes function Library library and VirtualMachine, which is developed by C++. The upper layer is a variety of application software, including call program, short message program, etc. The application software is developed by each company, and Java is used as a part of programming.

The first layer, application layer, provides some core application packages, such as email, SMS, calendar, map, browser and contact management. At the same time, developers can use Java language to design and write their own applications, which are equal and friendly with those core applications.

The second layer is the application framework layer, which is the foundation of Android application development, and most developers are dealing with it. The application framework layer includes ten parts: activity manager, window manager, content provider, view system, package manager, phone manager, resource manager, location manager, notification manager and XMPP service.

The third layer, system library and Android runtime, the system library includes 9 subsystems, which are layer management, media library, SQLite, OpenGLEState, FreeType, WebKit, SGL, SSL and libc. Android runtime includes core libraries and Dalvik virtual machine. The former is compatible with the functions that most Java languages need to call, and also includes core libraries of Android, such as android.os, android.net, android.media and so on. The latter is a Java virtual machine.
based on registers, and Dalvik virtual machine mainly completes important functions such as life cycle management, stack management, thread management, security and exception management and garbage collection.

The fourth layer, Linux kernel, Android core system service depends on Linux2.6 kernel, such as security, memory management, process management, network protocol stack and driver model. Linux kernel is also the abstraction layer of hardware and software stack. Drivers: display driver, camera driver, keyboard driver, WiFi driver, Audio driver, Flash memory driver, Binder (IPC) driver, power management, etc.

2.5. WiFi network communication

WiFi is a short-range wireless transmission technology, which can support wireless signals for Internet access within hundreds of feet (about 100m) [10]. The system uses WiFi network communication technology to realize wireless transmission of control command signals. WiFi router is used as wireless signal transmitter, and mobile phone and embedded home gateway are used as AP access [11]. Therefore, both the mobile phone and the embedded home gateway are assigned dynamic IP addresses, and both parties must identify each other's IP address before communicating, so as to carry out subsequent control instruction transmission. The network communication flow is shown in Figure 5.

![Figure 4. Workflow of embedded home gateway](image)

![Figure 5. Flow chart of network communication](image)

When the mobile phone first accesses the network, it broadcasts and sends the probe packet through the bound port number. After receiving the probe packet, the embedded home gateway bound with the corresponding port number replies a response packet. The response packet contains the IP address of the home gateway. The mobile phone can obtain the IP address of the home gateway through the response packet, and then can unicast the control instruction packet to the home gateway.
3. System test

After the development and debugging of sound control system, the control interface of mobile phone is shown in Figure 6. The mobile phone includes voice command control and interface command control. Using the voice command control mode, the user only needs to click the voice button and then speak the command. The mobile phone collects voice commands and controls the operation of the sound system. Using the interface command control mode, the user needs to click the corresponding command button to control the sound work.

![Figure 6. Interface of mobile phone APP](image)

System testing tests speech recognition in online engine and offline engine, and records matching rate and average response delay. The test results are shown in Table 3.

| Recognition mode | Matching rate/% | Average response time /s |
|------------------|-----------------|--------------------------|
| Online engine    | 98              | 1.2                      |
| Offline engine   | 97              | 1.0                      |

It can be seen from the test results that both online engine and offline engine can obtain better matching rate after adding similar syllables list of instructions. The average response delay of online engine is lower than the evaluation response delay of offline engine, because online engine needs to send data to the cloud for processing. No matter which recognition method is used, it can meet the system requirements for speech recognition.

4. Conclusions

In this paper, based on the mobile voice terminal SDK provided by IFLYTEK development platform, combined with WiFi network communication technology and embedded system, an audio control system controlled by voice is developed. The system uses mobile phone voice recognition as the control terminal, and sends control instructions to the embedded home gateway through WiFi network, so as to achieve the effect of voice control of audio equipment operation. At present, this system has passed the experimental test, and can well control the operation of audio equipment. In the future, we will further study and extend the system architecture to smart home control and other embedded device control.

References

[1] YE G, RAN B. Design and implementation of smart home system based on Internet of Things [J]. Journal of computer applications, 2014(S1):323-324.
[2] TANG Guiyao, WAN Xin. Research and basic realization of speech recognition technology [J]. Electronic technology & software engineering, 2015(15):128.

[3] YANG Yefen, YE Chengjing. Intelligent home voice control system based on GSM [J]. Computer systems & applications, 2017, 26(2):68-72.

[4] WANG Aiyun. Intelligent home speech recognition system based on NL6621 [J]. Computer engineering & software, 2015, 36(7):104-107.

[5] YU Yixiao, ZHENG Qilin, CHEN Xinyu. Implementation of intelligent home system based on speech recognition and MQTT protocol [J]. Internet of Things Technologies, 2017, 7(11):97-99.

[6] IFLYTEK. IFLYTEKMSC Development Guide: Engine Types [EB/OL]. [2018-06-19]. https://doc.xfyun.cn/msc_android/%E5%BC%95%E6%93%8E%E7%B1%BB%E5%9E%8B.html. IFLYTEK. Development guide of IFLYTEK MSC: engine type [EB/OL]. [2018-06-19]. https://doc.xfyun.cn/msc_android/%E5%BC%95%E6%93%8E%E7%B1%BB%E5%9E%8B.html.

[7] IFLYTEK. IFLYTEKMSC Development Guide: Overview [EB/OL]. [2017-07-04]. https://www.kancloud.cn/iflytek_sdk/msc_manual_andorid/299547. IFLYTEK. Development guide of IFLYTEK MSC: outline [EB/OL]. [2017-07-04]. https://www.kancloud.cn/iflytek_sdk/msc_manual_andorid/299547.

[8] Deep analysis of Weiss IFLYTEK's latest speech recognition system and framework [EB/OL]. [2016-08-25]. http://www.sohu.com/a/112088969_114877. WEI Si. Deep analysis of IFLYTEK's latest speech recognition system and its framework [EB/OL]. [2016-08-25]. http://www.sohu.com/a/112088969_114877.

[9] Baidu Encyclopedia. STC15F204EA [EB/OL]. [2018-05-10]. https://baike.baidu.com/item/STC15F204EA/6902552?fr=aladdin. Baidu Baike. STC15F204EA [EB/OL]. [2018-05-10]. https://baike.baidu.com/item/STC15F204EA/6902552?fr=aladdin.

[10] LI Qing, ZHANG Ya. System of wireless voice- controlled house [J]. Manufacturing automation, 2011, 33(3):218-220.

[11] YAN W, WANG Q, GAO Z. Smart home implementation based on Internet and WiFi technology [C]// Proceedings of 34th Chinese Control Conference. Hangzhou: IEEE, 2015: 9072-9077.