

# Design of Song Ordering System Based on Speech Recognition

Yutian Huang, Donglin Tang, Zhaoming Zhou

Southwest Petroleum University, Chengdu 610500, China

**Abstract:** In this paper, the song-ordering system designed based on speech recognition uses STM32 single-chip microcomputer as the main control unit to realize the corresponding functions. The designed system includes the design of hardware and the realization of software functions. The rest have LD3320 speech recognition module, OLED display module, and voice broadcast module as song playback module. The working process is mainly completed by sending the protocol, that is, the voice recognition module receives the voice command, and the serial port sends the protocol to trigger the microcontroller control part. The system meets the following functions: it has the function of voice recognition and analysis, and completes the remote control of the corresponding tasks by recognizing voice commands; it realizes the basic voice song request function, which meets some basic requirements and can display the relevant information of the currently playing song. After continuous debugging and functional testing, the system successfully realized the functions of using voice to control the completion of song ordering, displaying the basic information of the song on the display screen, and successfully playing the song. Finally successfully completed a language song ordering system with stable operation, correct functions, convenient, quick and intelligent.

**Keywords:** Speech Recognition; STM32 microcontroller; voice chip; artificial intelligence.

## 1. Introduction

In recent years intelligent tools and other increasingly replace the traditional meaning of some tools, such as typing in the input often have direct voice recognition<sup>[1]</sup> input, that is, directly to the device to say a paragraph, the device will be able to intelligently recognize the content of the speech, converted into text input. So it is necessary to design an intelligent system based on the technology of voice recognition to be used in the entertainment life. The current technological development seems to be the faster development of speech recognition technology, AI, deep learning gradually improve the speech recognition technology from scratch to achieve a zero breakthrough, slowly from only in the laboratory to develop the theory and experimental data, to have some small physical applications, and then maturely developed application products and go to the market, and gradually go to the people's lives. Now used voice input method, voice-controlled smart home, behind are inseparable from the application of voice recognition technology[1]. And daily life, the current stage of society in all classes of people, all receive from all sides of a variety of pressures, everyone is caught in the busy life pressure, in the spare time to find a suitable form of entertainment to relax the mind and body is necessary. Therefore, based on voice recognition to design a song system, in line with the trend of modern intelligence, but also to facilitate the leisure and entertainment of households, not only individuals need such a system at home, outside such as some stores or other public places, there is a voice recognition of the song system can also quickly meet the requirements of the people, i.e., people are busy with their respective daily life, direct voice to the system to send instructions That is, when people are busy with their daily life, they directly send commands to the system by voice, and the system receives intelligence through voice recognition, and the control part completes the corresponding control function realization, and then carries out the

corresponding playback.

Combined with the above said, music is closely connected with people's life. Various kinds of Bluetooth speakers and smart speakers are also the market that is quickly occupied. This has to do with the fact that voice recognition technology is starting to be integrated into actual products, with a huge variety of smart speakers using this as a point of attraction. For different music, the audience is different, and different users have a variety of needs, so as the market for voice smart speakers grows bigger and bigger, all kinds of speakers continue to improve the user experience, adding rich music types and a variety of intelligent features. Therefore, there is great value in both research and application of the design of speech recognition based song ordering system.

## 2. Overall System Design Program

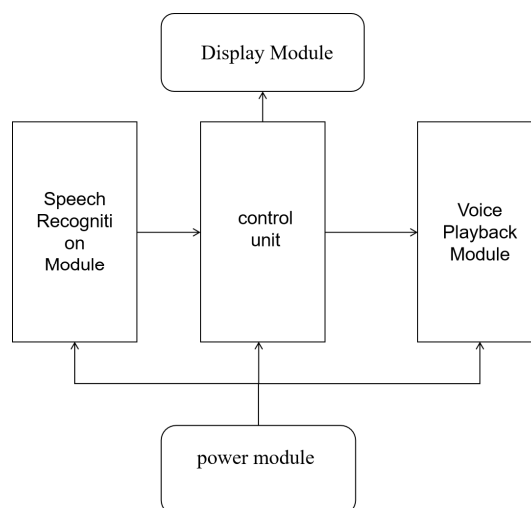


Figure 1. System flow chart

The system consists of five modules shown in the figure

above, the selection of microcontroller as the main control part [2] with speech recognition chip module for recognition, plus display module for display, voice playback module for playback. The speech recognition module converts the sound analog signals into digital signals, then recognizes the signals and sends the information to the microcontroller control module via serial protocol[3]; The control unit receives the corresponding information according to the protocol sent from the serial port to carry out the relevant command operations, such as song selection, playback, etc.; the display also receives the corresponding instructions to display the information of the song ordered by the user; the voice announcement module will play out the corresponding song.

#### (1) Program based on stm32 microcontroller master control

STM32F103C8T6 is used as the main control board, as well as speech recognition module LD3320 module and speech playback module JQ8400 module to form the system [4].

Speech recognition module for the LD3320, in fact, it is a simple 51 microcontroller, and then through a simple code can realize the function, communication can be achieved by connecting the serial port, that is, connecting GND, VCC, TX, RX several serial ports; LD3320 speech recognition chip, the internal integration of the speech recognition processor and signal modulation circuits, A/D converter, D / A converter, microphone input port, music output port, etc., are some high-efficiency, high-performance devices. LD3320 can recognize up to 50 key words and phrases [5]. Users only need to use the software program code to want to voice input keyword list compiled in advance, the sequence of these words and phrases and so on in the form of characters transmitted to the chip, the chip to store these key words and phrases, the user can be on the recognition of the part of the voice conversation, the chip's speed and accuracy are very high, and can be very good to recognize the information.

The choice of voice broadcasting module there are many on the market today, the more commonly used TTS voice modules, such as XFS5152, SYN6288, JQ8400 voice broadcasting chip and so on. The system uses JQ8400, the first feature is small size, the code is simple to realize the music playback, the second feature is cheap.

The STM32F103C8T6 of the main control board is a widely used, one of the common types of microcontrollers on the market. STM32 has many advantages, such as its ultra-high performance, excellent communication interfaces, and also contains many other functional modules.

#### (2) Program based on Sunplus MCU master control

This program uses a 16-bit Sunplus microcontroller as the main control chip, with control functions, can handle data as well as processing digital signals, comes with a voice recognition module, including general recognition devices have the AD module, DA module, automatic playback function and manual playback function. This makes it not only a microcontroller, but also can be used as a voice system, the voice can be recognized and played, functional diversity, which makes it can be used for multi-purpose, used to design the voice recognition song system quickly, the existing functions can be realized together, which is why there are a variety of modern microcontrollers on the market now applications. At the same time the microcontroller also has a

display module.

In order to play the effect of music and the accuracy of speech recognition can be added under the premise of the main control with the Sunplus microcontroller voice recognition module and voice broadcasting module, respectively, with WTK6900B02 and MP3 decoder chip module [6]. The former is a voice recognition module developed by Shenzhen Wei chuang Zhi yin Electronic Co, Ltd, which can realize the function offline, and the high recognition rate is the selling point of this product.

#### (3) Program Comparison and Selection

There are two premises for designing this system, one is to reduce the cost and power consumption and the other is that we want to have an extremely high performance of the speech recognition song ordering system. Therefore, we start discussing the advantages and disadvantages of the two solutions and how to choose them around the two premises.

STM32 in the first program has the mainstream Cortex core, rich and reasonable in terms of appearance and design, power consumption, price and other practicality and actual production costs are also reasonable. STM32 is also very powerful in terms of software, there are STM32 file libraries that contain comprehensive and rich technical documentation. The chip model of this microcontroller is many, covering a wide range. The chip's full functionality can perfectly meet the user's needs, and the overall view of the cost-effective is extremely high.

The second solution of the Sunplus microcontroller itself with the function of speech recognition [7], its model is SPCE061A microcontroller. From the above, it can be seen that it comes with a speech recognition module, which reduces the steps that the user has to do when setting up the speech part of the function. Since modern microcontrollers integrate more than one function, their learning cost is high, and users can learn their use in conjunction with the product information. However, due to the complexity of its functional modules, it would be more troublesome to use this microcontroller in the physical production.

This design uses a microcontroller to realize the control of intelligent voice system point song. STM32 as well as LD3320 and JQ8400 voice module use can be fully applicable to the processing and operation in this scenario. Hardware circuit connection is simple, the required devices are also easy to buy, and can be used together. In summary, it is considered more feasible to choose the first STM32-based master control scheme [8].

## 3. Design of Hardware Circuits

### (1) Microcontroller Control Module

STM32 is launched by STMicroelectronics (ST), is a cortex-m3 core 32-bit microcontroller, STM32F103C8T6 hardware using LQFP48 package, is the basic enhancement should also have 48 feet, 64k bytes flash these features. System design selection of the STM32F103C8T6 entire structure can be seen as ARM's cortex-m3 core and ST company in the two synthesized bus matrix, direct memory read DMA, AHB, APB1, APB2 mounted peripherals and so on the two parts. Fully understand the structure of the STM32 after the pins and their functions to learn, better realize the connection of the circuit and the completion of the hardware! The STM32F103C8T6 pinout is shown in the figure.





L is the output of the left channel, SPK is the part of the external speaker. The circuit schematic of the voice

announcement module is shown in Figure 6.

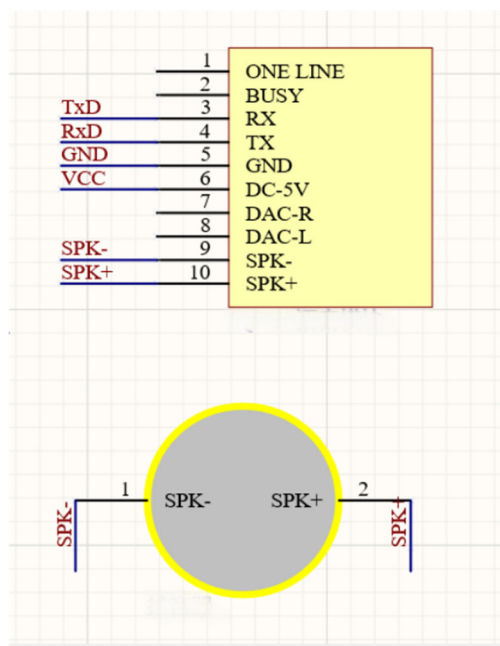


Figure 6. Circuit diagram of voice announcement module

#### (4) OLED Display Module

OLED, the same by this English abbreviation, that is, organic light-emitting diode, also known as organic laser display. OLED display module has been in the market to occupy a significant position, a variety of small mobile devices, such as MP3, MP4 and so on the player on the application of a great deal. With the more backward development, more and more large display on the TV and

computer also use OLED technology, compared to the traditional LED display, OLED display is only when the current through, so that the internal organic material luminous, the viewing angle is larger, the view and practicality is greatly enhanced. Adopts 0.96-inch SSD1306 display. Users need to pay attention to the use of the screen: OLED after powering up no indicator light or the like to remind whether the power, only when the program drive, the screen displays pre-set information about the relevant program.

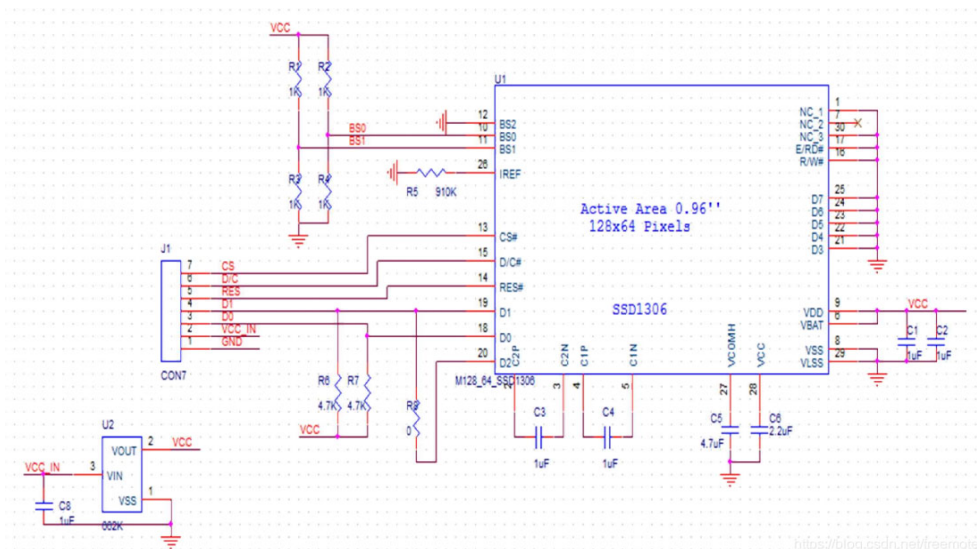


Figure 7. SSD1306 circuit schematic diagram

The SSD1306 circuit schematic is shown in the figure, and the dedicated chip in the common cathode OLED panel is the SSD1306. Since there is an internal crystal oscillator, RAM, and contrast controller, the power consumption is reduced to a very low level, and the number of externally mounted devices is also reduced. The brightness of the OLED display can be divided into 256 levels.

## 4. System Software Design

When entering the software part of the design, the main thing is the communication part. Because the functions of voice broadcasting and voice recognition both need the serial port protocol to give a combination.

For the voice recognition module, with the corresponding code, by burning into the voice recognition chip, and then the

function of this program is to go to identify the words spoken by the user, for the main control part of the STM32 microcontroller, when the information A is sent through the serial port, which is equivalent to letting the main control part of the information that the user said the song, the main control will use its own function to let the OLED screen to display, other parts are also the same, for the OLED module's communication protocol is IIC, you can build the connection

with other modules according to its definition and operation instructions. For the voice broadcast module, it is written according to the communication protocol of JQ8400. For example, AA070200AC, AA07 is the packet header, the last bit is the end, it sends to the voice module is actually the middle two, and then the voice broadcast module should know what the user sends.

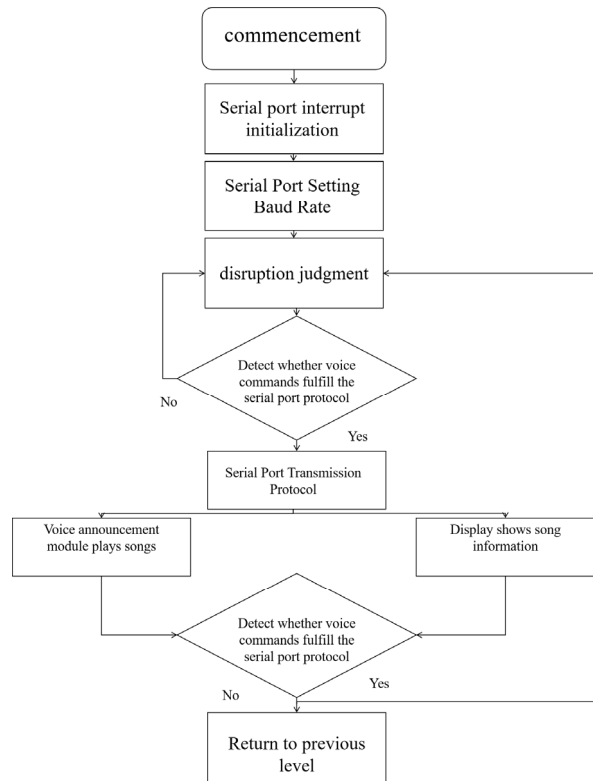


Figure 8. Software flow chart

### (1) Speech Recognition Module Software Design

The song protocol of this voice recognition song ordering system is as follows: start for A, pause for B, stop for C; the protocol for the song name is as follows: old boy for D, lone brave for E, thank you for ever being here for F, good luck for G, full of stars are you for H, the southern hemisphere and the Hokkaido for I; the definition of the song style is as follows: the style of love is for H, the style of blood is for E, the style of pleasantness is for G, the style of motivation D;

Voice Recognition Part When some information and commands of the song are recognized, the STM32 master control unit, in addition to receiving the serial protocol to get the information, also sends the serial protocol to the voice announcing and OLED display part to drive it to realize the corresponding song announcing and song information display. The corresponding information code is also written under this file. As shown in Figure 9 and Figure 10 for the command information program and its serial port communication and song information.

```

92 void USART1_IRQHandler(void)
93 {
94
95     if(USART_GetITStatus(USART1, USART_IT_TXE) != RESET)
96     {
97         USART_ClearITPendingBit(USART1, USART_IT_TXE);
98     }
99
100    if(USART_GetITStatus(USART1, USART_IT_RXNE) != RESET) {
101        if(USART1->DR) {
102            Usart1_RxBuff[Usart1_RxCounter]=USART1->DR;
103            Usart1_RxCounter ++;
104
105            Usart1_RxCounter=0;
106            //// ul_printf(Usart1_RxBuff);
107
108
109
110
111            if(Usart1_RxBuff[0] == 'A')
112            {
113                Usart1_RxCounter=0;
114                USART_SendData(USART2,0xaa); delay_ms(20); //AA 07 02 00 01 B4
115                USART_SendData(USART2,0x02); delay_ms(20);
116                USART_SendData(USART2,0x00); delay_ms(20);
117                USART_SendData(USART2,0xAC); delay_ms(20);
118                OLED_ShowChinese(0,0,35,16);
119                OLED_ShowChinese(17,0,36,16);
120                OLED_ShowChinese(17+16,0,34,16);
121                OLED_ShowChinese(17+32,0,34,16);
122                OLED_ShowChinese(17+32+16,0,34,16);
123                OLED_ShowChinese(17+32+16+16,0,34,16);
124                OLED_ShowChinese(17+32+16+16+16,0,34,16);
125            }
126        }
127    }
  
```

Figure 9. Command Message Program

```

208
209
210 if(Usart1_RxBuff[0] == 'G')//好运来
211 {
212     Usart1_RxCounter=0;
213     USART_SendData(USART2,0xaa); delay_ms(20);           //AA 07 02 00 01 B4
214     USART_SendData(USART2,0x07); delay_ms(20);
215     USART_SendData(USART2,0x02); delay_ms(20);
216     USART_SendData(USART2,0x00); delay_ms(20);
217     USART_SendData(USART2,0x03); delay_ms(20);
218     USART_SendData(USART2,0xB6); delay_ms(20);
219
220     OLED_ShowChinese(0,0,17,16);
221     OLED_ShowChinese(17,0,18,16);
222     OLED_ShowChinese(17+16,0,19,16);
223     OLED_ShowChinese(17+32,0,34,16);
224     OLED_ShowChinese(17+32+16,0,34,16);
225     OLED_ShowChinese(17+32+16+16,0,34,16);
226     OLED_ShowChinese(17+32+16+16+16,0,34,16);
227
228     OLED_ShowChinese(0,20,51,16);
229     OLED_ShowChinese(17,20,52,16);
230     OLED_ShowChinese(17+16,20,34,16);
231     OLED_ShowChinese(17+32,20,34,16);
232
233 }
234 if(Usart1_RxBuff[0] == 'H') // 满目光芒皆是你
235 {
236     Usart1_RxCounter=0;

```

Figure 10. Song Information

## (2) Software Design of Voice Announcement Module

Voice broadcast module using the JQ8400 broadcast function is very convenient, the user will broadcast the audio information in advance to copy the module, when the STM32 sends the serial port protocol, triggering the voice broadcast module, the module receives the specified protocol, go to perform the corresponding functions. Serial port 2 voice

broadcast program shown in Figure 11.

In addition to the STM32 microcontroller through the serial port 2 to send commands to the voice broadcast module, the realization of the function of the voice broadcast module also needs to be set by the user to its audio information, voice recognition playback engineering.

```

65
66 __align(8) char USART2_TxBuff[USART2_TXBUFF_SIZE];
67
68 void u2_printf(char* fmt,...)
69 {
70     unsigned int i,length;
71
72     va_list ap;
73     va_start(ap,fmt);
74     vsprintf(USART2_TxBuff,fmt,ap);
75     va_end(ap);
76
77     length=strlen((const char*)USART2_TxBuff);
78     while((USART2->SR&0X40)==0);
79     for(i = 0;i < length;i ++){
80
81         USART2->DR = USART2_TxBuff[i];
82         while((USART2->SR&0X40)==0);
83     }
84 }
85

```

Figure 11. Voice announcement program

## 5. System Debugging and Analysis

### (1) Hardware Circuit Debugging

According to how to realize the function of the two modules two by two or each other for serial communication, after designing the circuit, start in the PCB board to select the appropriate location for the insertion of the board, at the same time, the pin between the module to carry out the corresponding welding work. After the above operation, the entire circuit of the line is cumbersome and messy wiring problems, PCB board of each module of the welding problem is eliminated. After a number of functional tests, and constantly adjust the various parts of the poor contact and other issues, to determine the functions to be achieved by each independent module and the requirements of the function is consistent with the realization of the function, can be a stable realization of the function of each module. Under the control of the STM32 microcontroller for voice recognition point song function test. As the voice recognition module, voice broadcast module, and OLED display can intuitively respond to the module operation or not, as well as the red LED indicator of each module, so when the external USB interface

to provide power, the physical picture is shown in Figure 12.

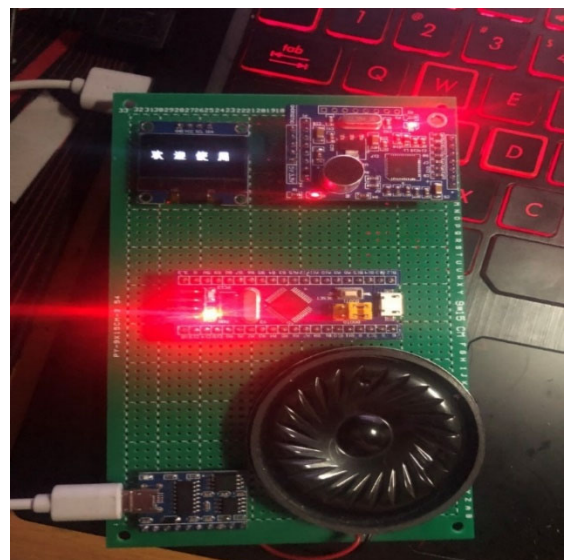


Figure 12. System Physical Diagram

(2) Software Function Debugging

Use the software KEIL5 to write the program, this project is based on the STM32 microcontroller to set up the control function of each module, so we need to create the main function, OLED display function, delay function, bit manipulation function, and then consider the voice

recognition and the module for serial communication to create two USART.C, respectively, for the operation of voice recognition and voice broadcasting. Find the STM32 microcontroller library file from STC-ISP software and complete the compilation. The environment for compilation is shown in Figure 13.

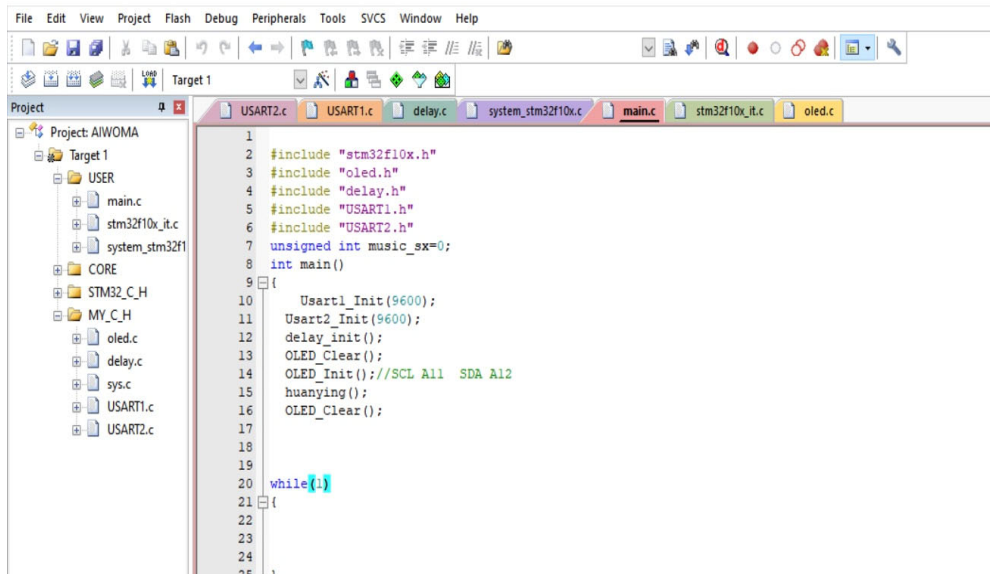


Figure 13. Program compilation environment

Need to generate HEX files, HEX files through the software downloaded to the hardware object, so that the hardware object has the function written by the software. Speech recognition part of the project file with this design is STC-ISP, this software supports now on the market by the KEIL software development projects developed by the STC series of microcontrollers developed by the project, based on the C language assembly language.

Speech recognition part of the project file, first open the STC-ISP download program, enter the interface to connect the microcontroller and the computer, to ensure a good hardware connection, set the COM port and baud rate. After all the preparations are completed, click the power switch to download.

(3) Test results

Circuit power normal, the software runs normally, the entire design of the external USB power supply, test whether its function can be realized. That is, through the voice input information, voice speak the user in advance to copy into the song track, to see if the voice pointing function can be realized, and then voice speak pause, stop, to see if the function can be realized, with the singer's name and style of the track to point to the song, to see if the function can be realized.

The physical realization of the function is shown by Fig. 14. Where Figure 14 indicates that the user orders a song with

voice input commands, and the song name and singer information are displayed on the OLED screen.

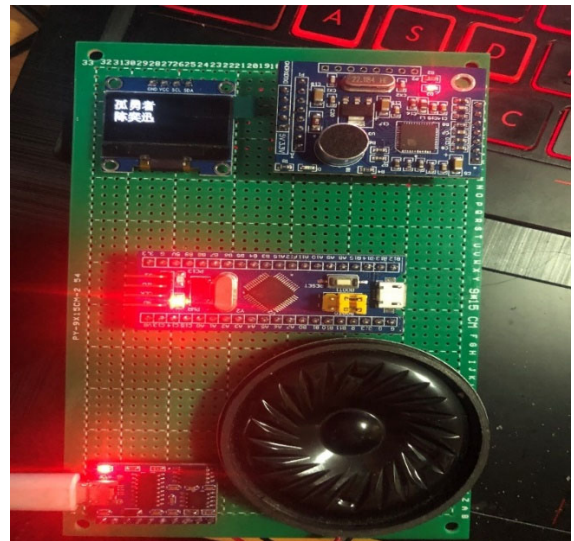


Figure 14. Function Display

The specific implementation of the function is shown in Table 1.

Table 1.

Design Requirements	Realization
After power on, the display starts to work normally showing: Welcome to use!	Yes
Supports display of song information (artist and song title)	Yes
Support voice-recorded song title ordering	Yes
Support voice-recorded singer name song ordering	Yes
Support voice-recorded song style song ordering	Yes
Supports pause, stop, etc.	Yes
Songs play clearly	Yes

## 6. Conclusion

This voice recognition based song ordering system design, STM32F103C8T6 microcontroller as the main control chip, and LD3320 voice recognition module, JQ8400 voice announcement module, OLED display module is connected to three modules, a total of four modules plus external power supply to form a common system of this design. The voice recognition based song ordering system is simple and fast, with the voice on the control of the realization of its functions, in people's usual leisure and entertainment time, used to listen to songs to relax, very intelligent.

## References

- [1] Hong Q. Yang et al. Speech recognition principles and applications [M]. Beijing: Electronic Industry Press, 2020.
- [2] Han J. et al. Embedded system design and application [M]. Wuhan: Huazhong University of Science and Technology Press, 2019.
- [3] P.Y. Wan, M.S. Wu, R.S. Liu. Speech Recognition and Speech Prompting for Monitoring Systems [J]. Electronic Technology Application, 1999(10):17-18.
- [4] HouChenxiang. Application of Sunplus microcontroller in KTV voice song ordering system[J]. Heilongjiang Science and Technology Information, 2014(01):167.
- [5] G. Li, H. F. Li. Content-based audio retrieval: concepts and methods[J]. Small Microcomputer Systems, 2000(11):1173-1177.
- [6] Liu L. et al. STC microcontroller project example tutorial [M]. Wuhan: Huazhong University of Science and Technology Press, 2019.
- [7] Wang, X. TensorFlow speech recognition in practice [M]. Beijing Tsinghua University Press, 2021.
- [8] Y. Li, L. Q. Cai, G. Y. Tan, H. Chen, C. Chiu. Voice control of home devices based on STM32 microcontroller[J]. Technology and Innovation, 2018(05):130-131. DOI:10.15913/j.2018.05.130.