

WTK6900B01 Speech Recognition Module

Note:

WAYTRONIC ELECTRONIC CO.,LTD. reserves the right to change this document without prior notice. Information provided by WAYTRONIC is believed to be accurate and reliable. However, WAYTRONIC makes no warranty for any errors which may appear in this document. Contact WAYTRONIC to obtain the latest version of device specifications before placing your orders. No responsibility is assumed by WAYTRONIC for any infringement of patent or other rights of third parties which may result from its use. In addition, WAYTRONIC products are not authorized for use as critical components in life support devices/systems or aviation devices/systems, where a malfunction or failure of the product may reasonably be expected to result in significant injury to the user, without the express written approval of WAYTRONIC.

1. Features

- Voice recognition rate up to 90% within half a meter.
- Control output through voice command.
- Play the specified voice through MCU serial communication control.
- Three output modes: serial output, IO output, speaker output.
- PWM output can directly drive 0.5W/8Ω or 1W/8Ω speaker.
- Differential amplifier circuit.
- Voice recognition distance up to 2~3m in a quiet environment.
- Can add SPI flash to increase the storage capacity.
- Can store many groups of words.
- Speaker-independent recognition.
- Support 32 languages, such as Chinese, English, Cantonese, Japanese etc.

2. Function description

WTK6900B01 voice module is a voice recognition module, small size, voice recognition distance up to 2~3m in a quiet environment. You can add SPI Flash to increase memory size. Following are the main functions:

- Voice recognition function: Can identify the default speech words and phrases, output both through serial port and IO port, identify the result.
- Voice play function: Through MCU instructions, play the specified voice, at the same time, through the BUSY pin mark, judge whether there is voice play.

3. Application occasions:

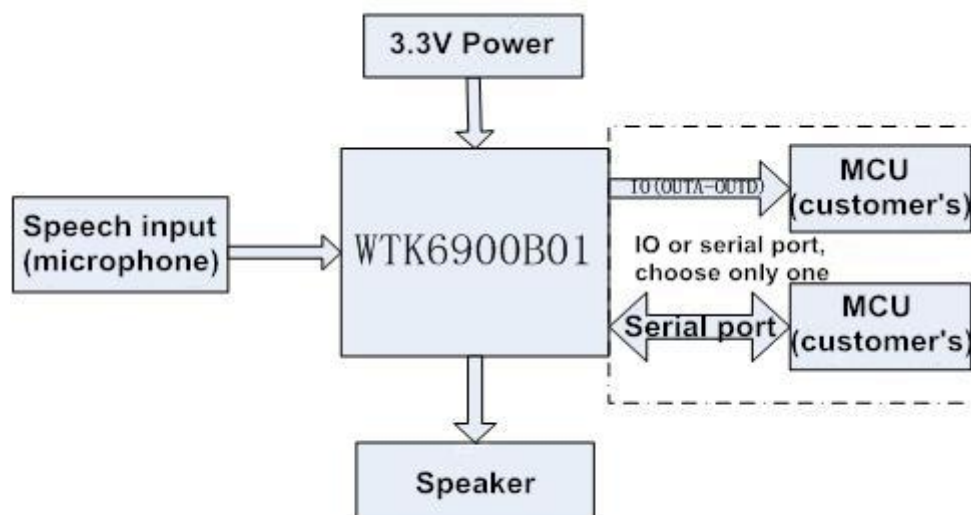
High-end toys: parrot talk, story teller, music player, etc.

Switches: light switch, brightness adjustment, etc.

Voice broadcast: prompt tone for small household appliances.

Voice navigation: giving directions in hospital, shopping mall, etc.

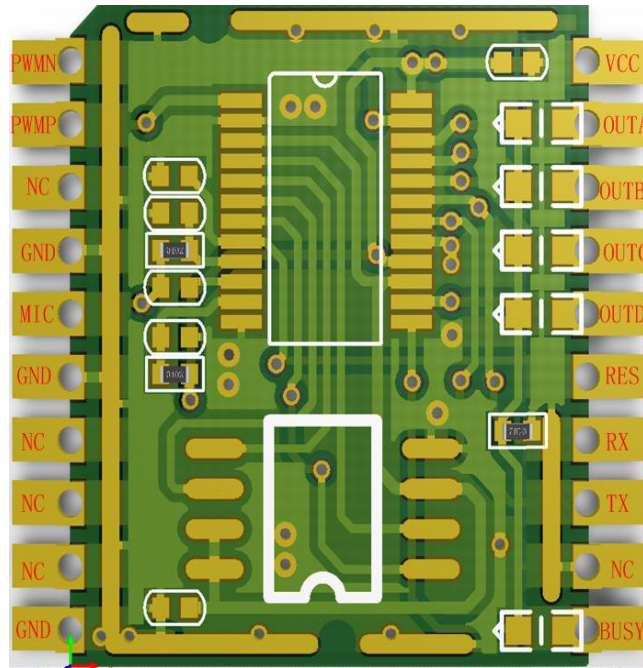
4. Block diagram





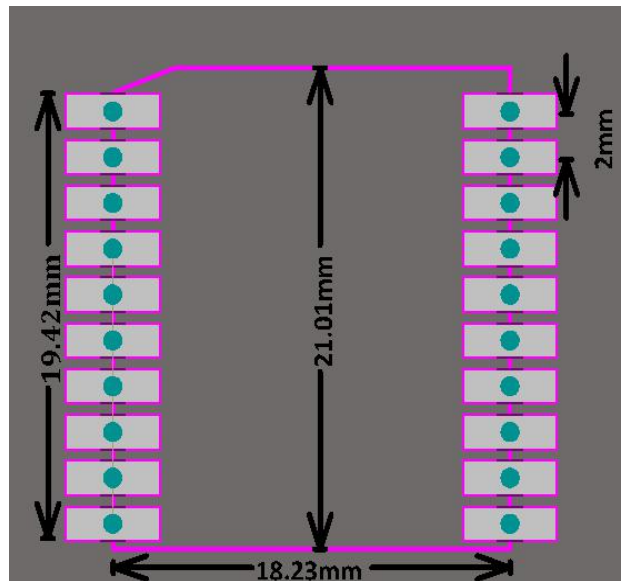
Internal structure of module: WTK6900A-24SS speech recognition chip, SPI-flash; Can match with the microphone, speaker, MCU or other peripherals. In the dashed box, serial port and IO are used to detect WTK6900B01 module whether can correctly identify, so when connecting (customer's) MCU, can connect the IO port or serial port, choose one.

5. Module pin introduction



Pin	Pin name	Description
1	PWMN	Digital PWM output(-), can directly drive the speaker if used with PWMP
2	PWMP	Digital PWM output(+), can directly drive the speaker if used with PWMN
3	NC	Null (reserved)
4	GND	Digital Ground
5	MIC	Microphone positive input
6	GND	Digital Ground
7	NC	Null (reserved)
8	NC	Null (reserved)
9	NC	Null (reserved)
10	GND	Digital Ground
11	BUSY	BUSY pin, used to prompt whether voice is playing
12	NC	Null (reserved)
13	TX	Serial port, can connect with RX pin of MCU
14	RX	Serial port, can connect with TX pin of MCU
15	RES	Reset pin(no need to connect, reserved)
16	OUTD	IO port
17	OUTC	IO port
18	OUTB	IO port
19	OUTA	IO port
20	VCC	Power input

6. Module size



7. Electrical parameter

Symbol	Parameter	Min	Max	Type	Unit
VCC	Working voltage	1.8	5.5	3.3	V
I _{halt1}	Working current	--	--	7	mA
T _a	Working temperature	-20	75	--	°C
IO	Output current	-16	16	--	mA

8. Function

8.1 IO port function

- RXD/TXD standard serial communication, baud rate: 9600, as shown below.
- No. 11 pin is BUSY indication, high level when play, low level when not play.
- MIC Microphone positive input, can connect with common microphone
- PWMP/PWMN pwm output, can directly drive 0.5W/8Ω or 1W/8Ω speaker



Port configuration

8.2 Command format

Command format of serial port: start code+parameter+inverse code, like below:

Start code	Parameter	Inverse code
AA/BB	See below	See below

Note: ① parameter+inverse code = FF. For example, the command of TX output is “AA 01 FE”. Start code is AA; parameter is 01; inverse code is FE. 01+FE=FF; other data is also by such analogy.

② Start code of TX output command is AA; start code of RX input command is BB.

8.3 Introduction of vocabulary entry

WTK6900B01 speech recognition module can identify the fixed vocabulary entry, speaker-independent recognition.

For the fixed vocabulary entry, you can make changes. You can replace the words according to the demand, stored to WTK6900A-24SS voice chip storage or external SPI Flash. Voice recognition will be successful when the similarity of input speech and stored words is up to the standard.

8.3.1 The memory capacity of vocabulary entry

There are two choices for you to store the vocabulary entry: voice chip inside, SPI Flash. These two has different capacity, so how many vocabulary entries can be stored, that's naturally different. Following is a storage capacity comparison of four-word vocabulary entry (four words in each vocabulary entry).

Storage location	The number of vocabulary entry
WTK6900A-24SS voice chip	About 20 pieces
SPI Flash	Depends on Flash capacity

Although WTK6900A-24SS voice chip can store 20 pieces of vocabulary entry inside, more it store, lower operating rate is, and bad recognition effect.

8.3.2 Storage way of vocabulary entry

There are two ways:

① Stored in one group

If the number of vocabulary entry is less, and vocabulary entry is in the same type, such as the color words, that can be selected in the same group, no need to switch, more saving time.

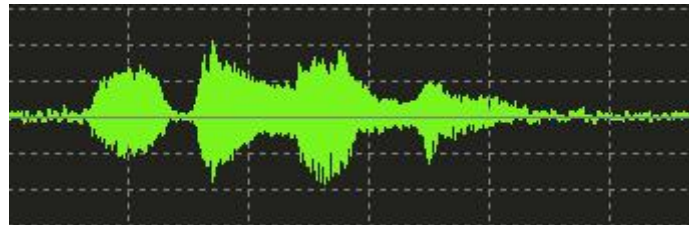
② Stored by group

If there are many vocabulary entries, and in different types, you can store them into groups. For example, there are three groups of different vocabulary entries: stories, ancient poetry, music.

Stories	Ancient poetry	Music
The daughter of the sea	Pastoral Poem	Descendants of the Dragon
Harry Potter	The Pagoda on the Fly-in Peak	Because Of Love
Snow White	The Pond After the Rain	I Cant Help It
The Wolf and the Lamb	This Summer	Sole Memories

8.3.3 The process of voice recognition

The recognition process is waveform. In fact, the sound will produce the waveform, shown below. No matter it is human voice or the voice of the fan, there are waves, but if different people say the same words, sound waves will have a little difference. Even the same person said the same words, sound waves will be different, because tone, speed and so on may have little changes, which influence the sound waveform.



In the design process, set a threshold. If the input speech waveform and the stored entry waveform has a higher similarity than the threshold, the recognition will be successful. The threshold level can be changed (user cannot change). The threshold increased, the similarity higher, it will be more difficult to identify, the recognition rate reduced; if the threshold is lower, it will be easier to identify, recognition rate increased, but there is also a problem, that is error identification.

In a noisy environment, the waveform of many voices superimposing together, waveform similarity of one entry in the storage exceeds the threshold value, then it will identify it as the entry in the storage, and speech recognition success, which causes the false identification. If set the threshold lower, error recognition rate will be higher.

Therefore, the recognition rate and the false recognition rate are consistent. The recognition rate is increased, the rate of false recognition will be increased, so design recognition rate according to the occasion requirements. Of course, false recognition rate can be reduced. For example, when you do not need to identify, make module in a sleep state, not only power saving, but also reduce the rate of false recognition.

8.3.4 The design of vocabulary entry

Recognition accuracy is not only related to the recognition threshold, but also related with the length of vocabulary entry.

Vocabulary entry	Recognition effect
Open	So short, uneasy to identify
Light on	Recognition effect not so good
Light off	Similar with other phrase, recognition effect not so good
Aladdin, turn on the light	Moderate length, low similarity, good recognition effect
Tired, turn off the light	Moderate length, low similarity, good recognition effect

Therefore, when you replacing the vocabulary entry, please consider the length, similarity. Length should be moderate; similarity should be low.

8.3.5 The operation after recognition success

WTK6900B01 speech recognition module has three output modes: serial output, IO output, speaker output. You can choose how many output mode you need, to check whether the recognition is successful. What the speaker plays can be changed according to your needs, but you cannot change it by yourself.



8.4 Recognition function introduction

Input entries from the microphone (that is voice command), and stored in SPI Flash, comparing vocabulary entry. If consistent, that is correct, then the serial port TX will output the corresponding instruction, and the three IO ports (OUTA, OUTB, OUTC) will output the corresponding level, at the same time the speaker playing.

Command of IO terminal	OUTC	OUTB	OUTA	TX output	Implementing result
None	0	0	0		
Pearl light red	0	0	1	AA 01 FE	The 1st light on, voice play "Pearl light red"
Sky blue	0	1	0	AA 02 FD	The 2nd light on, voice play "Sky blue"
Lavender purple	0	1	1	AA 03 FC	The 1st and 2nd light on, voice play "Lavender purple"
Peppermint green	1	0	0	AA 04 FB	The 3rd light on, voice play "Peppermint green"
Cherry loose powder	1	0	1	AA 05 FA	The 1st and 3rd light on, voice play "Cherry loose powder"
Jasmine white	1	1	0	AA 06 F9	The 2nd and 3rd light on, voice play "Jasmine white"
Please turn on the light	1	1	1	AA 07 F8	Three lights on, voice play "all lights on"
Please turn off the light	0	0	0	AA 08 F7	Three lights not on, voice play "all lights off"

Note: "1" means high level, "0" means low level.

8.5 Play function introduction

Play function is achieved together with the MCU(user's). Through MCU, play the specified voice (this function can be achieved if only use the module).

After power on, MCU (user's) send commands the module through the serial port, play the audio of SPI Flash. If the play is successful, voice recognition module output "01" through the serial port. If the input error, unable to play, it will output "02", and will pull up the BUSY pin when playing. When the play is finished or unable to play, the BUSY pin will be pulled low. Therefore, according to the BUSY pin, you can calculate voice playback time.



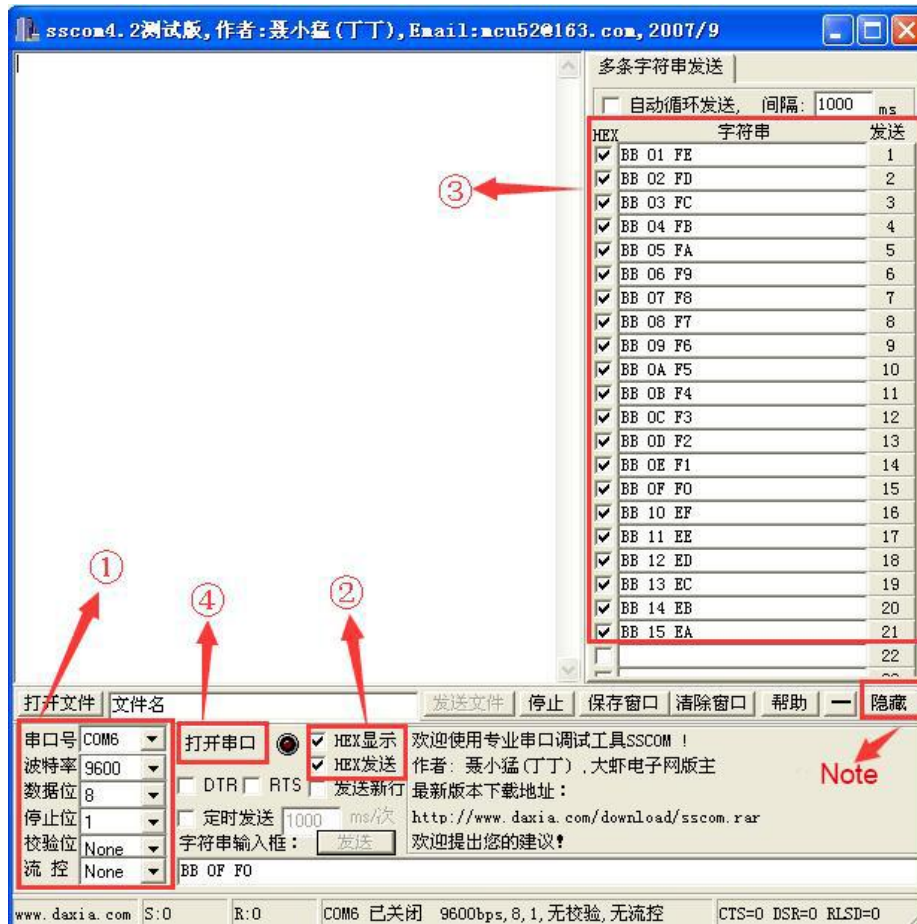
Note: "1" means high level, "0" means low level.

RX input	Play the audio of SPI Flash (the audio in SPI Flash can be modified)	Play success (TX output return command)	Input error Cannot play (TX output return command)	BUSY (when playing)	BUSY (playing ends)
BB 01 FE	1 st audio in SPI Flash	01	02	1	0
BB 02 FD	2 nd audio in SPI Flash	01	02	1	0
BB 03 FC	3 rd audio in SPI Flash	01	02	1	0
BB 04 FB	4 th audio in SPI Flash	01	02	1	0
BB 05 FA	5 th audio in SPI Flash	01	02	1	0
BB 06 F9	6 th audio in SPI Flash	01	02	1	0
BB 07 F8	7 th audio in SPI Flash	01	02	1	0
BB 08 F7	8 th audio in SPI Flash	01	02	1	0
BB 09 F6	9 th audio in SPI Flash	01	02	1	0
BB 0A F5	10 th audio in SPI Flash	01	02	1	0
BB 0B F4	11 th audio in SPI Flash	01	02	1	0
BB 0C F3	12 th audio in SPI Flash	01	02	1	0
BB 0D F2	13 th audio in SPI Flash	01	02	1	0
BB 0E F1	14 th audio in SPI Flash	01	02	1	0
BB 0F F0	15 th audio in SPI Flash	01	02	1	0
BB 10 EF	16 th audio in SPI Flash	01	02	1	0
BB 11 EE	17 th audio in SPI Flash	01	02	1	0
BB 12 ED	18 th audio in SPI Flash	01	02	1	0
BB 13 EC	19 th audio in SPI Flash	01	02	1	0
BB 14 EB	20 th audio in SPI Flash	01	02	1	0
BB 15 EA	21 th audio in SPI Flash	01	02	1	0

9. Typical application circuit

Microphone access, used as voice input terminal, speaker, serial port, IO are output terminal.

- When need to play voice, add speaker.
- WTK6900B01 voice recognition module can be connected with user's MCU, can connect to the IO port of MCU, can also connect to serial port of MCU, but choose only one of them.



Note: if open the software and not appear the input box ③, please click "Hide", then turn into "Extend" 扩展 click it, turn into "Hide" 隐藏.

- Parameter setting in ①:
Serial port number is the current port in use, (user needs to set the parameter based on the actual port).
Baud rate: factory default value is 9600, if the customer has modified before, use the modified value.
Data bit: 8 Stop bit: 1 Check bit: None Flow Control: None
- After the setting, check ② HEX display, HEX send. Check whether current serial port is not occupied by other device, click ④ "open serial port"
- The commands in ③ are responding to the RX input command in play function. You can directly copy the commands in RX input column, refer to the above "8.5 Play function introduction", tick HEX options, click the numbers behind the command, then can send this command.

10.3 How to use

According to the wiring steps, serial port settings, then can be used (power on, into the working mode).

- If you want speech recognition, just read out the words to the microphone, such as "pearl light red". When the speech recognition is correct, IC will play voice; electric level of IO port changes; serial port will output the corresponding commands at the same time. For more recognition results, please refer to "8.4 Recognition function introduction".
- Through MCU control module, to play the corresponding voice, send commands through serial port, it will play the corresponding voice according to the commands.



10.4 Notes

- 1) Environment: quiet environment
- 2) Distance: the best effect is in half a meter, the farther from the module, the recognition rate will be decreased; the farther the distance, the more difficult to spread the sound to the microphone.
- 3) Speak into the microphone, so microphone can receive the sound better; the recognition effect will be better.
- 4) When reading the entry, do not need to put your head so close to the microphone.
- 5) Speaking speed should be moderate, not so fast, not too slow, in a normal speaking speed.
- 6) It is best to leave the module in the air. If the weld of microphone is pressed with hand or pressed on the conductive object, it will recognize there is waveform input, affecting the recognition rate.
- 7) If connect a speaker to play voice, and you want voice louder, more clearly, need to connect the cavity, to prevent the voice spread, the speaker can be placed in a small box, or wrap around the speaker with your hand.