Speech Synthesis module communication protocol

1.Communication protocol

The module communication mode is I2C communication, the module I2C address is 0x50, and all data frames are sent through I2C communication.

2. Voice broadcast

2.1 Voice synthesis support control commands

Sent to the host computer XFS5152CE All speech synthesis chip command need to be packaged after transmission 'frame" mode.

The frame structure is composed of three parts: the frame header mark, the length of the data area and the data area.

Header	The length of the data area	Data Area
0xFD	0xXX , 0xXX	Data

The following is the description of the command frame.

The frame header and the end are fixed at 0xFD, and the data length is divided into two bytes to send. Finally, the data area is less than 4k in size.

name	length	Explanation
Header	1 Byte	Defined as hexadecimal " 0xFD "
The length of the data area	2 Bytes	Is two bytes, high byte first, followed by the low byte
Data Area	Less than 4k Bytes	Command word and command parameters, and the length of "data field length"

The data area is composed of command words and command parameters.

The upper computer uses the command words to realize various functions of the voice broadcast chip.

Different command words have corresponding command parameters.

As shown below.

name	Data transmission	Explanation
Command word	0x01	Speech synthesis command
	0x02	Stop synthesis command, no arguments
	0x03	Synthesis pause command, no arguments
	0x04	Restore synthesis command, no arguments
	0x21	Chip status inquiry command
	0x88	The chip into power-saving mode
	0xFF	Chip returns to normal operation from the power saving mode
Command parameters	Different command words have different parameter lists, see the explanation of each command word	

2.2 Speech synthesis command

The speech synthesis module uses I2C to fix the frame header to 0xFD.

The length of the data area is the length of the combined text + 2 (the command word and the text format are one byte).
The text length represents the byte length of the text.
For example, when the text encoding is GB2312, the length of the string "123" is 3, and the length of the string "Hello" is 4, because a Chinese character in GB2312 Chinese is 2 bytes length.
Ordinary English and numbers are 1 byte.

<table>
<tr><th>Name</th><th>Data sent</th><th colspan="6">Explanation</th></tr>
<tr><td>Command word</td><td>0x01</td><td colspan="6">Text with text encoding settings of the play command</td></tr>
<tr><td rowspan="6">Parameter list</td><td rowspan="5">0Xxx</td><td colspan="2" rowspan="5">1 Byte Text encoding format.
The value is o~3</td><td>Parameter value</td><td colspan="2">Text encoding format</td></tr>
<tr><td>0x00</td><td colspan="2">GB2312</td></tr>
<tr><td>0x01</td><td colspan="2">GBK</td></tr>
<tr><td>0x02</td><td colspan="2">BIG5</td></tr>
<tr><td>0x03</td><td colspan="2">UNICODE</td></tr>
<tr><td>Data</td><td colspan="5">The binary content of the text to be synthesized</td></tr>
<tr><td rowspan="3">Command frame format structure</td><td>Frame header</td><td colspan="2">Data area length</td><td colspan="3">Data area</td></tr>
<tr><td rowspan="2">0xFD</td><td>High byte</td><td>Low byte</td><td>Command word</td><td>Text encoding format</td><td>text to be synthesized</td></tr>
<tr><td>0xHH</td><td>0xLL</td><td>0x01</td><td>0x00~0x03</td><td>.....</td></tr>
</table>

Tips:
When the speech synthesis module is synthesizing text, if it receives another valid synthesis command frame, the chip will immediately stop the text currently being synthesized, and instead synthesize the newly received text.
In order to avoid this problem, we need to use the command to query the chip status and send a new command frame only when the chip is detected to be idle.

2.3 Synthesis stop command

Sending the following four data to the module through I2C can stop the module from synthesizing.
In order to avoid continuously sending data too fast, the internal data processing is not timely enough, which leads to errors.
We should send a data between a start bit and a stop bit, and leave a little time.

Name	Data sent	Explanation	
Command word	0x02	Stop synthesis	
Parameter list	No		
Command	Frame header	Data area length	Data area

frame format structure	0xFD	High byte	Low byte	Command word
		0x00	0x01	0x02

2.4 Synthesis pause command

Name	Data sent	Explanation		
Command word	0x03	Pause synthesis		
Parameter list	No			
Command frame format structure	Frame header	Data area length		Data area
	0xFD	High byte	Low byte	Command word
		0x00	0x01	0x03

2.5 Synthesis recover command

Name	Data sent	Explanation		
Command word	0x04	Pause synthesis		
Parameter list	No			
Command frame format structure	Frame header	Data area length		Data area
	0xFD	High byte	Low byte	Command word
		0x00	0x01	0x04

2.6 Read chip status

Send the following 4 data to the module through I2C communication, and then read one byte through I2C communication to get the chip status value.
In order to avoid continuously sending data too fast, the internal data processing is not timely enough, which leads to errors.
We should send a data between a start bit and a stop bit, and leave a little time.
When reading, a complete I2C sequence is required, that is S, I2CADDR+1 (read), ACK, DATE, NACK, and P.
The default state value is 0xff.

Name	Data sent	Explanation	
Command word	0x21	Obtain the corresponding parameters through this command to determine whether the TTS chip is in a synthetic state. If it return 0x4E indicates that the chip is still being synthesized; if it return 0x4F indicates that the chip is In idle state.	
Parameter list	No		
Command	Frame header	Data area length	Data area

frame format structure	0xFD	High byte	Low byte	Command word
		0x00	0x01	0x21

Except for the default value 0xff, other chip status values are shown in the following table.

Return data type	Return data	Triggering conditions
Initialization successful	0x4A	Chip Initialization successful
Receive the correct command frame	0x41	Receive the correct command frame
Receive an error command frame	0x45	Receive an incorrect command frame
Chip busy	0x4E	Receive a "status inquiry command", the chip is synthesized text state return 0x4E
Chip idle	0x4F	When a data combining finished, return to the idle state chip Ox4F ; When the chip receive a "status inquiry command", the chip is idle return 0x4F.

3. Text control mark

3.1 Text control method using labeled

Speech synthesis function supports a variety of text control marks, which can set the speaker, volume, speaking rate, and intonation of the speech synthesis.

In generally, the format of the text control mark is"[]".
For example, [m3]. The use of the mark is exactly the same as the composite text.

The user can send the mark as a text to the chip separately.
The mark is only used as a control mark to realize the setting function.

3.2 Text control list of tags(Just for English)

Function	Control	Explanation	chip
Synthesis of style settings	[f?]	?=0, Stressing each syllable style	[f1]
		?=1, Synthesis normally.	
Synthetic language settings	[g?]	?=0, Automatically determine the language	[g1]

		?=1, Arabic numerals, units of measure, special symbols, Chinese synthetic	
		?=2, Arabic numerals, units of measure, special symbols, English synthesis	
Set word pronunciation way	[h?]	?=0, Automatically determine the pronunciation of the word	[h1]
		?=1, Letters pronunciation way	
		?=2, Word pronunciation way	
Select speaker	[m?]	?=3, Set pronunciation artificial Xiaoyan (female,recommended speakers)	[m3]
		?=51, Set pronunciation artificial Xujiu(male,recommended speakers)	
		?=52, Set pronunciation artificial Xuduo(male)	
		?=53, Set pronunciation artificial Xiaoping (female)	
		?=54, Set pronunciation artificial Donald Duck	
		?=55, Set pronunciation artificial XuXiaobao (kids)	
Set the digital processing policy	[n?]	?=0, Automatically determine	[n0]
		?=1, For the number of digital processing	
		?=2, Digital value for processing	
" 0 " pronunciation method	[o?]	?=0, Read “zero”	[o0]
		?=1, Read Read as "o”	
Synthesis process stalled for some time	[p?]	? Means that Unsigned integer indicating the length of a pause time in milliseconds (ms)	
Speed setting	[s?]	Set speaking speed (0~10)	[s5]
Intonation setting	[t?]	Set speaking intonation (0~10)	[t5]
Volume setting	[v?]	Set speaking volume (0~10)	[v5]
Restore the default parameters of synthesis	[d]	All settings will become the default values	

Note the use of text control marks:

1. All control marks are half-width characters.
2. The control mark needs to be sent in the format of the voice broadcast command, and the control mark is synthesized as text, that is, the synthesis command is in the format of "frame header + data area length + synthesized command word + text

encoding format + control mark text".
3. The control mark is a global control mark.
4. When the chip is powered off or reset, the original set logo will lose its effect, and the chip will restore all default values.