

广州唯创电子有限公司

Guangzhou Waytronic Electronic Co., Ltd

WT3000T8 语音合成芯片使用文档

V1.01

免责声明:

广州唯创电子有限公司申明: 说明书以官网资料为准, 如若资料内容有更新, 不会一一进行通知。如若使用 IC 时导致侵犯到第三方专利或其他权利, 不承担任何责任。如若使用我司 IC, 在航空卫星军事设备, 人身安全等领域, 造成了重大财产损失或生命伤害, 甚至生命死亡, 我司不承担任何责任。

目 录

版本更新	1
1. 产品简介	2
2. 产品特点	2
3. 引脚描述	2
3.1. 封装管脚图	2
4. 控制命令	错误！未定义书签。
4.1. 固定语音播放功能	错误！未定义书签。
4.2. 语音合成播放功能	错误！未定义书签。
5. 语音合成功能使用方法	4
5.1. 简单调用方式	4
5.2. 标准调用方式	4
5.3. 查询芯片工作状态的方法	错误！未定义书签。
6. 串口通信说明	4
6.1. 协议命令格式	4
6.2. 命令码对应表	5
7. 文本控制标记	5
7.1. 文本控制标记的使用方法	5
7.2. 文本控制标记列表	6
7.3. 控制标记使用示例	6
7.3.1. [m?]标记的使用	6
7.3.2. [s?]标记的使用	6
7.3.3. [t?]标记的使用	6
7.3.4. [v?]标记的使用	6
7.3.5. [=?]标记的使用	7
7.4. 固定语音功能	7
7.5. 发码举例	7
8. 附录	7
8.1. UNICODE 编码简介	7

版本更新

版本号	修改说明	修改日期
V1.00	原始版本	2023-12-26
V1.01	定制版本	2024-01-30

1. 产品简介

WT3000T8 是一款功能强大的高品质语音芯片，采用了高性能 32 位处理器、最高频率可达 240MHz。WT3000T8 高集成度的语音合成芯片，可实现中文、英文语音合成；并集成了语音编码、解码功能，可支持用户进行语音合成和语音播放，具有低成本、低功耗、高可靠性、通用性强等特点，现有 WT3000T8-32N QFN32（体积小 4*4MM）的封装芯片。

2. 产品特点

- 控制方式：UART，默认波特率 9600；
- 上电默认不播放；具备 BUSY 状态指示、BUSY 播放时为低电平；
- 音频输出方式，样品默认 DAC 输出；
- 支持语音高品质音频格式，（8kbps~320kbps）声音优美，.MP3、.WAV 格式；
- 最大可以支持 128Mbit 的 Flash；
- 大功率 IO 驱动能力，最高可直接驱动 32mA；

芯片支持任意中文文本的合成，可以采用 UNICODE 编码方式。每次合成的文本量最多可达 512 字节。

芯片对文本进行分析，对常见的数字、号码、时间、日期、度量衡符号等格式的文本，芯片能够根据内置的文本匹配规则进行正确的识别和处理；对一般多音字也可以依据其语境正确判断读法；另外针对同时有中文和英文的文本，可实现中英文混读。

- 支持语音解码功能，用户可以使用芯片直接播放音频文件
- 支持多种控制命令

如合成文本、停止合成、暂停合成、恢复合成、状态查询、进入省电模式、唤醒等。

控制器通过通讯接口发送控制命令可以对芯片进行相应的控制。芯片的控制命令非常简单易用，例如：芯片可通过统一的“合成命令”接口播放提示音和中文文本，还可以通过标记文本实现对合成的参数设置。

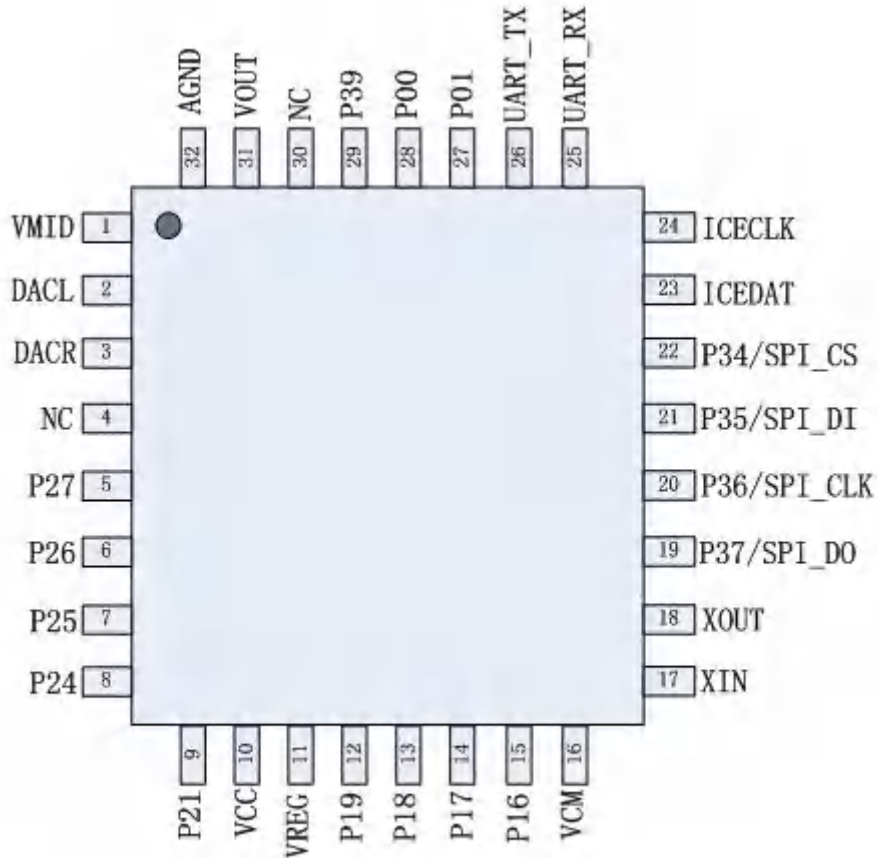
- 支持多种方式查询芯片的工作状态

包括：查询状态管脚电平、通过读芯片自动返回的工作状态字、发送查询命令获得芯片工作状态的回传数据。

- 单芯片使用（使用内置容量）时,内置语音需出厂前写入。

3. 引脚描述

3.1. 封装管脚图



WT3000T8-32N

PIN-NO.	名称	功能说明
0	GND	散热地
1	VMID	连接一个1uF 外部电容到地
2	DACL	DAC 左声道音频输出
3	DACR	DAC 右声道音频输出
4	NC	保留此引脚悬空
5	P27	IO 口
6	P26	IO 口
7	P25	IO 口
8	P24	IO 口
9	P21	IO 口
10	VCC	电源输入
11	VREG	接一个1uf 电容到地
12	P19	IO 口
13	P18	IO 口
14	P17	IO 口
15	P16	IO 口
16	VCM	参考电压输出,接一个0.1uF 外部电容到地
17	XIN	无

18	XOUT	无
19	P37/SPI_DO	SPI Flash 数据输出
20	P36/SPI_CLK	SPI Flash 时钟
21	P35/SPI_DI	SPI Flash 数据输入
22	P34/SPI_CS	SPI Flash 片选
23	ICEDAT	烧录口
24	ICECLK	烧录口
25	UART_RX	UART_RX, 通信端口
26	UART_TX	UART_TX, 通信端口
27	P01	Busy 显示
28	P00	IO 口
29	P39	IO 口
30	NC	保留此引脚悬空
31	VOUT	IO 电源3.3V
32	AGND	模拟地

4. 语音合成功能使用方法

4.1. 简单调用方式

简单调用是指用户不用关心芯片当前的工作状态，只需要发送文本到芯片合成为语音输出。

在简单调用情况下，上位机只要与芯片之间建立起 UART 通信方式，即可发送合成命令来实现文本的合成，上位机不需要判断芯片的回传数据或状态引脚的输出。

注意:如前一帧文本还没有合成完，就再次发送文本到芯片会中断前次合成，而执行新的合成。

4.2. 标准调用方式

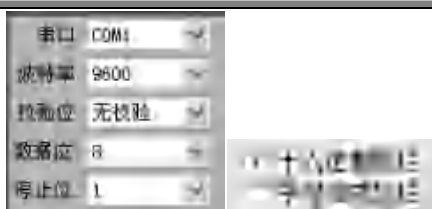
若上位机需要确保上次文本被完整合成之后，再发送合成命令帧合成下一段文本，则需要通过回传确定芯片的工作状态。具体方法举例如下：应用中需要合成的文本为 5k 字节，超过了 WT3000T8 芯片一个命令帧所能容纳的最大文本长度（512 字节），这时需要分多次给 WT3000T8 芯片发送文本信息。程序过程如下：

- 1、上位机先给 WT3000T8 芯片发送一个文本合成命令帧，携带不超过 512 个字节的文本；
- 2、上位机等待 WT3000T8 芯片自动返回的回传信息，直到收到回传数据，说明前面的文本已合成完毕；或使用查询芯片的状态引脚、发送查询命令的方法，确认上一帧文本合成完毕。
- 3、上位机向芯片再次发送一个文本合成命令帧，发送出剩下的文本。

5. 串口通信说明

5.1. 协议命令格式

芯片内置标准 UART 异步串口接口，默认波特率 **9600**，属于 3.3V TTL 电平接口。通讯数据格式是：起始位：1 位；数据位：8 位；奇偶位：无；停止位：1 位。使用电脑串口调试助手，需要正确设置串口的参数，设置如图：



指令格式

起始码	帧长度	流水号	应答标志	数据帧来源	N 个 KEYID 信息	累加和校验	结束码
7E	见下文	见下文	见下文	见下文	见下文	见下文	EF

注意：帧长度：2 字节，指起始码+帧长度+命令码+N 个 KEYID 信息+累加和校验和+结束码的长度，帧长度低位在前高位在后；

流水号：1 字节，每次一帧数据自动加 1，避免接收重复的数据，相同流水号的数据为重复数据应做丢弃；

应答标志：1 字节，固定填 00；

数据帧来源：1 字节，02 为 TTS 芯片端，03 为 MCU 芯片端；

N 个 KEYID 信息：由 N 个 KEYID 信息组成，1 个 KEYID 内容为 2 字节 ID 地址+1 字节数据长度+N 字节数据，单个 KEYID 信息最大支持 255 个字节数据，但支持同时传递多个相同的 KEYID 携带不同信息；

累加和校验：是指帧长度+流水号+应答标志+数据帧来源+参数的累加和的低字节。

5.2. 命令码对应表

基础播放指令：

KEYID	对应功能	数据内容	备注
03E8	TTS 合成	(输入合成的具体内容)	数据格式默认为 GBK 编码
03EA	TTS 播放编码	01: ASCII; 02: GBK; 03: BIG5; 04: UTF16LE; 05: UTF16BE; 06: UTF8	
03ED	TTS 停止合成	无, 固定 00	
03EE	芯片休眠	无, 固定 00	
0001	固定音播放	01+曲目地址	曲目地址范围: 0001-0008, 均为提示音效
00CD	音量加	无, 固定 00	六级音量可调, 默认音量是最大音量
00CE	音量减	无, 固定 00	

6. 文本控制标记

6.1. 文本控制标记的使用方法

WT3000T8 芯片的语音合成功能支持多种文本控制标记，可以满足用户对语音合成发音人、音量、语速、语调等的设置。

文本控制标记的格式一般是半角中括号（即“[]”）内一个小写字母、一个阿拉伯数字，如：[m3]，标记的使用方法和合成文本完全一致。

用户可以把标记作为文本单独发送到芯片上，如：只发送“[v3]”到芯片上设置合成音量为 3 级，或者把标记和其他要合成的文本放在一起发送给芯片上，如：“[v3]我在小声说话，[v10]我在大声说话”。

标记只是作为控制标记实现设置功能，不会合成为声音输出。如：“[s1]我慢条斯理。[s8]我快言快语”中，经过标记的设置，前一句合成语速会很慢，后一句合成语速会很快，但不

会读出“s1”和“s8”。

6.2. 文本控制标记列表

作用	控制标识	详细说明	芯片默认
选择发音人	[m?]	?为发音人，取值 1~57	[m2]
设置语速	[s?]	?为语速值，取值：0~10	[s5]
设置语调	[t?]	?为语调值，取值：0~10	[t5]
设置音量	[v?]	?为音量值，取值：0~10	[v5]
为单个汉字强制指定拼音	[=?]	?为标记前一个汉字的拼音+声调（1~5 分别表示阴平，阳平，上声，去声和轻声）5 个声调。 例如：“着[=zhuo2]手”，“着”字读作“zhuó”	

注意：

1、所有的控制标识均为半角字符。

2、控制标识需要按照语音合成命令的格式发送，控制标识作为文本进行合成，即合成命令是“帧头+数据区长度+合成命令字+文本编码格式+控制标记文本”的格式。

6.3. 控制标记使用示例

6.3.1. [m?]标记的使用

示例文本	芯片解释
[m3]唯创知音	女声合成：“唯创知音”
[m4]唯创知音	男声合成：“唯创知音”

6.3.2. [s?]标记的使用

示例文本	芯片解释
您好，[s9]欢迎使用唯创知音 [s5]语音合成芯片	芯片语速的调节功能，提供 11 级的语速调节，芯片默认的语速为 5。本句合成时“您好”为 5 级语速，“欢迎使用唯创知音”为 9 级语速，“语音合成芯片”为 5 级语速。

6.3.3. [t?]标记的使用

示例文本	芯片解释
您好，[t10]欢迎使用唯创知音 [t0]语音合成芯片	芯片语调的调节功能，提供 11 级的语调调节，芯片默认的语调大小为 5，最小为 0，最大为 10。本句合成时“您好”为 5 级语调，“欢迎使用唯创知音”为 10 级语调，“语音合成芯片”为 0 级语调。

6.3.4. [v?]标记的使用

示例文本	芯片解释
欢迎[v3]使用[v8]唯创知音语音合成芯片	芯片音量的调节功能，提供 11 级的音量调节，最小 0 为静音，芯片默认的音量大小为 5。本句合成时“欢迎”为 5 级音量，“使用”为 3 级音量，“唯创知音语音合成芯片”为 8 级音量。

语言，如英语，也没有哪一个编码可以适用于所有的字母，标点符号，和常用的技术符号。

这些编码系统也会互相冲突。也就是说，两种编码可能使用相同的数字代表两个不同的字符，或使用不同的数字代表相同的字符。任何一台特定的计算机(特别是服务器)都需要支持许多不同的编码，但是，不论什么时候数据通过不同的编码或平台之间，那些数据总会有损坏的危险。

而在 UNICODE 标准中，提供了 1,114,112 个码点，不仅可以包含当今世界使用的所有语言文字和其他符号，也足够容纳绝大多数具有历史意义的古文字和符号。并且，UNICODE 给每个字符提供了一个唯一的数字，不论是什么平台，不论是什么程序，不论什么语言。

UNICODE 标准已经被工业界的领导们所采用，例如：Apple, HP, IBM, JustSystem, Microsoft, Oracle, SAP, Sun, Sybase, Unisys 等等。最新的标准都需要 UNICODE，例如 XML, Java, ECMAScript, LDAP, CORBA 3.0, WML 等等，并且 UNICODE 是实现 ISO/IEC 10646 的正规方式。许多操作系统，所有最新的浏览器和许多其他产品都支持它。UNICODE 标准的出现和支持它的工具的存在，是近来全球软件技术最重要的发展趋势。