



**本科学士毕业论文**

基于STM32的智能语音播报系统的设计与实现

姓 名： 毛思耀

学 号： 20151104803

院 系： 计算机科学技术学院

年 级： 2015级

专 业：计算机科学与技术（嵌入式）

指导教师： 李红霞

**毕业论文目录**

[1 绪论 1](#_Toc3931169)

[2 系统整体方案分析设计 2](#_Toc3931170)

[2.1 功能需求分析 2](#_Toc3931171)

[2.2 技术需求分析 3](#_Toc3931172)

[3 硬件电路设计 4](#_Toc3931173)

[3.1 系统电路设计 4](#_Toc3931174)

[3.2 功能模块设计 5](#_Toc3931175)

[3.2.1 STM32主控制器电路设计 5](#_Toc3931176)

[3.2.2 XFS5152CE语音合成模块电路设计 6](#_Toc3931177)

[3.2.3 DHT11温湿度传感器电路设计 7](#_Toc3931178)

[3.2.4 音频输出电路设计 9](#_Toc3931179)

[4 软件设计 9](#_Toc3931180)

[4.1 整体框架程序设计 9](#_Toc3931181)

[4.2 功能模块程序设计 9](#_Toc3931182)

[4.2.1 温湿度传感器节点程序设计 10](#_Toc3931183)

[4.2.2 STM32程序设计 11](#_Toc3931184)

[4.2.3 语音合成节点程序设计 13](#_Toc3931185)

[5 主要实验流程 14](#_Toc3931186)

[5.1 软件测试 14](#_Toc3931187)

[5.2 硬件测试 14](#_Toc3931188)

[6 总结与展望 16](#_Toc3931189)

[致谢： 16](#_Toc3931190)

[参考文献： 16](#_Toc3931191)

基于STM32的智能语音播报系统的设计于实现

计算机科学技术学院 2015级嵌入式一班 毛思耀 20151104803

指导教师 李红霞 讲师

摘要 科技的不断迭代进步，催生了人工智能等前沿技术的兴起，在其以一种不可思议的速度发展的同时，带动了大批与之配套的周边技术的发展。在此背景下，已有两百多年历史的语音合成技术在近几年得到了爆发式的发展。故而，语音合成技术被普遍运用于智能交互式系统等领域。基于此设计实现了以32位微处理器STM32芯片为控制核心，并考虑在不同的应用场景模式，通过XFS5152CE语音合成芯片将文本信息转换为语音音频信息，再通过功放电路利用扬声器或喇叭进行播报的智能语音播报系统。

关键词 语音合成；STM32；信号处理；串口通信。

# 1 绪论

在当今节奏快，压力大的社会生活中，由于智能手机的发展迅速导致其普及范围极广，人们将过多的精力放在了智能手机等设备上面。通过观察我们的身边可以发现，很多人甚至在走路的时候也一直低着头在看手机，我们称呼其为“低头族”，也正因此造成其无法分出更多的精力去看身边发生的事情,因此听觉交互成为了一个十分重要的途经来帮助人们获取周边环境的其他重要信息。而传统的语音交互通常需要依靠相关专职人员进行人工播报，但是这些专职人员大多来自于普通的务工者，他们的发音可能不如经过专业训练后的播音员那样精准，而且每个人的文化水平也存在差异，这样就导致了人工语音播报会存在发音不标准甚至偶尔读错字的情况发生，而且在地铁站这种旅客集中的地方，在早晚高峰时段，车站广播员要在4个小时内重复1800多次车辆到站提醒，而且还不能播错，一旦出现播错、漏播的情况对于播音员来说就是一场“事故”。在这种高强度的人工播报负荷下，广播员每次广播到最后，他们几乎都已经头晕眼花，所以在这些场所通常需要安排多名专职专业的双语播音人员来进行轮流播音，可以说极大的增加了人力成本。同时，在一个拉着1000多人的嘈杂的地铁车厢编组里，车厢内最高噪音甚至能达到70分贝的情况下如果仅仅依靠人工播音提示站点信息，会导致很多乘客因为环境嘈杂而听不到站点信息。而智能语音播报系统通过采用先进的嵌入式微处理器这种小型MCU以及借助语音合成技术，不但能够降低系统成本，避免了不必要的浪费，而且这种小型嵌入式系统还具有功耗低等特点，同时还能降低人力成本节约资源，并且还避免了人工播报的一系列问题，提高了服务效率以及用户体验度。智能语音播报系统不仅可以使发音效果清晰可辩，而且相比于人工播报错误率更低，更加准确，同时音色和真人几乎无异。因此该系统非常适用于公共服务、交通服务以及智能家居等领域。

在此背景下，本设计采用ST公司生产的STM32芯片以及语音合成芯片设计开发了一款智能语音播报系统，该系统可以将文本信息转换成语音输出。本文首先分析了智能语音播报系统的研究背景，并在对该系统的应用场景做了详细需求分析以及技术分析的基础上，完成智能语音播报系统总体框架的方案分析设计，主要分成前期对选材时各种硬件芯片的分析选择，相关技术理论分析以及不同场景模式的需求设计实现。之后研究了语音合成芯片的使用，并在STM32芯片的基础上连接语音合成芯片，通过设计程序驱动语音合成模块工作，使其可以将不同输入源信号转换为语音输出。

# 2 系统整体方案分析设计

## 2.1 功能需求分析

在现有的公共服务领域、智能家居等领域，智能语音播报通常只播放已经存储在系统中的固定文字内容。但在交通服务以及行政服务等领域，存在一些特殊情况需要人工进行临时播报，在智能家居领域中，用户甚至想要由自己来制定一些播报内容，为了兼容智能语音播报系统，就需要动态更改系统播放文字内容。同时在环境检测等其他领域，人们需要能够在忙其他事情的同时实时感受到当前天气的温湿度变化并且还不需要为此分出精力，此时将环境检测传感器与智能语音播报系统相结合便完美的解决了这些问题，因此对环境检测数据进行实时播报也成为了人们的迫切需求。

本设计结合上述情况，设计了三种不同应用场景输入源语音播报。其一是社会环境下的公共场所，例如火车站、公交车站、地铁站等需要长期固定播报的地方，播报直接存储在内存中的文字内容，其二是临时播报，由人工通过串口由电脑端输入，系统接收信号解析转换后播报，其三是由其他传感器（如DHT11温湿度传感器）定时采集相关数据参数传入语音合成模块后合成语音信息，进行实时播报。

## 2.2 技术需求分析

智能语音播报中最为关键的一项技术就是语音合成技术。这项技术拥有着十分漫长的历史，最早甚至可以追溯到由Kratzenstein在240年前研制出的机械式语音合成器，然而其真正开始具有实用性价值还是在计算机技术和DSP技术长足发展的近现代。语音合成的方法也经历了从刚开始的机械式合成到电子技术兴起时的电子语音合成，再到七十年代的参数合成方法以及发展至今的波形拼接方法，其中参数合成中比较有代表性的就是串并联模型，然而，虽然这种方法合成的语音十分逼真，但是因为其提取共振峰参数困难，所以该方法实用性不高，而沿用至今的波形拼接中比较有代表性的就是LCP、POSLA、LMA这三种合成技术。该技术还涉及到多种学科的应用，包括声学，即将文字序列经过语义和韵律处理后生成类似于真人发声的波形信号；语言学，即语法分析、断句分词等；数字信号处理等技术的应用。究其本质，我们可以发现语音合成技术主要解决的问题就是可以将任意文字信息转化成为音频信息。并且合成的音频信息还应具有能够正确表达语意、自然流畅、清晰度高等特点。

在语音合成技术中，文本信息都需要按照标准进行规范化，即文本信息的标点符号、特殊符号、计量单位等都需要按照标准转换为可被语音合成芯片识别的信息。语音合成非常关键的是文本信息的分词模型、韵律预测以及协同发音。其中分词模型决定了如何有效地划分文字段落，韵律预测用来确保各个词组的发音不是生硬的机器音并且韵律符合语境，各词组间的连接关系则由协同发音来保障。在上述三部分完成后，语音合成技术将从词库中选择对应词的发音，并进行语音重构，使得文本信息的语音波形能够构建出来。

目前，基于语音合成的技术已经在国内外得到了广泛研究，不少公司成果斐然，其技术已经相对成熟，其商用前景广阔。例如国外的微软、AT&T、Lyrebird等，而国内公司在语音合成、语音识别技术方面的产品比较成熟的有科大讯飞、百度、以及北京宇音天下等。其中北京宇音天下和科大讯飞生产的语音合成芯片是目前国内用户使用最多的，较为常见的芯片。

因此，本设计实现的智能语音播报系统将采用ST公司基于Cortex-M3内核架构上设计生产的STM32F103C8T6芯片作为主控制器，语音合成芯片选用科大讯飞公司研发生产的XFS5152CE芯片，以此来构成整个智能语音播报的主干电路系统。其中STM32控制器的输入源主要包含了程序内嵌的文本信息以及UART的人工输入文本信息和温湿度传感器采集发送来的信号数据。同时STM32控制器将在规定的串口通信协议下实现数据的交换。

# 3 硬件电路设计

## 3.1 系统电路设计

本设计的硬件电路主要包括四个主要部分，分别是STM32 主控制器部分、XSF5152CE语音合成芯片模组部分、DHT11温湿度传感器电路部分和LCD液晶显示屏部分。其中DHT11温湿度传感器可以实时检测出当前环境的温度和湿度，在获取到检测数据参数后，再打印到LCD液晶屏上显示，同时在液晶屏上显示温湿度播报的倒计时，方便调整倒计时时间。STM32主控制器部分用来控制调度各模组的输入输出信号有序进行，即输入源上可以接收内存中的数据、DHT11温湿度传感器数据以及人工发送的串口文本数据。STM32控制器在收到不同场景下的待合成语音播放文本后，将文本数据信息按照语音合成芯片的通信协议组成命令帧格式然后将数据帧发送到XFS5152CE语音合成芯片上，语音合成芯片解析数据帧，然后对命令帧中包含的待合成文本进行语音波形构建，并最终通过音频功放电路以及喇叭完成语音播报。由此设计出整体电路，如图3.1所示：

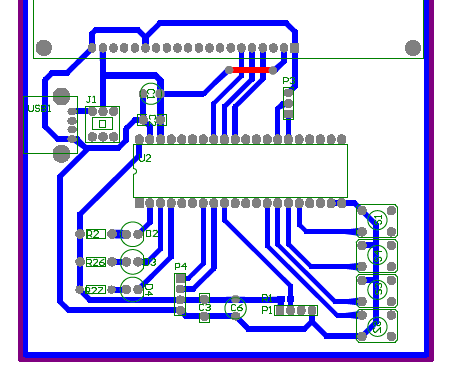


图3.1系统电路PCB接线图

## 3.2 功能模块设计

## 3.2.1 STM32主控制器电路设计

本设计采用的嵌入式控制器STM32F103C8T6通用增强型芯片，其主频能达到72MHz，并且提供了多种总线接口，比如SPI、IIC以及USART，还有两个12位数模转换器等。考虑到本设计中智能播报系统的低功耗，即对运行主频要求不高，因此STM32F103C8T6符合本系统的设计要求。其电路设计如下图3.2所示：

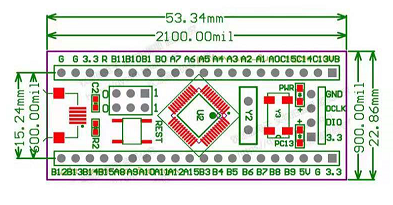


图3.2主控PCB结构图

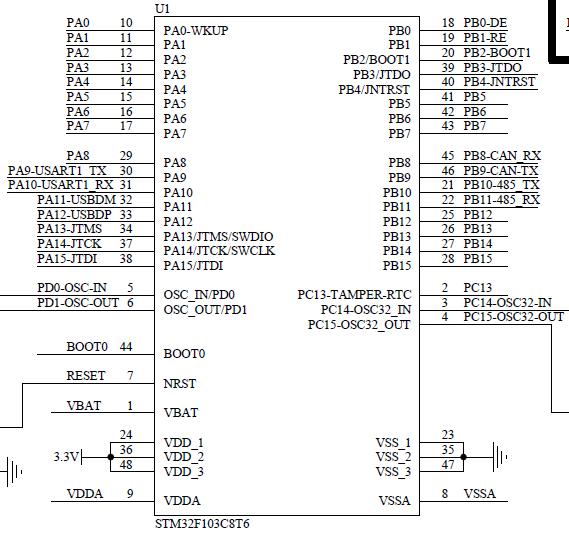


图3.3 STM32F103C8T6引脚封装图

由于STM3控制器的输入电压的产生是由两部分组成，分别为外接的电源到5V的转换电源以及5V到3.3V的供电电压。所以首先要将输入电压转换输出至5V，之后再经过焊接到电路板上的AMS1117-3.3V芯片，即LDO线性稳压芯片输出稳定电压3.3V。即可当做STM32的工作电压使用。在此同时，因为XFS5152CE语音合成芯片内部会自动输出所需要的3.3V和1.2V电压，因此我们只需要将3.3V工作电压接入语音合成芯片对应针脚即可。

## 3.2.2 XFS5152CE语音合成模块电路设计

按照语音合成芯片开发指南，本智能语音播报系统中语音合成芯片的电路原理图如图3.4所示：

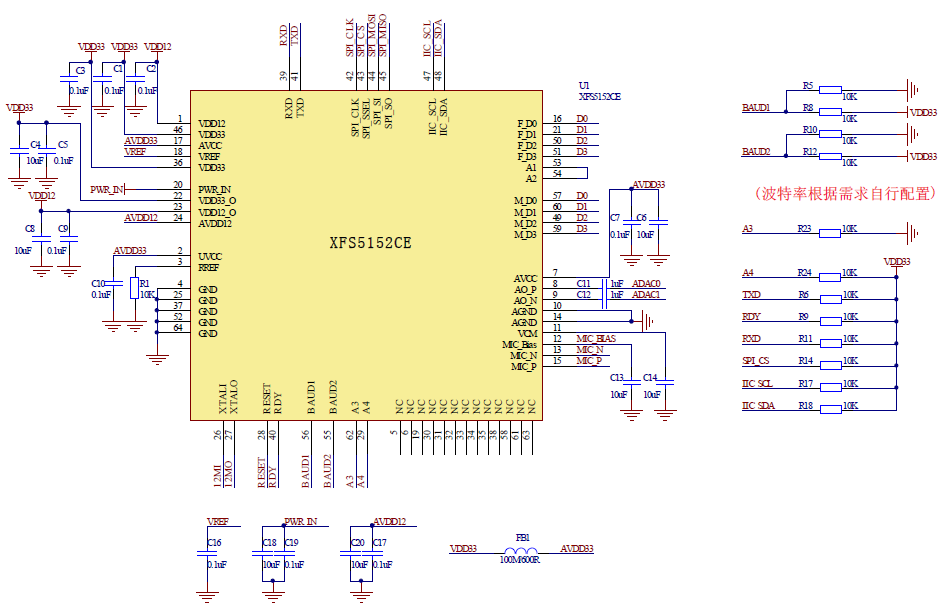


图3.4 XFS5152CE电路图

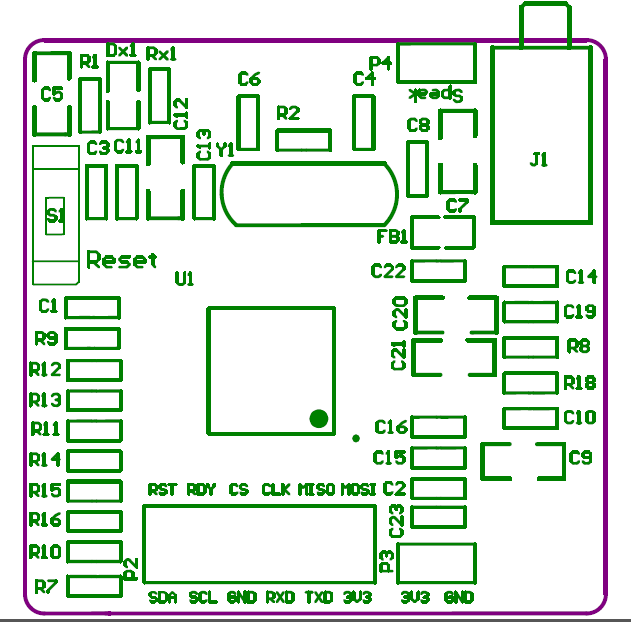


图3.5语音合成模块结构框图

通过查阅指南，可以了解到XFS5152CE芯片的数据读取方式有三种，分别为SPI、USART以及I2C，并且该芯片支持多种控制命令，并且芯片允许发送的数据最大为4K字节。本系统采用了常用的USART串口通信方式，即STM32芯片处理完不同场景下的输入数据后，将通过USART串口按照语音合成芯片的通信协议将数据发送至XFS5152CE芯片。STM32在将数据发送至XFS5152芯片合成语音之前，可以通过语音合成芯片的RDY引脚判断其当前是否正在合成语音，以用来避免因打断芯片的工作，造成合成的语音发生错误的情况发生。当其引脚电平为低电平时，表明了该芯片的状态为空闲，否则表明其正在合成文本，因此可以连接一个发光二极管，在其低电平时，点亮LED，即可查询其状态。

在之前的STM32控制电路设计中，已正确输出了可供XFS5152CE语音合成芯片的工作电压。但由于语音合成芯片还需要模拟电源，因此在电路设计过程中需要引入电感，用来实现高阻抗路径，并将数字电源地以及模拟电源地接地，使其隔离分开。

## 3.2.3 DHT11温湿度传感器电路设计

DHT11温湿度传感器，拥有产品稳定性高、响应速度快、测量精准、自动校准等特点，考虑到传感器采集数据的可靠性以及精准度，因此本设计采用该类型传感器采集空气中的温度和湿度信息。该传感器量程参数如图3.7所示：

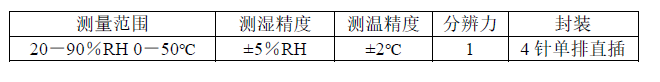


图3.6 性能参数

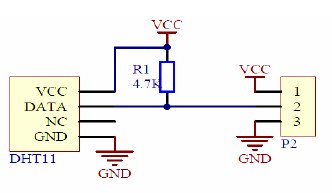


图3.7 接线图

DHT11传感器采用单总线通信，所以其中的NC为空脚，主要使用其他3个引脚，其中DATA接口为数据的引出脚，即数据交换、控制都将通过该针脚进行。通过DHT11的数据手册，可以了解到DHT11的数据输出每次有40bit数据。这些数据由三部分组成，分别为湿度参数、温度参数以及校验和。其中校验和是温度参数和湿度参数相加后计算出的值的末8位为CRC值。数据格式如下图3.7所示：

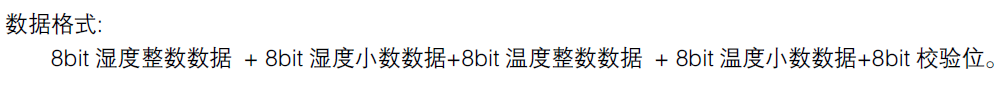


图3.8 数据位定义

在温湿度采集数据的语音播报场景中，由于采集数据的不可见性，因此需要将STM32控制器与12864液晶屏相连，该液晶屏可以动态显示当前采集的温湿度参数以及设定定时倒计时的时间变化和定时播放的剩余时间。LCD12864液晶，也就是屏幕上共有128\*64个只有亮或不亮两种状态的点。由于汉字字形呈现方形的特点，所以目前该液晶屏可以最多显示出32个汉字。因为本系统只需要显示简单参数，所以该液晶屏应用于本系统上完全足够。液晶屏如下图3.9所示：

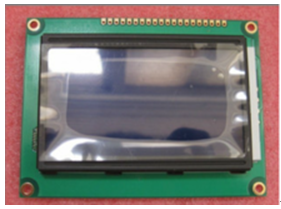


图3.9 LCD液晶显示屏

## 3.2.4 音频输出电路设计

在语音合成芯片完成语音波形构建后，将音频接口线连接语音合成芯片，即可将语音音频信号传送至功率放大器以及喇叭，完成语音播报。其功率放大器为HXJ8002单通道，BTL桥连接的高保真、大功率音频功率放大器，非常适用于本智能语音播报系统的低功耗系统。

# 4 软件设计

本节将详细介绍智能语音播报系统中的软件整体框架设计以及功能模块程序设计。

## 4.1 整体框架程序设计

本文的软件结构主要是通过各数据获取模块将文本信息发送到语音合成芯片，即STM32芯片接收到数据后对文本进行解析，解析后STM32将文本信息发送到XFS5152CE语音合成芯片，此时即完成整个软件系统流程。其中温湿度传感器模块程序在延时函数的作用下控制模块用来定时采集温湿度数据，之后将采集到的数据写入字符串中发送至STM32控制器以用来将采集到的温湿度数据发送到语音合成芯片中；LED函数为显示各状态下的指示灯效果；中断函数为数据的通讯提供中断服务；按键函数控制系统的运行；液晶显示函数将相关数据显示在液晶屏上。

## 4.2 功能模块程序设计

软件功能模块主要包含了温湿度传感器节点程序设计、STM32主程序设计以及语音合成节点程序设计等。接下来将对功能模块进行详细设计。

## 4.2.1 温湿度传感器节点程序设计

传感器节点模块主要负责监测实时温湿度的采集，并将通过主控板引出的I/O接口将温湿度数据发送到STM32控制器，通讯流程如下图4.1所示：

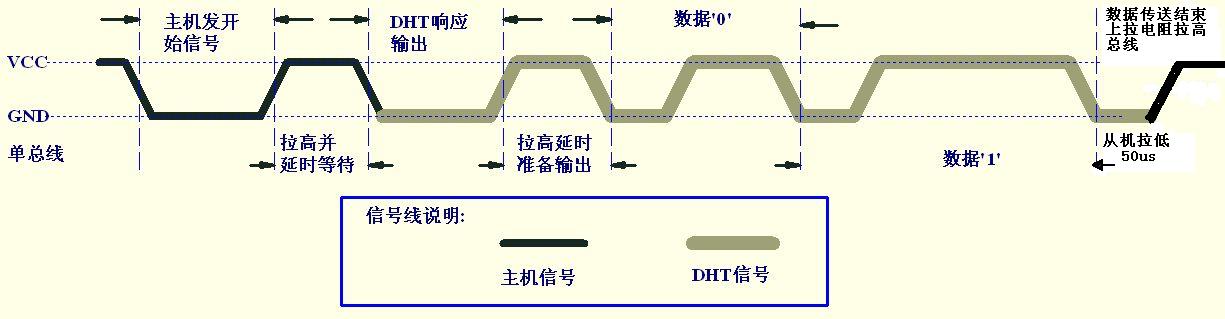


图4.1 温湿度传感器通讯过程

如上图4.1所示，传感器接收到STM32芯片发来的开始信号节点后等待该信号结束,然后向主控制器发送响应信号并延时80us，之后在程序指定的I/O接口下发送按照指定格式编排好的采集数据。其传感器连接的IO口初始化代码如下图4.2：

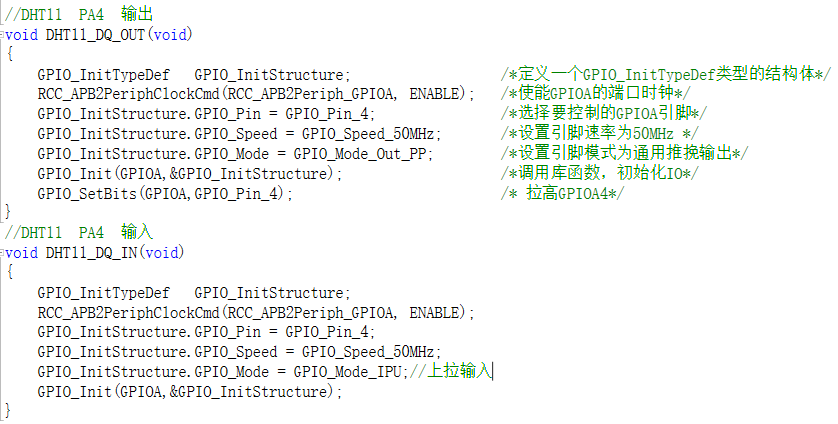


图4.2 GPIO初始化

传感器初始化端口成功后，将复位DHT11并读取数据并发送到STM32控制器中。读取数据的关键代码如下图4.3：

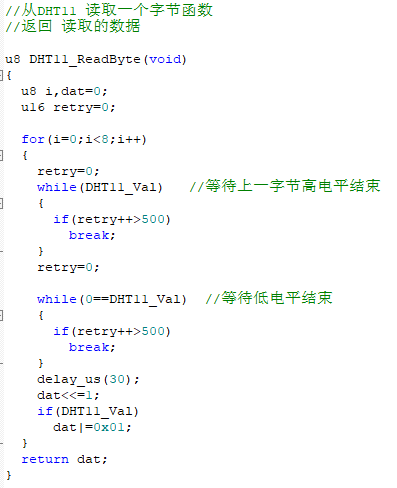


图4.3 从DHT11读取一次数据

## 4.2.2 STM32程序设计

在本文的三种场景模式下，所有的语音预处理数据都需要经过USART串口发送至语音合成芯片中，这些语音预处理数据不仅包含人工输入的含文本语音合成数据，而且还包含了温湿度传感器发送来的环境检测数据，以及程序自身存储在字符串中的数据。通讯过程如下图4.4：



图4.4 USART串口通讯程序流程图

在USART串口通信程序中，STM32以及USART需要设置相同串口通信参数，方才能够建立连接并进行稳定的数据传输。串口通信参数通常包含了波特率、CRC校验等。本系统的串口通信格式采用了基于帧的数据格式。每一个帧的数据主要包含了三部分,分别是帧头、帧的数据区以及校验区。在系统运行过程中，待语音合成数据都需按照帧结构发送至STM32芯片。STM3控制器中串口初始化程序如下图4.5：

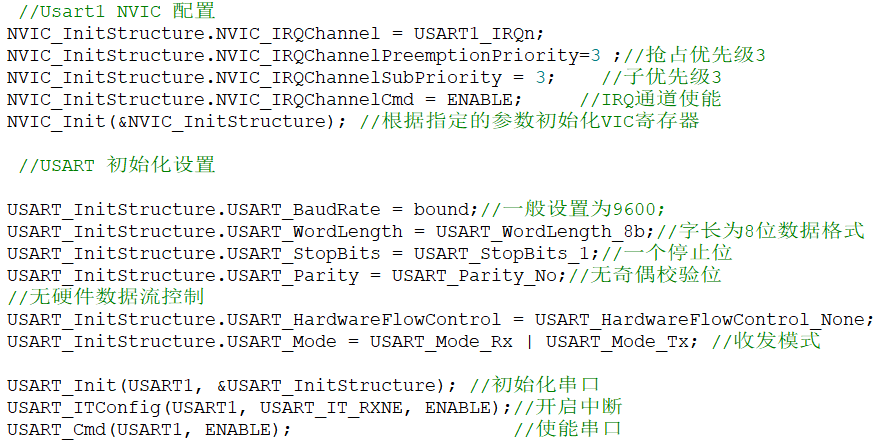


图4.5 串口初始化

## 4.2.3 语音合成节点程序设计

通过查看XFS5152CE的数据手册，可以了解到XFS5152CE有自己设定的通讯协议，并且接收的语音合成命令采用数据帧的方式进行封装。因此待合成语音的数据帧结构要严格按照统一的通信协议进行处理。

命令帧结构包含了三部分组成，从前到后依次是帧头标志、数据区长度以及存放待合成文本的数据区。其命令帧的帧头标志为0xFD；数据区长度为2Bytes，由高字节和低字节组成；数据区则包含了帧结构中的有效信息，具体来说就是命令控制字、文本编码格式以及待合成文本，其命令控制字中合成命令为0x01、停止命令为0x02、暂停为0x03、恢复为0x04。文件编码格式也有GBK、GB2312等国内常用的四种中文编码格式可供选择。STM32控制器串口发送文本信息到XFS5152CE语音合成芯片的关键代码如下图：



图4.6 参数命令帧关键代码段

# 5 主要实验流程

前面几章对本项目中的智能语音播报系统做了充足的环境和软硬件配置，使其能够进行有效编码工作。在完成编码工作之后，需要进行测试来检验开发好的系统是否能够正常工作，并通过测试来发现可能出现的代码错误或者逻辑漏洞。因此我们需要对该系统进行单元测试和集成测试来确认本系统是否具备了我们所期望达到的功能和目的。其方式又可以分为软硬件测试两种。

## 5.1 软件测试

单元测试，又称为模块测试。顾名思义，模块测试就是需要验证系统中每个基本模块的功能。因为一个完整的系统代码通常有多个不同功能模块代码组成，因此独立验证每个模块的功能可以及时定位出问题的模块，然后修改功能错误。在单元测试过程中，将着重考虑四个方面，一是模块接口，即测试模块的接口主要是确保模块的输入输出参数信息是正确的，测试所测模块或功能的数据流；二是局部数据结构，即检查代码中声明或者定义的数据类型是否是合法有效的，同时检查这些数据结构的初始化值是否正确；三是独立路径，即要合理地设计测试用例以便找出可能是由于错误计算、错误逻辑而导致的问题，这些错误的控制流通常包括了死循环、变量赋值类型错误等；四是错误处理路径，即一个完善的软件应该能够抛出异常或是停留在错误逻辑上，并在异常或者错误发生时，通过预先设计的判断代码合理安全的处理问题或是抛出问题。错误的类型有：野指针引用、数组越界、过度释放、获取到异常或是错误但没有足够的信息来表明问题出现的原因。经对每个模块的独立测试，本系统中的涉及的每个模块代码编译无报错，通过添加调试信息的方式，即打印调试信息到串口助手上或打印到液晶屏上来单独验证功能模块的程序运行正确。

集成测试，即将所有子模块单元组合后进行编译链接执行，来检验在设计阶段是否出现了不符合逻辑、耦合度偏高、内聚度偏低等问题，以用来找出单元测试中无法找到的BUG。经对整个系统的编译链接，本系统代码运行无误，零Error、零Warning。通过将程序生成的.hex文件烧录到STM32芯片上运行观察。本设计代码可以实现需求功能。

## 5.2 硬件测试

测试步骤如下：连接系统硬件模块，上电后分别对人工输入语音播报测试以及温湿度传感器播报测试。硬件连接及液晶界面显示如下图5.1：



图5.1 硬件连接及上电后液晶界面显示

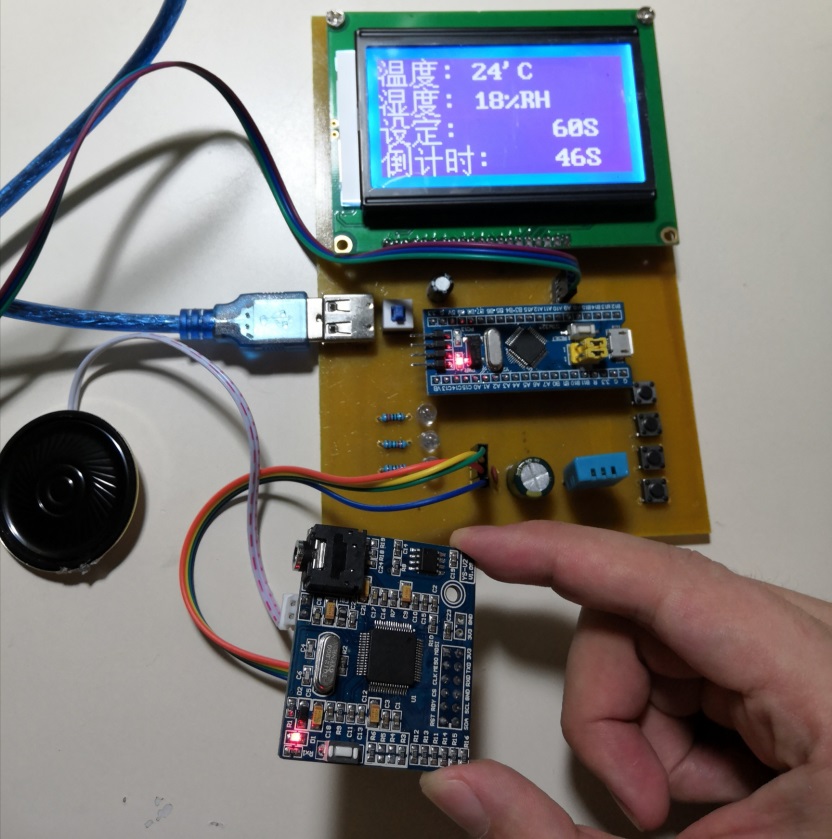


图5.2 进入温湿度播报界面

# 6 总结与展望

本设计在分析智能语音播报系统的需求基础上，提出了三种不同语音播报模式：人工输入模式、温湿度传感器语音播报模式以及内存读取语音播报模式。在不同模式下，本文充分考虑了其硬件以及软件实现的结构，最终采用了STM32控制器以及XFS5152CE语音合成芯片来接收不同场景的语音数据以及合成最终的输出语音音频信号。经过详细的硬件电路设计、软件实现以及综合测试后，本智能语音播报系统运行稳定，并且可以在三种不同场景模式下成功完成语音播报，其发音效果清晰、准确，取得了很好的语音播报效果。

但同时也需要注意到，对应人工输入场景，本系统未设计上位机程序，仅仅依靠串口调试助手来完成人工输入播报内容，这在实际场景中不是很方便。未来可以通过设计有效的上位机程序，来提高用户的交互式体验，同时还可以设计一个文件存储系统，这样本智能语音播报系统就可以从SD卡中的相应文档中读取文字来智能完成语音播报。

# 致谢：

在经过几个月的硬件电路设计以及软件代码编写后，毕业设计以及毕业论文得以顺利完成。首先，我要特别感谢我的论文指导老师李红霞老师。在这几个月里，从开始的选题、确定功能需求、收集相关资料、电路板的焊接、代码编写直至论文的最终完成，李红霞老师给予了我极大的指导和帮助。李老师不惜牺牲宝贵的课间休息时间来解答我的疑问，纠正我的错误。在这里我向您致以最真挚的感谢！同时也感谢大学四年来学院里所有帮助过、教导过我的老师们，你们辛苦了！正是你们的教导，让我不仅学会了专业技术知识，更懂得了什么才是正确的人生观、价值观。同时也要感谢我的同学们，正是你们陪我一起成长和学习，在这里向你们表示我深深的感激之情。

# 参考文献：

[1] 刘琦，刘滨，朱兆优.基于STM32的语音播放系统的设计[J]．科技广场．2014(94-98).

[2] 胡良焕,杨国涛,侯永春,郭晓学,秦铆.基于语音播放的环境检测仪的应用研究[J].信息通信, 2015.

[3] 王虎升,李金环,袁宪锋,张胜春. 基于STM32的嵌入式语音播报系统的设计[J].北京联合大学学报(自然科学版). 2011(03).

[4] 韩永刚. STM32在液晶显示模块上设计的研究[J]. 电子测试.2017(09).

[5] 张敏,石倩倩,张珊珊,王田,梁晓平. 基于语音识别和STM32的老年人健康状况监护系统的设计[J].电子质量.2017(05).

[6] 杨博,张加,李敏,等. 基于ARM 的多通道数据采集系统[J].仪表技术与传感器,2015(2):104-107.

[7] 齐亚萍,李亚,雷升杰.基于ARM 的远程数据采集系统设计[J].自动化与仪表,2015(3):57-60.

[8] 何凯.智能车载语音播报器的设计与开发[D].湖南:湖南大学，2012.

[9] 唐颖.公交车智能播报系统的研究与设计[D].程度：电子科技大学，2017.

[10] 董笑甜. 基于STM32的草莓温室控制系统[D].曲阜师范大学, 2018.

[11] 李雨轩.具有语音播报功能的水温控制系统 [D]. 商品与质量, 2018,147-149.

[12] 鹿晓茸.基于H9200CS的无线语音播报系统 [D].济南, 山东大学, 2013.

**Design and Implementation of Intelligent Voice Play System Based on STM32**

Computer science and technology College 2015Embedded class 1. Mao siyao 20151104803

Directed by Li hongxia lecturer

**Abstract** The continuous iterative advancement of science and technology has spawned the rise of cutting-edge technologies such as artificial intelligence, and at the same time it has developed at an incredible speed, and has driven the development of a large number of peripheral technologies. In this context, the speech synthesis technology with more than two hundred years of history has developed explosively in recent years. Therefore, speech synthesis technology is widely used in intelligent interactive systems and other fields. Based on this design, an intelligent voice broadcasting system with 32-bit microprocessor STM32 chip as the control core and different application scenarios is designed and implemented. Text information is converted into voice and audio information through XFS5152CE voice synthesis chip, and then broadcast by loudspeaker or speaker through power amplifier circuit.

**Keywords** speech synthesis; STM32; signal processing; serial communication