基于语音识别技术的智能家居主控设计

**2.系统方案的设计**

2.1 系统总体功能框架介绍

本课题旨在设计一套功能完善的智能家居系统，给用户提供一种简单、轻松、高效的生活方式。基本构想是实现用户家庭设备的无线连接，在家庭环境中可以通过简单的语音交互控制各个设备，同时远程可以通过手机客户端实时监测家庭环境。为了满足以上功能，我们设计的智能家居系统总体框图如图X所示，该智能家居系统是由阿里云服务器、具有语音识别功能的主控（语音主控）、手机客户端、智能设备四部分组成，其中阿里云服务器主要用于数据的存储与共享以及设备的远程监测与控制，手机客户端用于设备的统一管理，而语音主控除了普通的家庭网关的功能以外，还具有远声场语音识别的功能，智能设备主要涉及灯光、红外控制器、窗帘、燃气传感器等常用家庭设备。如图所示，系统以智能家居语音主控为核心，利用科大讯飞AIUI语音识别模块，实现远声场语音的采集与识别，通过ZigBee技术组建家庭无线传感控制网络，实现主控与家庭智能设备的互联。同时，语音主控在功能上相当于家庭网关，通过串口将ZigBee网络与万维网联系起来，实现异构网络之间数据的转发，可以及时地将智能设备的状态上传至云端服务器，与手机APP相配合实现数据的云共享以及家庭环境的实时远程监控。



2.2 智能家居语音主控功能设计

语音主控是智能家居系统的核心，是整个系统的大脑，处于系统的决策和控制层，在功能上需具备如下几点：

1. 语音识别功能，也是系统的核心特色功能，我们采用科大讯飞的AIUI语音识别模块，并配合6麦克风环形阵列，具有语音唤醒和3-5米远声场语音识别的功能；

2. 异构网络之间数据的转发和控制协议的解析，语音主控主要有AIUI语音模块与CC2530模块组成，其中AIUI模块上搭载有WiFi模块，可以通过WiFi实现与Internet网络的连接，而CC2530模块作为ZigBee网络的协调器，负责组建家庭内部的ZigBee无线传感网络并实现网络中数据的接收和发送，AIUI模块与CC2530模块通过串口进行联系；

3. 与阿里云服务器通信，语音主控需要及时地将家庭中智能设备的状态上传至阿里云服务器以便于用户家庭数据的云共享。同时，语音主控还需要接受用户的远程控制指令，当用户发送远程控制指令至服务器时，服务器通过推送将控制指令透传至相应主控；

4. 与本地手机APP客户端交互，当用户与语音主控处于同一个WiFi局域网内，除了通过语音直接控制外，还可以通过TCP/IP协议向语音主控发送控制指令，这就要求语音主控必须具有本地微型服务器的功能；

5. 与ZigBee网络中的其他智能设备通信， CC2530模块作为ZigBee网络的协调器，负责组建家庭内部的ZigBee无线传感网络并实现网络中数据的接收和发送。当语音主控接收到控制指令后，通过AIUI模块的串口发送至CC2530模块，CC2530模块通过ZigBee协议将控制指令发送至相应的智能设备。

综合以上分析，本设计的智能家居语音主控包含以下几个模块：AIUI语音识别模块、CC2530 ZigBee数据传输模块、本地微型服务器模块、阿里云服务器通信模块。

2.3 语音识别方案

2.3.1 语音方案的选择

语音识别技术的出现，让机器拥有了“耳朵”，自然语言交互被认为是人与机器之间最简单自然的沟通方式。从1952年贝尔实验室研制出世界上第一个能识别10个英文数字发音的语音识别系统，到如今百家争鸣的语音识别系统，语音识别技术的发展可谓是日新月异，与机器用自然语音对话的梦想也逐渐实现。远声场的语音识别技术包括前端声学处理和后端语音识别算法，涉及到声源定位、降噪、回声消除、语音唤醒、端点检测等多项技术。目前国内在中文语音识别领域做的比较好的有科大讯飞、百度、阿里巴巴、云知声、思必驰等。我们通过对比各大企业提供的服务和产品的实际性能，并综合我们的需求做出合理的选择。

国内各大企业语音识别方案对比

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 科大讯飞 | 百度 | 阿里 | 云知声 | 思必驰 |
| 接入方式 | 软件/软硬一体 | 软件 | 软件 | 软件/软硬一体 | 软件/软硬一体 |
| 产品类型 | 智能硬件、多平台软件、软核 | 多平台软件 | 多平台软件 | AI芯、多平台软件 | 智能硬件、多平台软件 |
| 提供服务 | 智能家居、音乐、故事、新闻、闲聊等100多个内容服务 | 依附于百度平台的各种服务包括音乐、天气等200多项技能 | 天气、附近地点、设定闹钟、查看股票等生活领域 | 智能家居、音乐、天气等日常多种服务 | 智能家居、音乐、天气等日常多种服务 |
| 远声场识别 | 最高支持8麦克风三维空间拾音 | 6麦克风二维空间拾音 | 无说明 | 二维空间拾音 | 6麦克风二维空间拾音 |
| 个性化 | 唤醒词定制、热词上传、自定义技能 | 唤醒词定制 | 用户自定义热词 | 上传个性化数据 | 唤醒词定制、场景定制 |
| 识别方式 | 在线、离线 | 在线、离线 | 在线 | 在线、离线 | 在线、离线 |
| 方言识别 | 客家语、东北话、天津话、粤语、河南话、四川话等20种方言 | 粤语和四川话 | 东北、河南、四川、粤语 | 无说明 | 无说明 |

目前而言，语音识别仍是一项高科技的领域，需要时间的积累和沉淀，通过上述的综合对比分析，可以看出科大讯飞在语音识别方面做得比较全面，有一套完善的体系，提供的产品种类更加丰富，识别准确率更高，接入方式简单且支持个性化定制，最终我们采用科大讯飞的AIUI智能硬件，并配合六麦克风环形阵列，可实现3至5米远声场92%以上的识别精度且支持多种方言识别。

2.3.2 AIUI智能硬件语音识别的工作原理简介

AIUI智能硬件主要由科大讯飞XFAI0801模块和6麦克风环形阵列组成，其工作原理如图所示，首先由6麦克风环形阵列对声场的空间特征进行采样，采集的声音信号进行模数转换至XFAI0801模块进行信号的预处理，XFAI0801模块中集成了配套的麦克风阵列算法，可实现360度声源的定位、回音消除、去除混响等降噪处理，处理过后的信号进行前后端点检测并送至云端连续识别远程引擎，该引擎为讯飞核心技术之所在，不仅包括语音信息到文字信息的转换还包括语义的理解以及内容平台和用户个性化系统，最终给我们返回的是讯飞平台自定义的Json格式的文本，而开发者的重心就是对Json文本信息的解析，专注于自己的业务逻辑。从中也可以看出，AIUI智能语音模块是科大讯飞提供的在语音识别方面软硬一体的解决方案，内部集成了语音采集与处理功能并封装了语音交互的复杂功能，旨在为开发者提供简单易用的交互接口。

语音识别工作原理框图

2.4 无线通信方案

智能家居领域由早期的有线连接方案已经逐步过渡到了无线连接，智能设备的无线连接因其灵活、易安装也促进了智能家居的推广。目前市场上，智能家居的主流产品大多有WiFi、蓝牙、ZigBee进行无线连接。在各类近距离无线通讯协议的选择中，我们通常会拿几个指标进行对比，包括传输速率、功耗、距离、组网等，其中WiFi高带宽和高功耗的特征，更适用于各种插电和需要高速数据传输的设备，而蓝牙和ZigBee则更适用于电池供电或自供电的设备，包括可穿戴设备以及大量传感器数据传输设备。

三种无线通信协议的对比

|  |  |  |  |
| --- | --- | --- | --- |
|  | WiFi | 蓝牙 | ZigBee |
| 通信距离 | 50-100m | 10-20m | 50-100m |
| 传输速度 | 11-54Mbps | 1Mbps | 256kbps |
| 功耗 | 高 | 中 | 低 |
| 组网规模 | < 36 | 点对点 | 理论上< 65535 |
| 主要应用 | 无线上网、PC、手机 | 可穿戴、多媒体 | 工控、医护 |

ZigBee技术具有低功耗、低成本、自组织、网络容量大等优点，在对传输速率要求不高且价格敏感的智能家居领域，将会提供比蓝牙和WIFI更有效、性价比更高的解决方案。同时，ZigBee无线局域网络不依赖与任何网路基础设施，由协调器自发地扫描并建立网络，终端节点加入网络后形成自组织的局域网络，网络的容量大（理论上一个网络可容纳65535个节点）且稳定性好。本系统采用TI的CC2530芯片并配合Z-Stack协议栈可以轻松实现家庭无线网络的组建。同时，由于ZigBee网络容量大，可以满足家庭中的几乎所有智能设备的入网要求。

特别说明，本课题开始于2015年，当时智能家居产品中的无线方案主流的也就是WiFi、蓝牙和ZigBee，我们是针对这三个无线协议的对比分析并结合我们自身需求做出的选择。但当今社会科技日新月异，2016年9月份随着NB-IOT技术开始商业化部署，低功耗广域网络（LPWAN）在逐渐普及，在智能家居领域又多了一个竞争者。低功耗广域网络在各项指标中都有着不错的表现。从传输距离角度来看，由于LPWAN网络广覆盖、穿透性强的特征，无处不在的网络可以让各类低功耗设备随时接入，连那些深埋在地下的管网和各种角落的计量表都可以实现覆盖和连接。从功耗角度来看，LPWAN具有更低功耗，对于一些电池供电的设备来说，甚至数年都无需更换电池。从组网角度来看，ZigBee一直引以为豪的是组网能力强，一个网关理论上可连接数万设备，且可以Mesh组网，而LPWAN作为广域网络，只要支持该网络协议的设备均可直接接入相应网络，就像手机一样，开机即可接入网络，提升用户体验。当然，低功耗广域网络将会是我们下一代智能家居无线通信的重点考虑对象，本课题所采用的无线通信协议是目前比较成熟的ZigBee技术。

**3 智能家居语音主控的硬件设计**

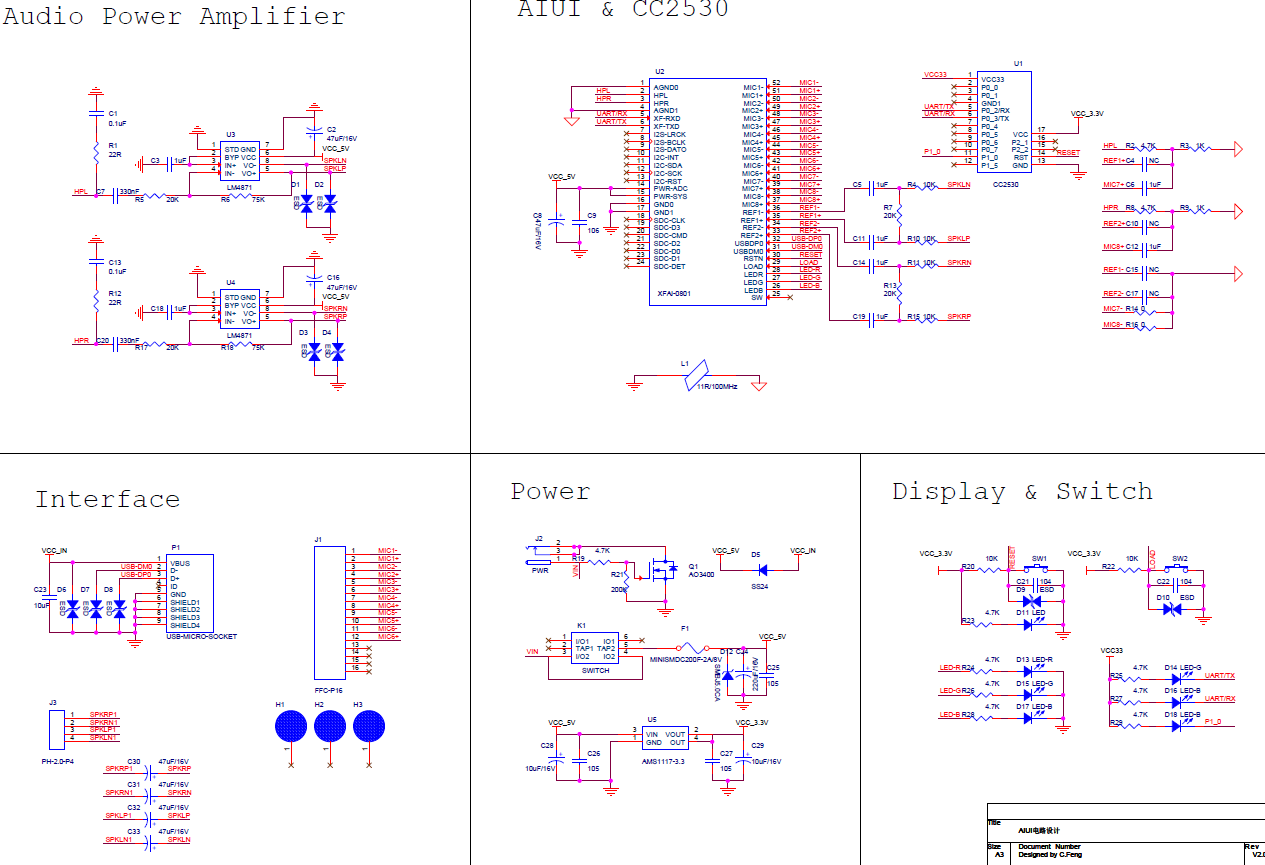
3.1 硬件总体设计

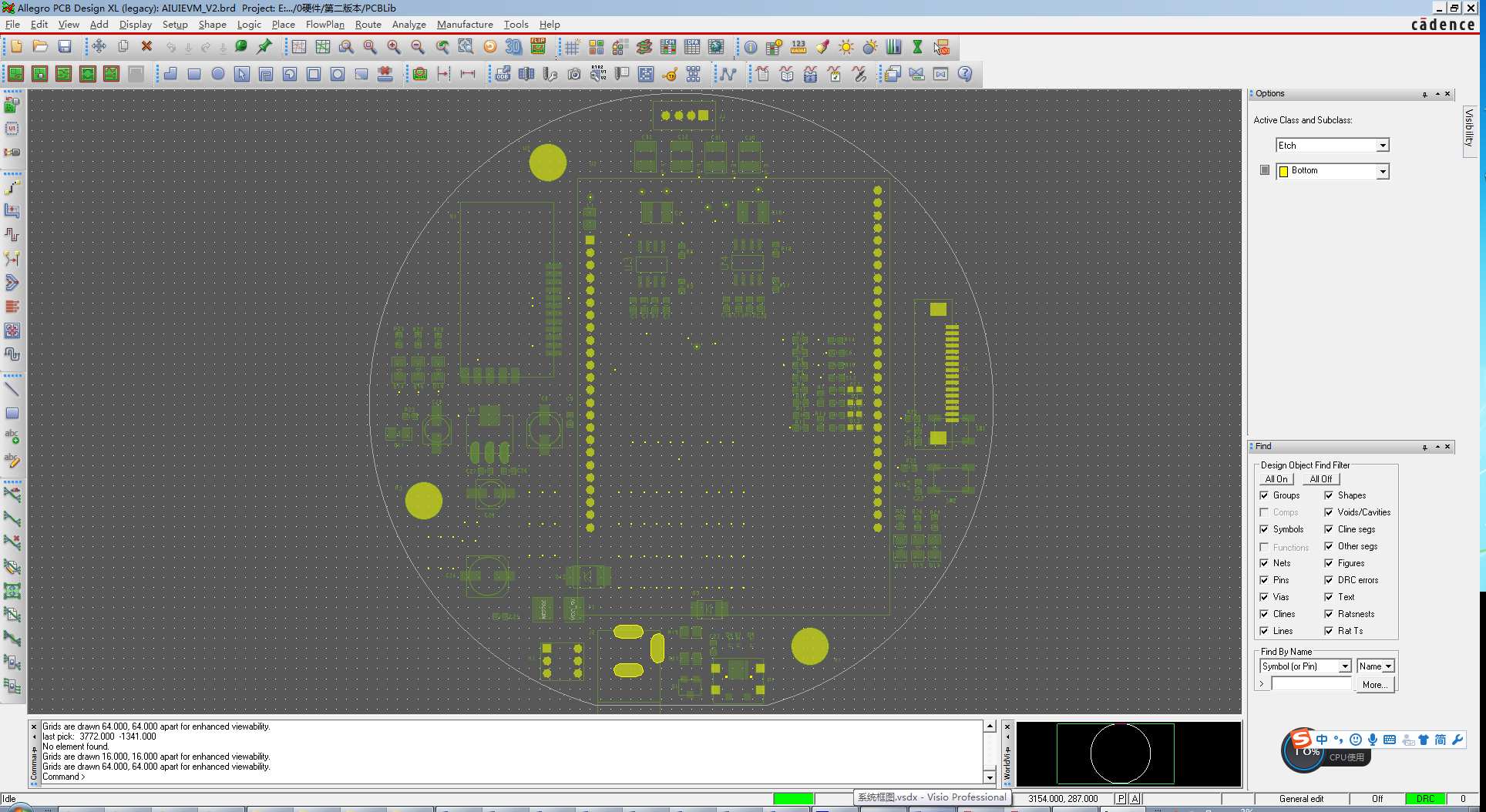
在硬件设计上，语音主控主要由AIUI语音识别模块、CC2530 ZigBee通信模块、电源模块、音频功放模块以及麦克风阵列几部分组成。如图Z所示，其中AIUI模块和CC2530模块是语音主控的核心部件，AIUI模块主要负责语音采集处理与语音识别，同时还负责与手机客户端和服务器通信。而CC2530模块主要负责组建家庭内部的ZigBee无线传感网络并实现网络中数据的接收和发送。



硬件总体框图

语音主控的硬件设计主要采用Cadence电子设计软件进行开发，Cadence软件功能很强大，不仅包括硬件设计中原理图与PCB的制作，还提供强大的电子仿真功能，有关Cadence的介绍和使用，目前市场上的书籍很多，此处不再累述，最终设计的原理图、PCB图以及整体实物图如图2所示。





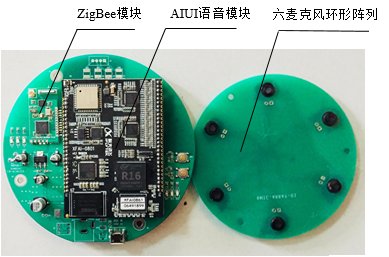


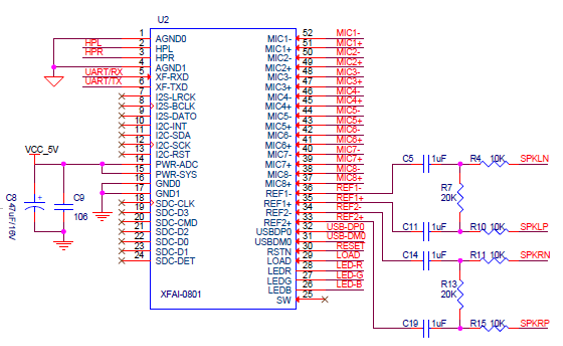
图2 语音主控实物图

Fig. 2 Architecture of smart home system

3.1 AIUI语音模块

AIUI语音模块是科大讯飞提供的在语音识别方面软硬一体的解决方案，其中CPU采用全志的R16一款4核cortex-A7智能硬件处理器；内存(RAM)采用三星DRAM型号为K4B4G16其内存大小为1GB ；Flash采用NCEMASD9-08G EMMC存储空间为4GB；ADC音频采集芯片为科胜讯CX20819-11Z一款4通道远场语音捕获高性能高清音频ADC；WiFi采用AP6210 WiFi /蓝牙合二为一芯片，WiFi方面支持b/g/n 2.4G，最高带宽72.2Mbps，其实物图如图2中所示。

为了语音模块的集成开发，需要清楚模块上各个引脚的定义与用途，如表x所示，其中输入引脚主要有电源输入、麦克风阵列信号采集输入、用于回音消除的参考信号输入和用于开发调试的USB输入，输出引脚主要有音频信号输出、串口输出，语音模块的外围电路如图所示，各部分的功能与连接状况会在后续的相应模块中详细介绍。



整个模块的操作系统采用Andriod4.4系统，可以根据业务需要支持二次开发，在功能方面可以灵活扩展；该模块在语音识别方面具有如下功能[6]：

1）远场精准识别

运用远场识别和降噪技术，使拾音距离达到5米。

2）360度声源定位

模块利用麦克风阵列技术，实现360°语音信号采集，并能通过声源定位来确定目标说话人的方向。

3）回声消除

在播放和录音同时进行的场景，模块通过回声消除技术，可以将扬声器的声音屏蔽，只接收用户的声音。

4）全双工持续交互

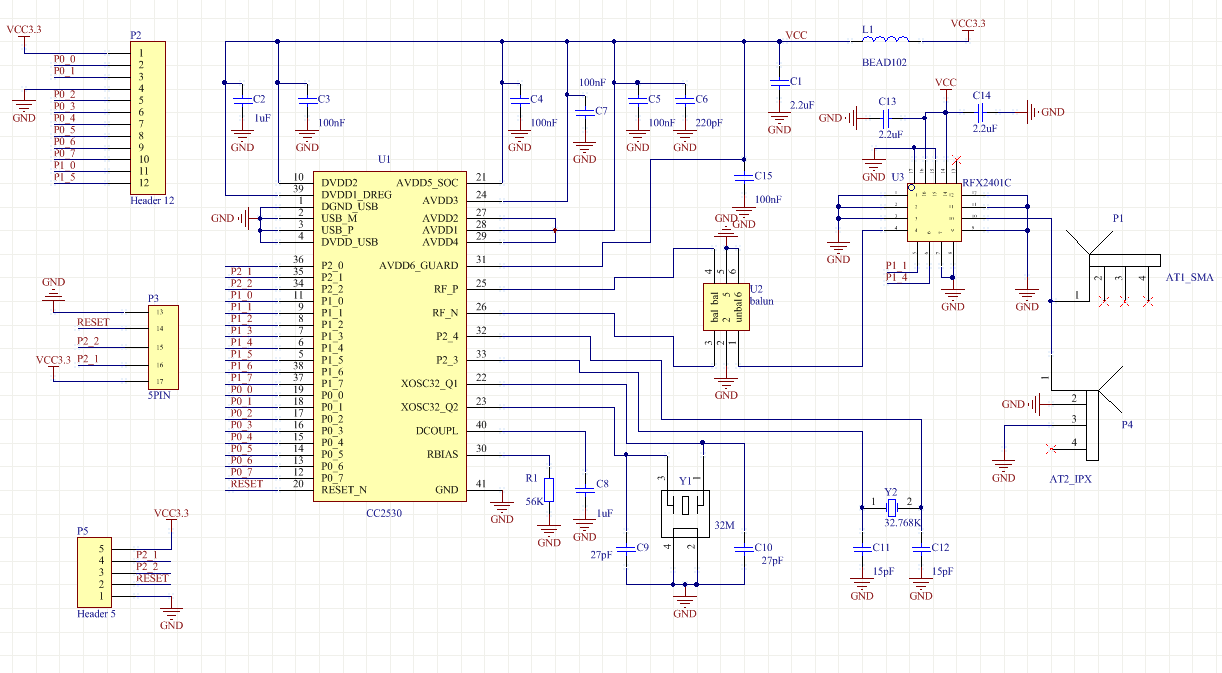
持续监听语音源，并将语音进行录音上传到云端。云端对接收的语音进行不间断的识别处理。

5）支持上下文对话

结合交互的上下文，准确理解说话人的意图，同时管理多个对话场景，各个场景之间根据用户意图进行场景自适应级场景切换；长时间的记忆用户交互历史，并将这些记忆利用在交互中，使交互更加智能。

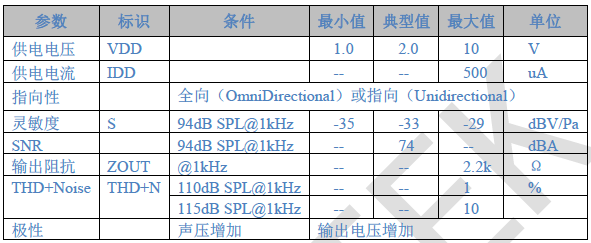
3.2 ZigBee通信模块

ZigBee技术使用2.4GHz频段，采用跳频技术，传输速率可达256Kb/s，因具有低功耗、网络容量大等特点，适合应用于智能家居领域[7]。ZigBee通信模块如图2中所示，主要采用TI CC2530芯片和2401前端功放进行设计，在增加传输距离的同时配合Z-Stack协议栈组建Mesh网络，可以实现底层智能设备的灵活扩展。CC2530芯片的外设包括：强大的5通道DMA、具有捕获功能的32kHz睡眠定时器、具有8 路输入和可配置分辨率的12 位ADC、2 个支持多种串行通信协议的强大USART、21 个通用I/O 引脚、看门狗定时器等。可以满足传感节点和各种扩展功能节点的外设需求。图x是基于TI CC2530芯片制造的ZigBee模块的原理图，该模块预留了必要的引脚，包括电源接口、IO口、下载口以及串口，以模块的方式集成起来，简化了电路的设计，在底板上只需要设计对应的接口即可。

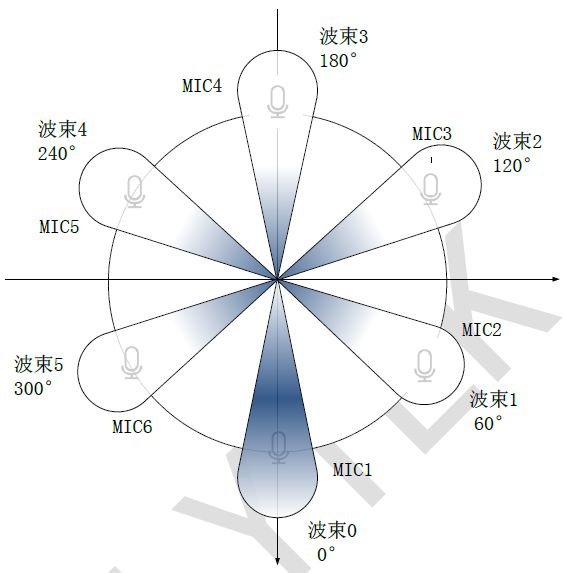


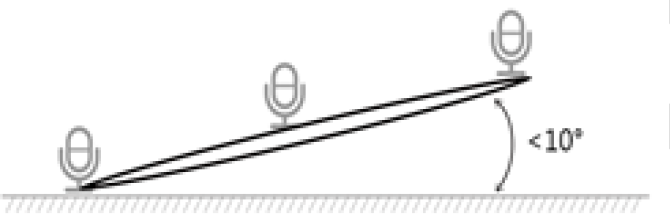
3.3 麦克风阵列

本系统采用6麦克风环形阵列并配合科大讯飞提供的麦克风阵列算法，可实现远场拾音降噪、回声消除、声源定位等功能[8]，其中麦克风采用大口径，高灵敏度的全向麦克风，如图2中所示。麦克风具体参数如下：



麦克风阵列需要与对应的算法配合，因此，各个麦克风的位置也有严格的要求。6麦环形阵列呈圆形布局，其中6个麦克风均匀分布在圆周，麦克风按圆形等距摆放，半径为35mm，如图6所示。圆平面和水平面之间可以有一定夹角，但夹角不能超过10°（图7）。麦克风阵列的零度方向必须和产品的正面朝向保持一致。



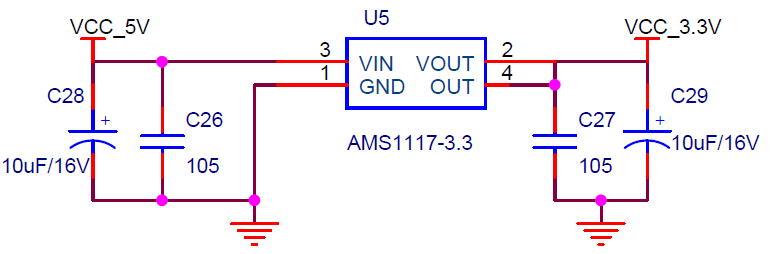


3.4 电源模块

电路中需要供电的主要有语音模块、LM4871音频功放芯片、ZigBee模块，各个器件对供电要求如表所示，因此，整个系统中只需要直流5V和3.3V两种电压。

|  |  |  |
| --- | --- | --- |
| 模块 | 电压/V | 电流/mA |
| AIUI语音模块 | 4.2-5.5 | 300-700 |
| LM4871 | 2.0-5.5 | 300-600 |
| ZigBee模块 | 3.3 |  |

5V的电压直接采用一个5V 2A的USB电源转换器直接供电，而3.3V采用5V转3.3V的线性电源稳压芯片AMS1117-3.3V，线性电源具有纹波小精度高的特点，原理图如图所示。



3.5 音频功放

音频功放电路设计如图4所示，采用TI的LM4871单声道桥接音频功率放大芯片，在5V的输入电压条件下，能够为4Ω负载提供3W功率的稳定输出，总谐波失真和噪声不超过10%，在左右声道上各加入一个功率放大，因此最大可提供6W的功率输出，可以满足普通家居环境中的音乐播放需求。

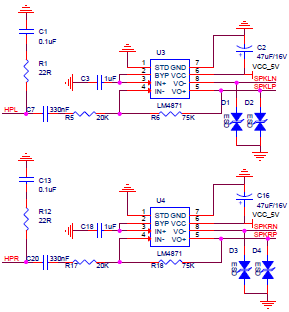
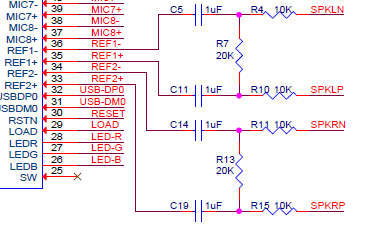


图4 音频功放电路

Fig. 4 The circuit of audio power amplifier

同时为了减小扬声器输出声音对麦克风阵列语音采集的影响，需要将扬声器的输出作为参考信号接入到AIUI语音模块的相应引脚，参考信号输入Vpp ≤150mV，当输出一般高于此值，需要分压时，分压电路如图所示。



3.6 串口通讯电路

AIUI模块与CC2530模块是系统的两个核心模块，两者之间通过串口进行数据交互，由于两者串口均为TTL电平，中间不需要转换，可直接连接，串口通信参数设置如下，采用波特率为115200 bps，8位数据位，1位停止位，无奇偶校验位。

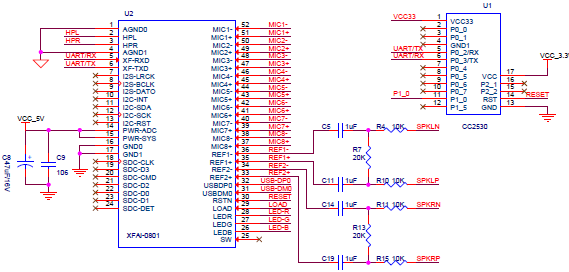
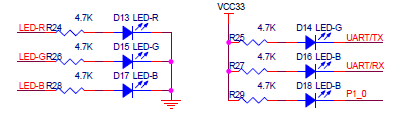


图3 串口连接电路

Fig. 3 The circuit of Serial communication interface

3.7 系统指示灯

本系统除了电源指示灯以外，还有一些用于指示当前工作状态的指示灯，如图所示，LED-R、LED-G、 LED-B主要用于指示AIUI语音模块在交互工作中的状态，而与TX/RX/P1\_0相连的LED灯主要用于指示ZigBee模块的这个状态，



|  |  |  |
| --- | --- | --- |
|  | 灯状态 | 设备状态 |
| AIUI模块 | 灯灭 | 休眠状态 |
| 红灯闪烁 | 未连接WiFi |
| 红灯常亮 | APPID校验不通过 |
| 绿灯常亮 | 录音状态 |
| 蓝灯亮 | 拒识状态 |
| ZigBee模块 | TX闪烁 | 串口发送数据 |
| RX闪烁 | 串口接收数据 |
| P1\_0常亮 | 网络连接正常 |

指示灯的一个重要作用是当开发中出现异常情况时，可根据指示灯的状态快速定位问题。

**4 智能家居语音主控的软件设计**

语音主控的软件设计主要是针对AIUI模块和CC2530模块的程序设计。其中AIUI模块CPU采用全志R16一款4核cortex-A7智能硬件处理器，其处理能力远超过CC2530模块。因此，语音主控的主要处理都放在AIUI模块中，包括语音识别、与阿里云服务器通信、与本地手机APP交互、数据的存储等，而CC2530模块只作为ZigBee的协调器用来组建ZigBee网络和负责网络中数据的传输。



如图Z所示，语音主控是智能家居系统的核心，是整个系统的大脑，处于系统的控制层，其具体功能需求在本论文第二章节的有详细分析，根据分析将语音主控中的程序按功能模块进行划分可分为三大功能，如图Z所示。语音交互功能，主要是根据科大讯飞提供的开发接口进行集成开发，核心任务是负责语音识别结果的解析，根据解析结果可以控制智能家居中的相应家电，同时还接入了科大讯飞的AIUI内容平台，可以为用户提供多项休闲娱乐等功能，如听音乐、查天气等方便日常生活；通信功能是系统的核心，功能上相当于一个家庭网关，负责异构网络之间数据的转发和控制协议的解析，语音主控主要有AIUI语音模块与CC2530模块组成，其中AIUI模块上搭载有WiFi模块，可以通过WiFi实现与Internet网络的连接，而CC2530模块作为ZigBee网络的协调器，负责组建家庭内部的ZigBee无线传感网络并实现网络中数据的接收和发送，AIUI模块与CC2530模块通过串口进行联系；更新与维护是该系统的后勤保障，主要用于维护当前系统的稳定和将来系统的升级，功能上包括版本的远程升级、应用APP的静默安装和崩溃日志的上传，当软件在运行过程中出现运行时异常，会导致程序不能正确执行，出现闪退现象，对此一方面我们采用进程守护的方案，会立刻重新启动程序，另一方面，我们会记录保存崩溃日志并上传至服务器，开发人员可以在第一时间，得到错误反馈，然后进行修复相应的Bug，修复后的程序安装包会以自动更新升级的方式对原先的应用程序进行替换，因此这也是系统的一套闭环的优化方案。



根据上述分析，最终我们将语音主控划分为5个子应用程序，分别为：AIUIProduct、MasterServer、UartService、DaemonService和AutoInstall。AIUIProduct负责语音识别与交换，MasterServer负责与手机客户端以及服务器交换，UartService为串口服务负责与CC2530模块通信，而AIUIProduct与MasterServer



4.1 软件开发环境介绍

由于AIUI语言模块的操作系统采用Andriod4.4系统，可以根据业务需要支持二次开发，在功能方面可以灵活扩展，因此我们上述的程序本质上就是Android应用程序的开发。

4.1 语音识别

AIUI语音模块是科大讯飞提供的在语音识别方面软硬一体的解决方案，内部集成了语音采集与处理功能并封装了语音交互的复杂功能，为开发者提供简单易用的接口。开发集成模式如图所示，开发者APP集成AIUIServiceKit，运行在AIUI模块上，从AIUI Service获取结果，进行解析处理。AIUIServiceKit中提供的AIUIAgent就是和AIUIService交互的桥梁，通过发送不同的AIUIMessage控制AIUI的运行， AIUI通过AIUIListener将不同的AIUIEvent抛出来给开发者程序解析。一般调用流程如下图所示。

开发者应用

AIUIServiceKit

AIUIService

AIUI

云端服务器

第三方

后处理服务

1. 软件开发集成模式
2. Integrated model of software development

在语音识别方面，AIUI模块支持听写识别、语义识别和离线识别三种方式。听写识别和语义识别都属于在线识别，需要将用户的语音信息采集并上传至讯飞云端服务器进行处理。区别在于听写识别是将采集的用户语音中包含的文字信息“提取”出来，把语音转换成对应的文字信息，开发者可以通过对文本进行解析处理。而语义识别是指将用户语音信息与讯飞语音平台上自定义的技能相匹配，在匹配成功的情况下会按照指定的协议返回Json格式的数据，开发者主要针对返回的Json格式的结果进行解析处理。离线识别不依赖于云端服务器，采用的是本地的离线识别引擎进行识别，与在线识别相比，离线识别功能上要弱一点，所以可能存在说法在离线命令词没有识别结果，在云端听写会正确听写结果。因此，为了提高体验效果，该系统采用在线与离线相结合的混合识别模式，在本地定义离线命令语法的同时在云端说法也有对应的语义，混合模式下，音频会同时送到本地离线引擎和云端解析，结果选择策略是哪个先返回有效结果，就采用哪个。



1. 混合模式语音识别策略
2. Speech recognition strategy in mixed mode

综上所述，语义识别功能最为强大，科大讯飞AIUI开放平台提供了一系列内置人工智能语义理解能力，称为开放技能，例如天气、音乐、闲聊、智能家居等。只要在云端配置了相应的技能，就会返回识别结果，本系统中对应的音乐、天气、智能家居中空调电视等均采用语义识别；听写识别将用户的语音信息直接转换为文本，开发者可以针对文本进行解析处理，识别处理更加灵活，例如当开放平台没有需要的技能时可以采用听写识别进行扩展；离线识别在功能上相对较弱，但最大的优点是不依赖网络，一些简单的指令可以利用本地识别引擎进行识别，保证系统即使在网络状态不佳的情况下，依然具有语音识别功能。

1）语义识别

语义识别功能最为强大，科大讯飞AIUI开放平台提供了一系列内置人工智能语义理解能力，称为开放技能，例如天气、音乐、闲聊、智能家居等。只要在云端配置了相应的技能，就会返回识别结果，语音主控中对应的音乐、天气、智能家居中空调电视等均采用语义识别。以智能家居空调为例，在云端配置了空调这一技能，当云端识别出用户该意图时会返回一段包含了语义分析结果的Json数据[9]。只需要调用其提供的接口并解析输出的Json数据即可，Json数据字段定义如表1所示。

1. 语义识别字段定义
2. Field definition of semantic recognition results

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名称 | 字段类型 | 是否必须 | 说明 |
| rc | Int | 是 | 应答码 |
| service | String | 是 | 区分不同服务 |
| operation | String | 是 | 设备的操作编号，取值只有SET |
| semantic | Object | 是 | 语义结构化表示各设备自定义 |
| text | String | 是 | 用户语音转写后的文字 |
| answer | Object | 否 | 对本次识别的回复 |

以“客厅空调温度设置为25度”这条语音指令为例，云端针对这一条指令识别后返回的Json数据如下：

{

"answer": {

"text": "已为您将空调温度设置为25度"

},

"operation": "SET",

"rc": 0,

"semantic": {

"slots": {

"attr": "温度",

"attrType": "Integer",

"attrValue": "25",

"location": {

"room": "客厅",

"type": "LOC\_HOUSE"

}

}

},

"service": "airControl\_smartHome",

"text": "客厅空调温度设置为25度"

}

可以从semantic解析语义中room字段得出设备的房间为客厅，attr属性名称得出本次操作为温度调节，attrValue属性值得出，温度调节为25度。拿到上述信息后，就可以得到对应的控制指令，并发送至对应设备。

2）听写识别

听写识别将用户的语音信息直接转换为文本，只需针对文本进行解析处理，识别处理更加灵活，例如当开放平台没有需要的技能时可以采用听写识别进行扩展。语音主控中智能家居中灯的控制、交互场景动态切换、网络配置与查询、唤醒词动态修改、以及APP的语音升级均采用听写识别。听写识别最大的好处在于对于家中的电器可以根据自己喜好命名，如将客厅灯命名为大白，用户说出语音指令“打开大白”，听写识别将语音转为文字，程序中只需进行文字匹配，即可准确识别，打开客厅灯，而采用语义识别甚至都无法弄清用户的意图。

3）离线识别

离线识别在功能上相对较弱，但最大的优点是不依赖网络，一些简单的指令可以利用本地识别引擎进行识别，保证系统即使在网络状态不佳的情况下，依然具有语音识别功能。离线功能的核心是编写的离线命令词语法，离线命令语法是使用巴科斯范式（BNF）描述语音识别的语法。语法文档被编译成识别网络后，将被送往语音识别器。语音识别器提叏输入语音的特征信息并在识别网络上进行路徂匹配，最终识别出用户说话的内容。因此语法是语音识别系统的输入之一，它是现阶段语音识别得以应用的必要条件。语法文件包括HEADER和BODY，具体格式如下所示，字段定义如表2所示。

!grammar main;

!slot <sleep>;

!start <sleep>;

<sleep>:你去睡觉吧!id(2001)|你去休息吧!id(2001)|睡觉去吧!id(2001)

1. 离线语法字段定义
2. Field definition of Offline syntax

|  |  |  |
| --- | --- | --- |
| 关键词 | 描述 | 示例 |
| grammar | 定义语法名称 | !grammar main |
| slot | 声明槽 | !slot < device > |
| start | 定义开始规则 | !start < device > |
| id | 定义说法所对应语义返回值 | < device >:打开客厅灯!id(2001) |

离线识别中一个强大的功能是声明的词槽支持动态修改，用户可以在文档主体定义规则，也可以通过开发接口中定义规则，即用户可以在语法中声明某个槽而无须在文档文体中定义，程序运行过程中通过Aitalk SDK进行动态添加和删除槽中所含的条目。语音主控利用该功能在每次更新家庭中的电器设备时自动更新离线识别的词槽。

4）提高语音识别准确度的措施

对于在线识别（包括听写识别和语义识别）可以通过上传识别热词的方式，提高识别的准确性。识别热词，顾名思义可以让AIUI在识别时优先识别成热词中的内容。AIUI热词分为两种，一种是用户热词，通过指定的命令上传， 仅对上传该热词的设备生效。另一种是应用热词，通过AIUI后台应用管理界面中上传，对使用APPID配置的所有设备生效。两种类型热词共同生效，互相不会覆盖，同类型热词第二次上传会覆盖生效。因此，在语音主控软件设计上我们充分利用两类热词。在应用热词方面，参考目前用户的命名习惯和目前市面上智能家居产品的命名规则，增加100+应用热词，涉及常用操作指令、楼层、房间、设备、情景模式和其他；在用户热词方面，在用户每次更新电器的时候，会将该用户下所有设备的名称和情景模式的名称作为热词，按照规定格式上传。在实际应用过程中，两类热词的使用可以显著提高识别的准确度，尤其对于一些偏僻的词汇，效果更佳。