

需求分析文档

0.市场调查

在项目开始前，我首先对同学们对实时语音助手的需求情况进行了市场调查，调查问卷设计如下：

10:38 4G HD

蓝牙 电量

预览问卷

...

🎯

智能语音助手需求分析

一款PC使用的定制化语音助手需求分析

* 1. 请问您之前对智能语音助手有了解或使用过相关产品吗，比如iPhone上的Siri,或者Windows上的CORTANA?

☐ 有了解，但没有使用过

☐ 有了解，并使用过

☐ 完全不了解

* 2. 如果有可能的话，您想要一个定制化的全功能语音助手吗，比如《钢铁侠》里的贾维斯或星期五?

☐ 想

☐ 不想

* 3. 如果您使用过Siri或其它语音助手，您希望在哪些方面有所改进?

☐ 我非常满意，不需要任何改进

☐ 我希望可以加入更加丰富的功能

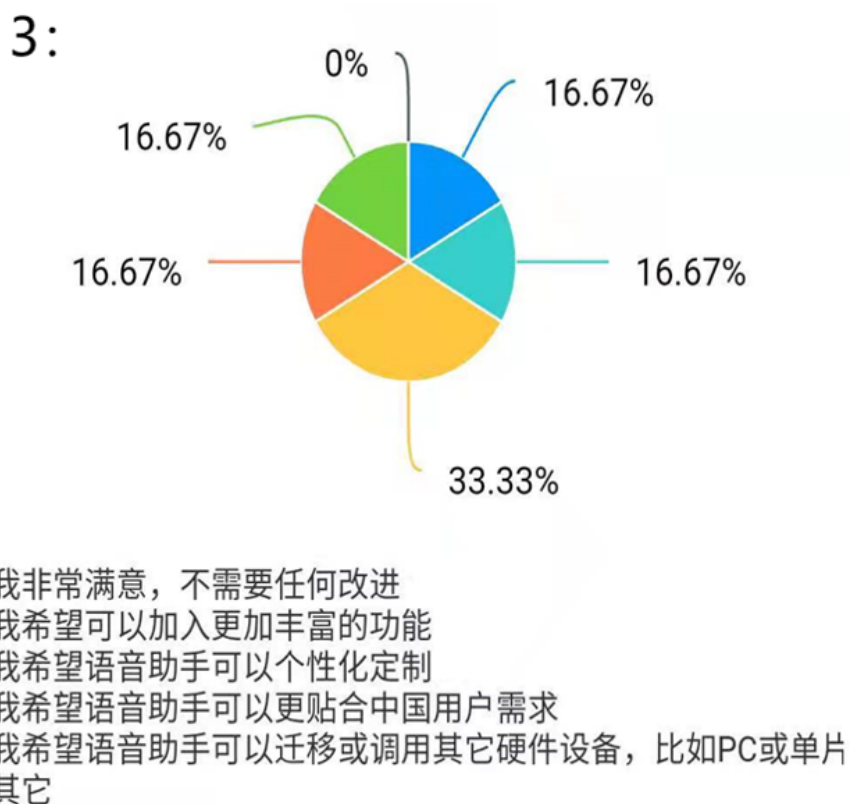
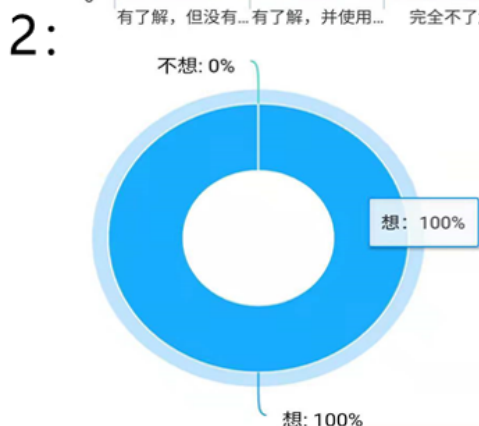
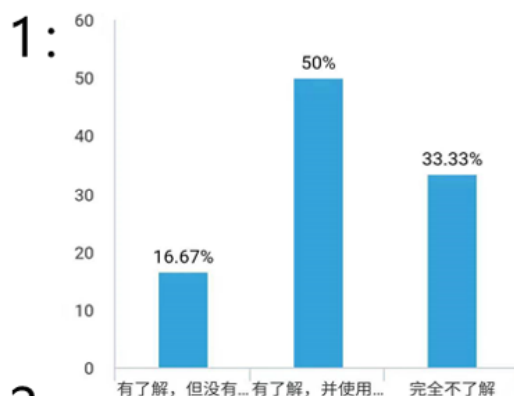
☐ 我希望语音助手可以个性化定制

☐ 我希望语音助手可以更贴合中国用户需求

☐ 我希望语音助手可以迁移或调用其它硬件设备，比如PC或单片机

☐ 其它

收到反馈结果如下：



从反馈结果中可以看出, 用户对更加智能, 更加定制化的语音助手有较强烈的需求。

1.文档概述

1.1目前iphone上Siri的功能:

- 实现收发语音消息
- 简单的打开指定应用
- 拨打电话, 联系其他用户
- 可以进行人性化聊天以及进行简单的信息搜索比如查天气

1.2 Siri的局限性:

- 目前的Siri只有手机版本, 并没有PC版本的, 因此也无法在硬件上有更多扩展空间
- Siri功能较为固定, 不能根据用户喜好进行拓展
- 不能照相以及进行图片识别等与SOTA的计算机视觉算法结合起来的高级图像应用
- 只能简单的打开应用, 不能对应用进行进一步复杂的操作

1.3搭建Siri Online的意义:

- 实现PC版本的Siri, 大大扩展Siri对软硬件的控制调用能力, 实现更多更丰富的功能

- 实现以语音交互为基础的平台化设计，实现定制化的语音调用功能，针对不同用户实现个性化设计
- 实现了Siri对复杂应用的深层次调用，提高了Siri的智能性，改善用户体验
- 为Siri增加图像识别模块，实现更高层次的人机交互
- 为Siri增加了富有中国特色的对对联、作诗功能，更加适应中国文化，更能赢得用户好感，更加适应中国市场需求
- 综合调用了语音识别，语音合成，自然语言处理，目标检测等多种人工智能模块，是对AI落地实践的一次小实验

2.项目简介

2.1整体说明

本项目实现了PC端的人工智能语音交互系统，通过调用科大讯飞，百度AI，腾讯云等多家平台的智能API实现了语音识别，语音合成，自然语言处理，目标检测等多种人工智能模块的整合改进，具体实现的 **语音功能** 有：

1. 实时语音聊天；
2. 通过语音计算数学题；
3. 通过语音对对联；
4. 通过语音作诗；
5. 通过语音查看天气

具体实现的 **视觉功能** 有：

- 在语音控制的基础上调用摄像头，实现物体种类识别。可分为实时视频识别与远程图像识别两种模式。

同时本设计基于平台化的设计思想，通过虚拟API接口的方式，实现了语音交互平台对多种功能的调用组合。由于PC端相较于移动电话端所拥有的巨大拓展空间，本设计实现了对具体应用的 **深度调用**，具体表现为

- 通过语音控制爬虫搜索并播放B站视频
- 通过语音打开QQ，并通过语音与指定对象进行聊天。

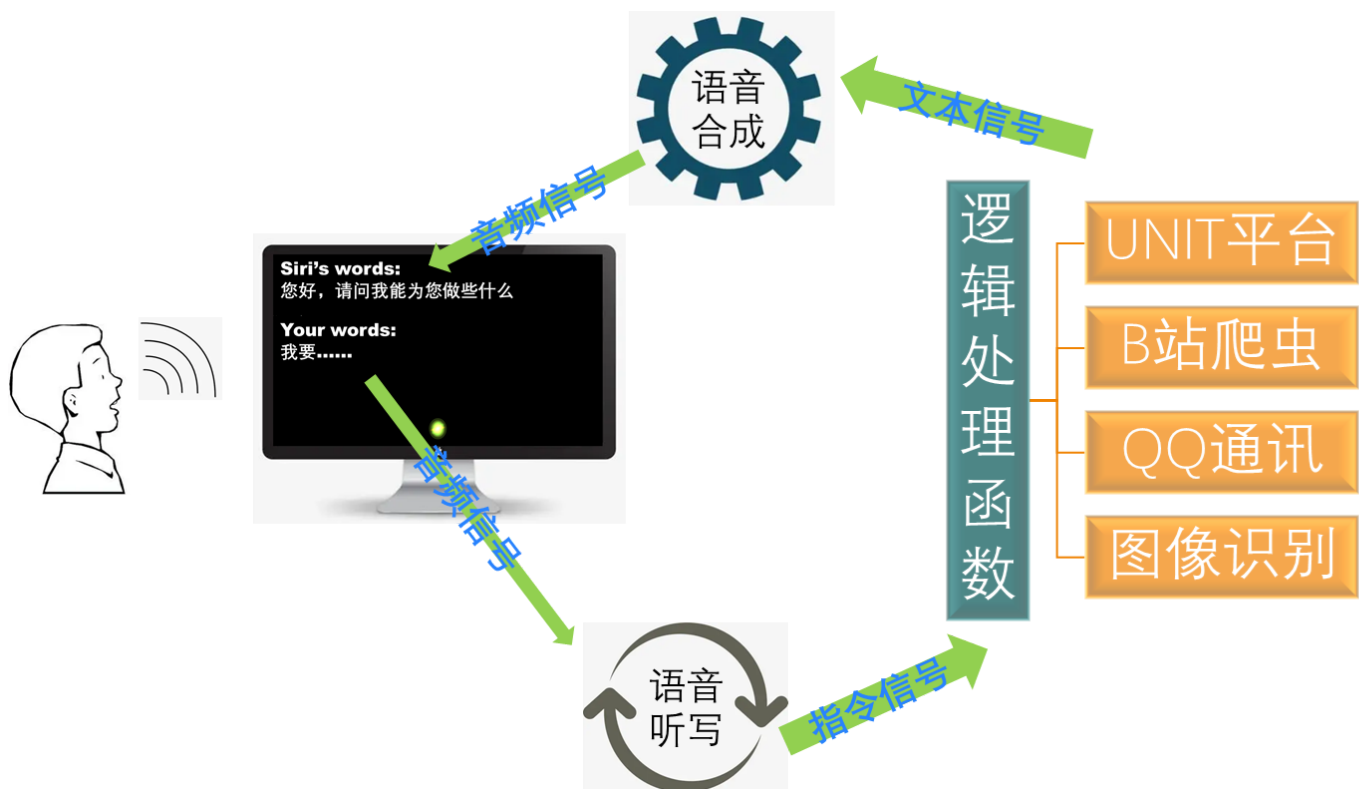
2.2 模块搭建：

本项目属于人机交互项目，主要包含 **语音模块**，**网站交互模块**，**应用交互模块**以及**前端展示模块**，具体模块组成如下所示：

本项目的基本模块调用情况如下：

- 讯飞开放平台（实现对话平台的搭建）：
 - 语音听写（流式版）
 - *语音合成
- 百度AI开放平台（实现聊天，天气查询，对对联等语音交互功能）：
 - UNIT平台
- 视觉模块：
 - 实时视频检测（通过yolo-tiny网络实现）
 - 实时图片检测（通过百度AI平台通用物体分析平台实现）
- 应用模块：
 - bilibili视频爬虫模块
 - QQ客户端消息收发模块
- 前端模块：
 - 前端语音交互信息可视化模块（通过html和css等前端语言实现）

2.3 项目的模块示意图



针对上面的模块设计图，有以下说明：

启动模块时首先打开交互界面，语音听写采用5秒一段的轮询调用，每次收到一段音频信号就转换成指令信号传递给逻辑处理函数。逻辑处理函数根据提取到的指令分别调用UNIT平台、B站爬虫、QQ通讯、图像识别等功能，之后调用的功能分别根据各自的模块执行操作，操作过程中产生的反馈信号以文本形式传递给语音合成函数，之后再由语音合成函数输出语音信号完成交互。

3.具体需求

3.1 需求总览

本项目初期定下了如下分析的需求：

1. 智能化语音交互平台
2. 美观简洁的交互界面（吸取Siri风格）
3. 强大的可扩展性和高可用性
4. 流畅的具体功能实现

3.2 需求实现优先级

根据需求对项目的重要程度，将所有功能分三个开发周期迭代实现

1. 语音交互功能的实现，完成讯飞开放平台语音合成和语义听写模块的操作，使得音频信号可以转化成文本信号，文本信号也可以转化成音频信号。
2. 搭建前端网页，完成逻辑控制，构建语音交互平台框架，实现main函数，形成功能函数的接入点，即虚拟API接口。
3. 分别完成UNIT平台、B站爬虫、QQ通讯、图像识别等具体功能，并将完成的功能函数接入虚拟API接口，实现平台对具体函数的调用。