奥图项目计划书

版本：0.1

[小i机器人](http://www.xiaoi.com/)

文档版本历史

|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 版本 | 内容说明 | 撰写人 |
| 2016-11-19 | 0.1 | 方案初稿 | 杨超 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**目 录**

[1. 方案目标 4](#_Toc341456838)

[2. 方案设计思路 4](#_Toc341456839)

[2.1. 网络拓扑图 4](#_Toc341456840)

[2.2. 业务流程示意图 5](#_Toc341456841)

[3. 业务功能 5](#_Toc341456842)

[3.1. 核心功能 6](#_Toc341456843)

[3.2. 硬件功能 6](#_Toc341456844)

[3.3. 增值功能 7](#_Toc341456845)

[4. 项目实施 7](#_Toc341456846)

[4.1. 工作任务 7](#_Toc341456847)

[附件一：服务器相关说明 8](#_Toc341456848)

# 方案目标

在奥图眼镜中集成智能语音控制系统，实现更智能更丰富的语音指令，使得指令内容不再仅仅局限于简单的几个硬件能力和和固定的语音指令。还能基于自然语言拥有更多样的指令和基础的生活服务，给予用户更好的使用体验。

# 整体说明

## 整体功能描述

|  |  |
| --- | --- |
| 功能 | 简介 |
| 拍照 | 拍摄照片，在相册中查看，能实现分享到微信和新浪微博 |
| 录像 | 录制视频，在视频中查看，能实现分享到优酷 |
| 打/接电话 | 连接手机后可以实现拨打和接听电话 |
| 通话记录 | 可以查看通话信息 |
| 发短信 | 连接手机后实现给指定用户发送短信 |
| 增强现实 | 特定场景实现AR功能 |
| 录音机 | 实现语言录制 |
| 导航 | 连接手机后可以实现导航功能 |

奥图

## 业务流程示意图

如上图所示，手机客户端采集到音频后，经过音频压缩后通过http方式传输到语音云接口进行识别，识别结果递交智能引擎进行语义理解和解析，智能引擎会根据理解结果生成相应的指令，手机客户端就可以根据指令执行相应的操作，并将需要音频合成的内容提交语音云接口进行语音合成，并播报相应的结果。

# 业务功能

IPTV第一期项目主要处理核心功能以及硬件功能，基于自然语言的切屏、点播、直播以及EPG信息的查询操作等，增值功能暂只提供电视节目表。

## 核心功能

**VOD点播：** 功能包括根据片名点播电影电视

播放指定电视剧某一集

播放指定演员出演的电影电视剧

动态更新片名信息，模糊匹配片名

进入一级、二级分类

动态电影电视剧关键字搜索

可购买需要收费的内容

**电视点播：** 频道上下切换

根据具体电视台名称切换

根据节目名称点播

返回先前观看的频道

根据时间+节目名称点播

## 硬件功能

根据最终确认的客户端中需要增加语音控制功能支持的相关功能和业务来定制特色功能，主要体现语音输入的便捷性；

**切换屏幕：**切换播放模式（手机IPAD等屏幕或电视屏幕）

**声音控制：**静音开关，支持上下文的音量控制（响一点/再响一点）

**开机：** 打开机顶盒

**关机：** 关闭机顶盒

**翻页：** 上一页、下一页、首页、最后一页、指定页

**菜单调用：**进入IPTV EPG 一级、二级分类

**播放控制：**快进、快退、暂停、播放、重放、退出、X倍快进/快退（2x、4x、8x、16x）

## 增值功能

**节目查询：**查询指定电视台指定日期节目，查询指定电视台指定时间的节目，查询固定节目的播出电视台和时间，接口由IPTV提供

# 项目实施

## 工作任务

此次采用为IPTV定制语音操控服务，具体的内容包括：

* **任务1：现有手机端增加语音智能的支持。**通过相应的手机版本sdk将小i的语音识别、语音合成、语义理解，智能交互等功能增加到客户端，以便为手机客户端增加语音智能控制功能。
* **任务2：知识和服务二次开发**：根据业务内容开发语法文件训练识别引擎，提高识别效果，协商制定语音操作指令库，整理相关指令，对接EPG信息，并维护整理后台相关知识库等。
* **任务3：机顶盒内嵌语音智能组件开发**：后续高端机顶盒内开发一组件用于通过语音直接操控机顶盒相关功能。(后期)

# 附件一：服务器相关说明

* 小i语音服务器由语音识别服务器+语音合成服务器+语义理解服务器+数据库服务器组成，考虑系统的健壮性和稳定性，后三类服务器都采用双机热备的方式（测试时可单机）
* 语音是电视机操控中的辅助手段，因此一般采用<5%的使用率来计算服务器配置。因此，若开放的用户数为4万，则服务器需配置4万\*5%=2000用户并发量，针对**限定领域（注意，目前的操控和定向的查询属限定领域，若非限定领域，服务器语音识别并发数会急剧下降）**，每台服务器可支撑150/200个并发流。