## 《盲人智能辅助系统》系统设计

**小组成员：郭艺君 何妍 刘恩嵩**

1. **系统规范**
2. **应用场景**

在有WIFI的室内环境下，为盲人提供体验性良好流畅的语音读书功能，缓解目前我国盲人图书资源稀缺以及不均衡的问题。

1. **系统功能**
2. 通过摄像头捕捉当前画面，利用深度学习技术提取文字并合成语音
3. 能对图书位置进行提示，以获得更完整的页面
4. 将读书记录数据传至云端，方便保存阅读记录
5. **系统指标**
6. **文字识别的准确率**

文字识别（包含标点符号等）的准确率大于等于98%

1. **语音合成的流畅度**

识别后的文字转换成的语音需要流畅、易于理解，评级划分为差、一般、良好、优秀。

差：语音流畅度差，无断句与节奏

一般：语音流畅度尚可，能进行基本的断句

良好：语音流畅，断句与节奏基本符合平时交流

优秀：在良好的基础上，富有情感，与现有听书软件的人工音频具有相近或同等水平

1. **辅助阅读的有效性**

对书籍页面提取的正确性（即在特殊情况下能否正确辅助阅读）大于等于98%。

1. **云端数据的正确性**

能够正确地把当前阅读记录信息上传到云端，正确性大于等于98%。

1. **用户手册**
   1. **引言**

该项目为为盲人设计的智能辅助系统，旨在为盲人提供体验性良好流畅的语音读书功能，缓解目前我国盲人图书资源稀缺以及不均衡的问题。

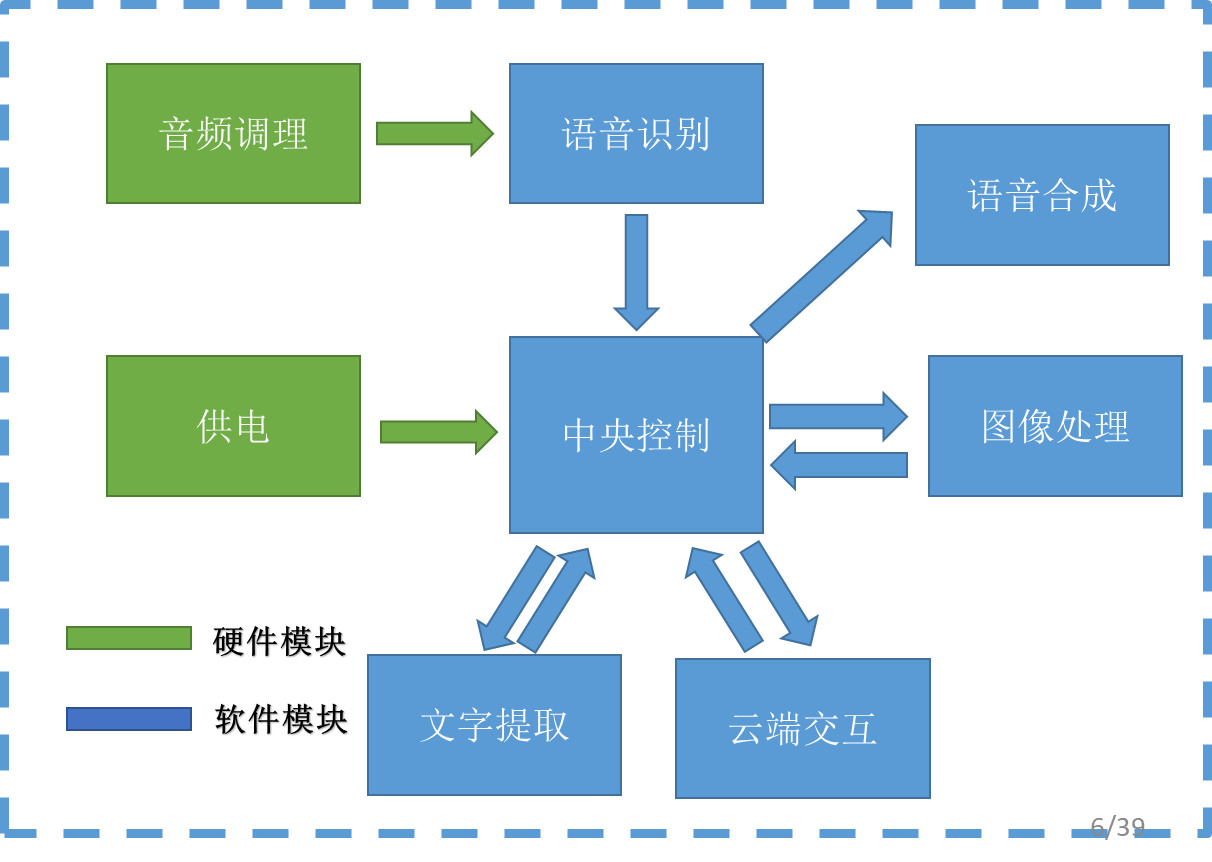
* 1. **软件概述**
     1. **目标：**为盲人提供便捷、流畅的智能听书服务。
     2. **功能：**引导盲人，使系统摄像头捕捉到正确、完整的图书页面，利用深度学习技术提取文字并合成语音，同时将读书数据上传云端。
     3. **性能：**文字识别准确率、页面提取正确性、云端数据正确性大于等于98%；语音流畅度优秀；续航时间大于等于1小时。
  2. **运行环境**

**硬件：**软件系统运行于树莓派开发板上，外存大小为32GB，使用USB接口输入输出。

* 1. **使用说明**
     1. **安装和初始化**：长按开机键开机，系统即开启。登陆账户后，用户数据将被同步。
     2. **功能选择**：使用语音识别来实现用户功能选择。即，用户说出“开始读书”、“暂停读书”、“结束读书”即可选择相应功能。
     3. **关闭：**长按关机键关机，系统即关闭。

1. **总体方案**

下图为我们的系统软硬件框图，其中绿色部分为系统的硬件模块，蓝色部分为系统的软件模块。语音通过音频调理电路转换为电信号，再经过语音识别传入中央控制进行功能选择。在成功识别指令后，图像处理模块将摄像头传入的图片进行预处理，以便提高深度学习文字识别API的准确性。在成功提取文字后，调用语音合成模块进行书籍的语音播放。云端交互模块将识别出的书籍相关信息以数据表的形式保存，并形成人机交互的网站界面。

****

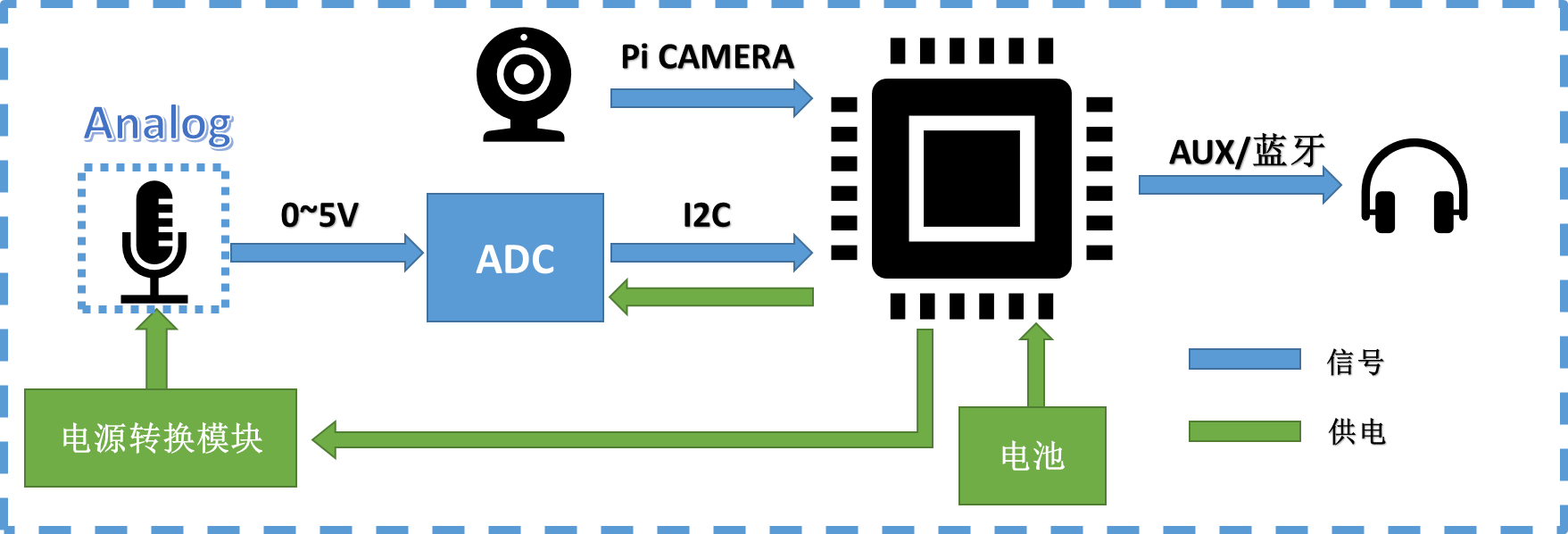
**系统软硬件框图**

1. **硬件详细设计**
2. **硬件模块框图**

对于开发板的选择，从系统功能出发，我们希望具有更加灵活的图像处理以及语音信号处理选择，因此选择了树莓派作为开发平台，同时树莓派内置了蓝牙、WIFI模块，满足系统通信要求。

作为可穿戴设备的产品定义，我们希望所有实体模块尽可能的小，而树莓派支持的微型麦克风种类少而且效果比较一般，因此将模拟电路设计放在语音调理电路。语音调理电路输出0~5V的电压信号给ADC模块，ADC模块通过I2C协议将信号传给树莓派，以便进行下一步的语音识别。在完成文字识别以及语音合成后通过AUX有线或者蓝牙无线进行语音播放。

系统计划使用电池为树莓派供电，同时由于语音调理电路使用到的运放需要双电源供电，需要一个电源转换模块将单电源转换为±12V的电源为模拟电路供电。

****

**硬件模块框图**

1. **模拟电路设计**

正常的音频信号频率范围为300Hz~3400Hz，根据奈奎斯特采样定理，理论上高于6800Hz的采样频率均可以满足音频信号的要求。下面的两个表为采样频率与采样精度和对应的语音信号要求的关系[[1]](#footnote-1)，对于本系统而言，最基本的8KHz与8Bit信号位宽可满足要求。

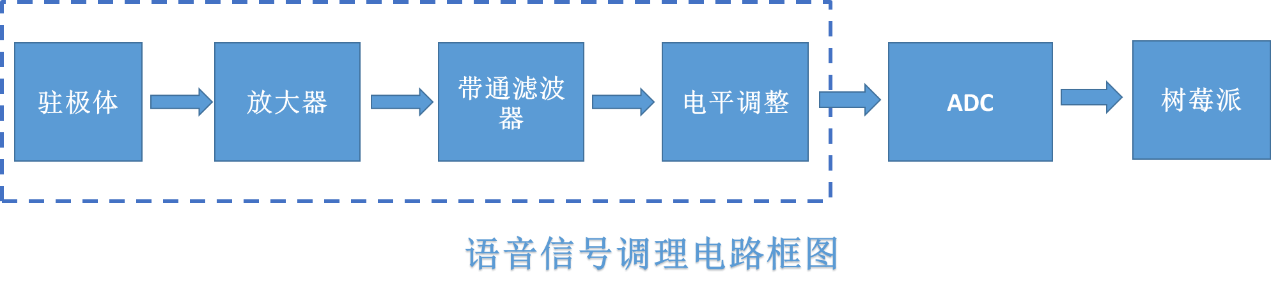
|  |  |
| --- | --- |
| 采样频率/Hz | 品质级别 |
| 8000 | 电话，对于人声已经足够 |
| 11025 | AM电台（低端多媒体） |
| 22025 | FM电台（高端多媒体） |
| 32000 | 好于FM电台（标准广播采样率） |
| 44100 | CD |
| 48000 | 标准DVD，及专业音频领域 |
| 96000 | 蓝光DVD |
| 192000 | 声卡支持的采样率 |

**常用的采样率及对应的品质级别**

|  |  |  |  |
| --- | --- | --- | --- |
| **位深度/位** | **品质级别** | **采样振幅值** | **动态范围/dB** |
| 8 | 电话 | 256 | 48.16 |
| 16 | 音频CD | 65536 | 96.33 |
| 24 | 音频DVD | 16777216 | 144.49 |
| 32 | 最佳 | 4294967296 | 192.66 |

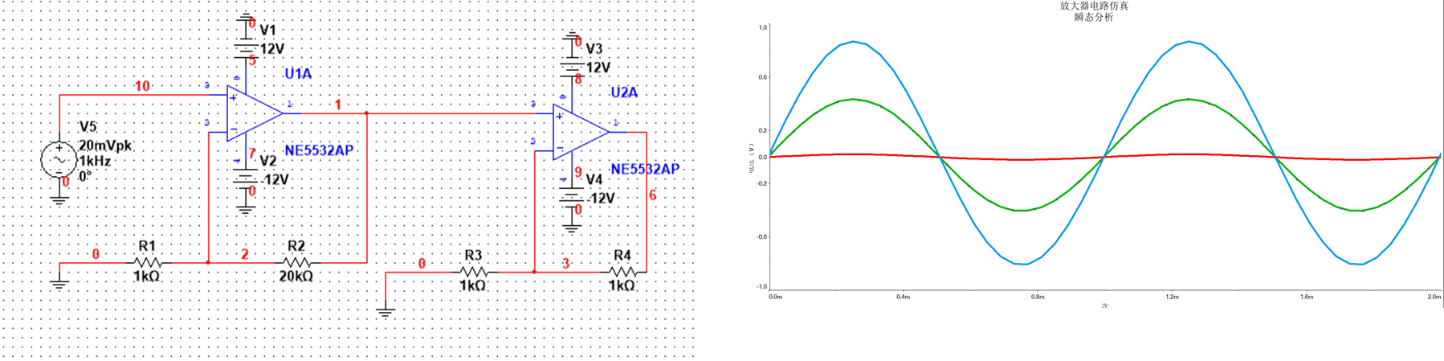
**采样位数及对应参数的关系**

语音调理电路可划分为以下模块，驻极体输出微弱的几毫伏到几十毫伏的电压，经过放大器放大200倍，再通过带通滤波器将信号频带限制在语音信号的300Hz~3400Hz内，由于ADC模块的输入是0~5V，再通过电平调整将输出信号调整至0~5V。ADC模块使用的是PCF7591，其采样频率为8KHz，信号位宽8Bit，据上述分析满足语音信号的要求。ADC模块输出通过I2C协议传至树莓派，保存为PCM文件，方便下一步的语音处理。



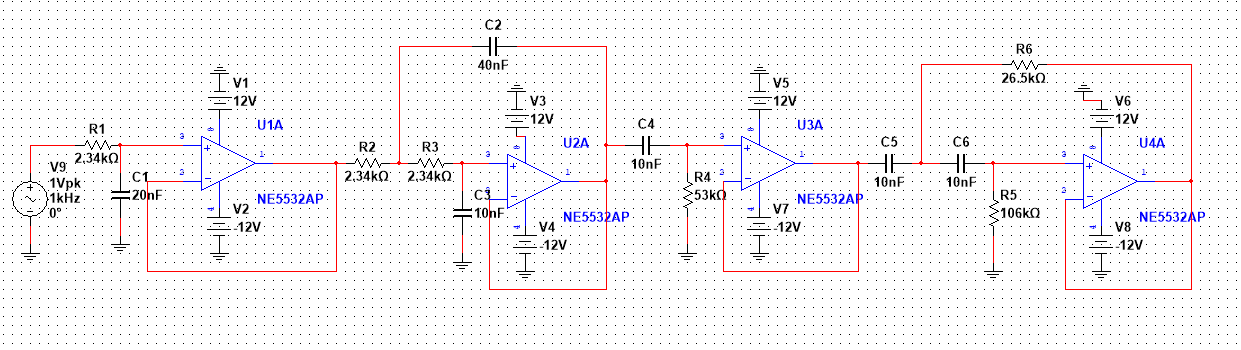
* 1. **放大器电路**

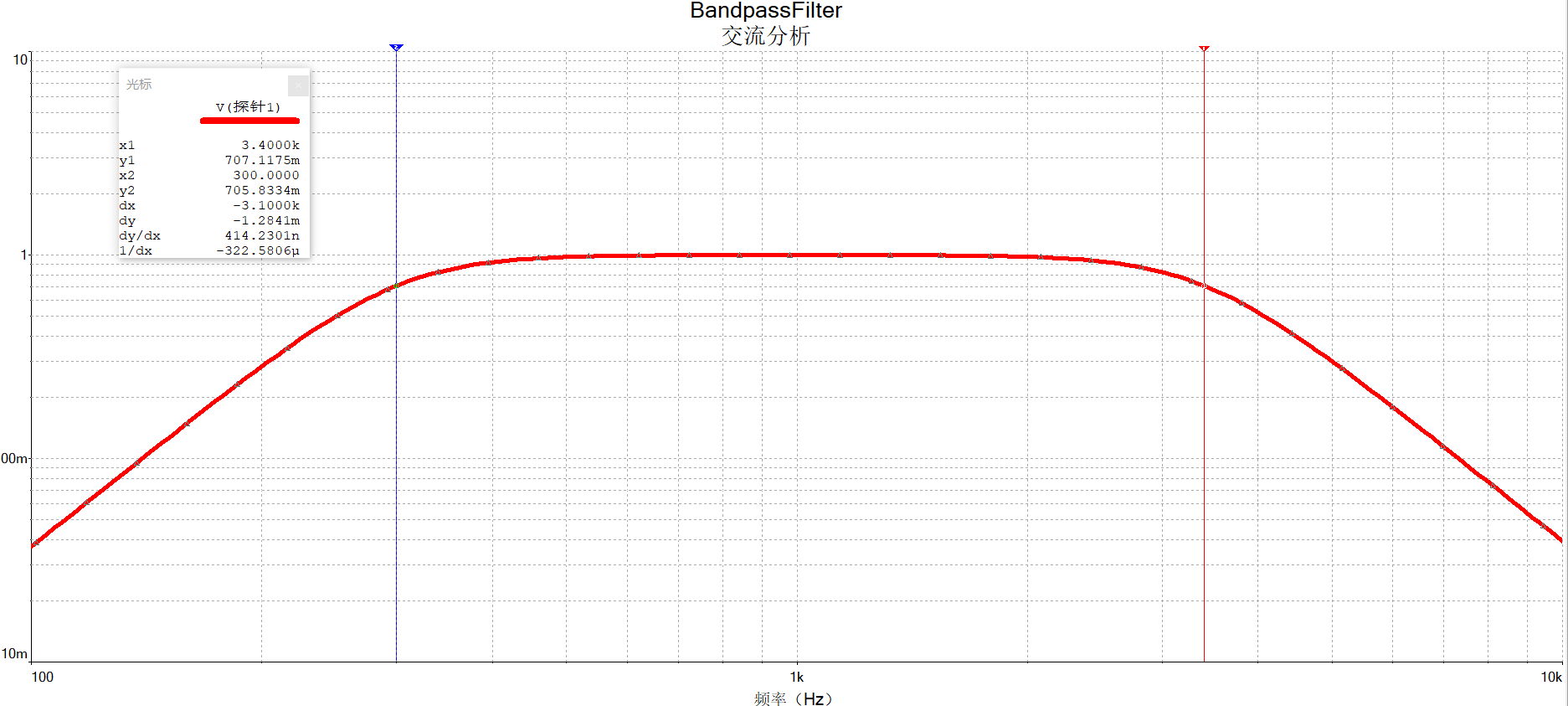
对于语音信号调理电路模块，运放需要选择满足音频信号处理要求的电路，这里选择NE5532作为语音信号调理电路的运算放大器。放大器电路采用两个同相放大器级联，放大倍数为200倍，电路原理图以及仿真结果如下图。



* 1. **滤波器电路**

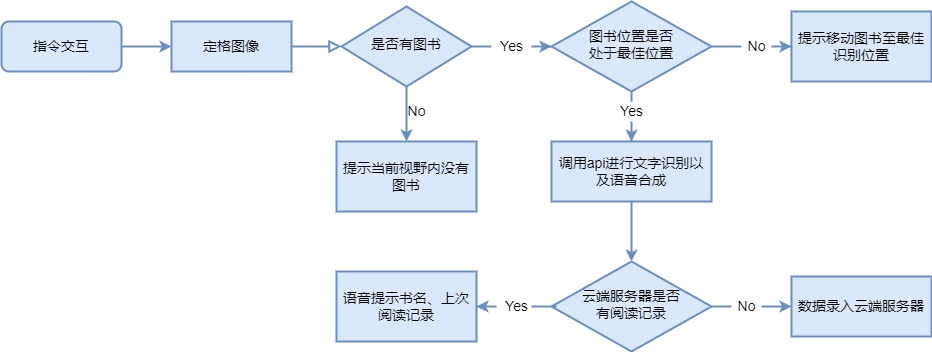
滤波器电路功能要求带通滤波器，其通带为300Hz~3400Hz，阻带衰减12dB/倍频。由于其通带跨度较大，因此选择低通滤波器与高通滤波器级联实现，而不是直接设计带通滤波器。这里采用三阶节的低通滤波器与高通滤波器级联实现此带通滤波器，元器件计算取值使用巴特沃斯多项式来实现，电路原理图即仿真结果如下图。可见仿真结果满足设计要求。





1. **软件详细设计**
   * + 1. **软件总体设计**

软件系统启动后，摄像头定格当前帧图像，判断当前视野内是否有图书，如果没有图书，即提示使用者当前视野内没有图书，如果有图书，进一步判断图书位置是否处于最佳位置，即当前书籍页面是否完整，如果不完整，则对移动图书进行提示，移动至页面完整后重新进行判断，进而调用API进行文字识别以及语音合成，再提取到当前阅读的书籍一些信息后，与云端服务器进行交互，判断是否有阅读记录，如果有阅读记录，则通过语音提示书名、上次阅读记录，如果云端服务器没有记录，则将当前书籍信息录入云端服务器。



* + - 1. **软件详细设计**

软件部分的详细设计主要分为四个子模块，文字识别、图像预处理、语音识别和语音合成，由于本系统软件部分较多，因此在系统设计中，我们对软件中的一些关键算法进行了详细的验证与比较。

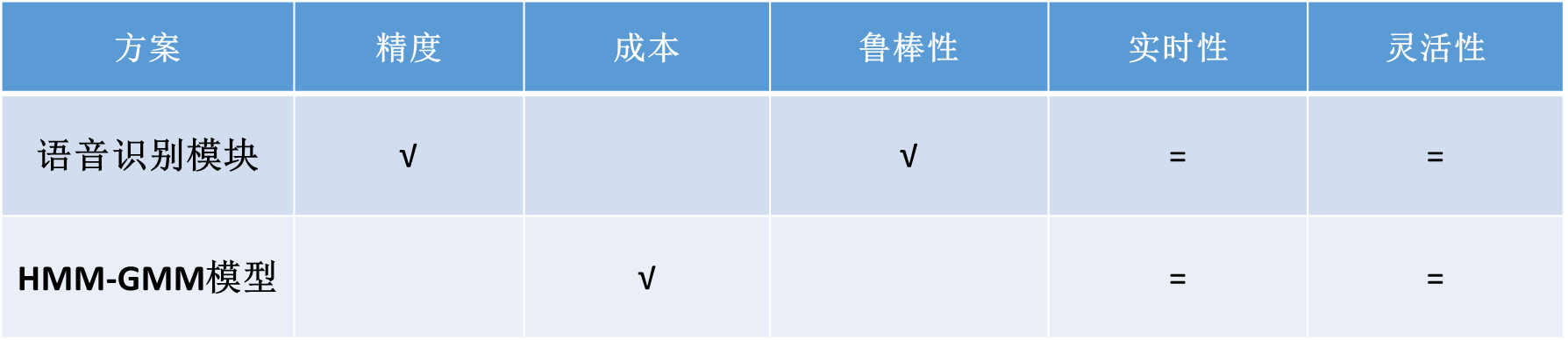
* 1. **语音识别**

对于语音识别模块，由于系统对于语音识别只是指令交互作用，因此并不需要通用的语音识别算法，只需要对固定的几条语音命令有效识别即可。这里暂时确定了两种方案，分别是语音识别现成模块以及HMM-GMM模型。语音识别现成模块指市面上现有的硬件模块，一般价格在几十块左右，可以手动写入识别命令。

HMM-GMM模型（混合高斯-马尔科夫模型）是一种基于概率的语音识别声学模型。传统的语音识别框架定义声学模型输入为，对应的句子，语音识别的任务就是求概率最大时对应的字序列。要计算，可以利用贝叶斯公式：

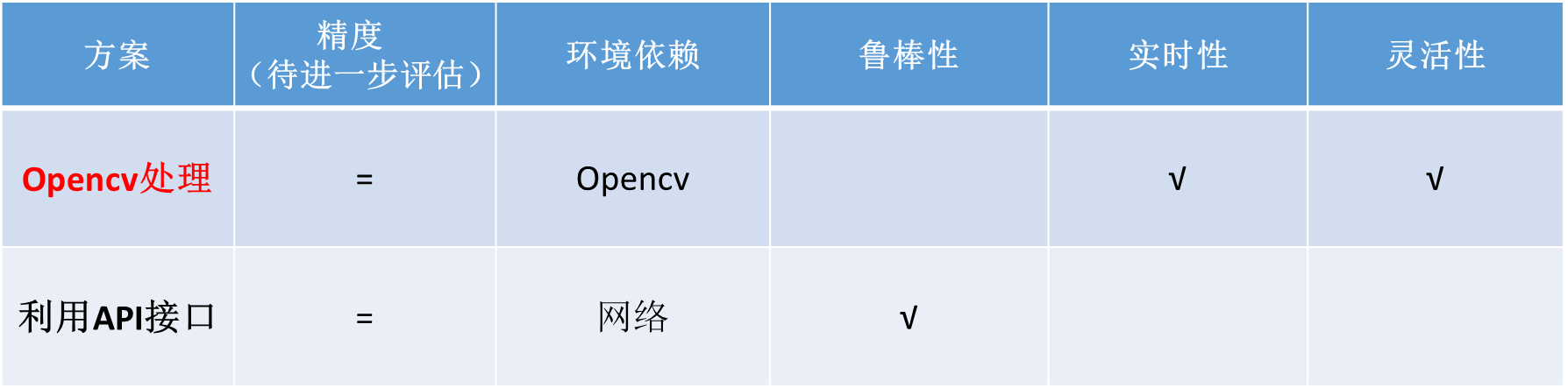
其中在计算时可以忽略，这样就得到了两部分 ，分别对应于传统语音识别框架中的声学模型和语言模型，前者的任务是计算给定文字之后发出这段语音的概率；后者表示某一字词序列发生的概率。HMM-GMM模型在此基础上进行相关操作去寻找对应的最大概率值，就获得了识别结果。

两种方案对比情况如下图。

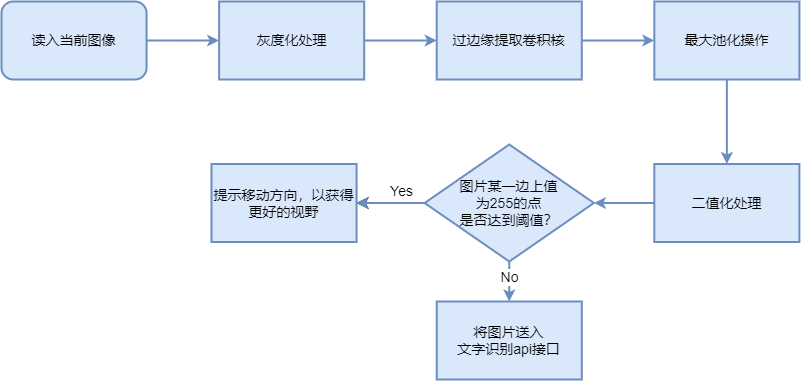


* 1. **图像预处理**

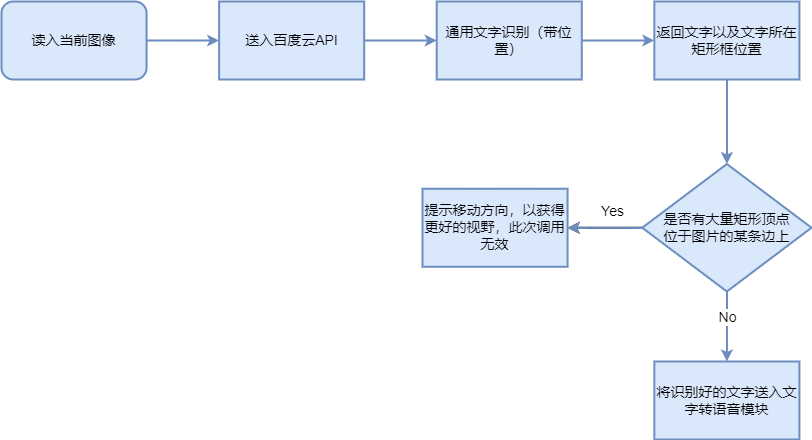
图像预处理的作用可以通过透视校正、二值化等操作提高文字识别的图片质量。在本系统中，图像预处理的另一个功能是提示书籍位置的移动，如何判断当前书的页面显示不完整是本系统的重要功能，这里也选择了两种方案，其原理都是类似的，即判断位于图片边界的文字数量是否达到阈值，如果达到阈值则认为其页面显示不完整，并提示朝着这一边的反方向移动图书。方案一是通过百度云文字识别API，其API包含返回文字位置，这样就实现了判断文字位于边界的问题；方案二是通过Opencv对本地图片进行处理，例如通过卷积、池化、二值化等操作，增强文字与背景的对比度，从而判断文字位于边界的问题。



两种方案的子程序图如下。



**方案一 Opencv本地处理**

****

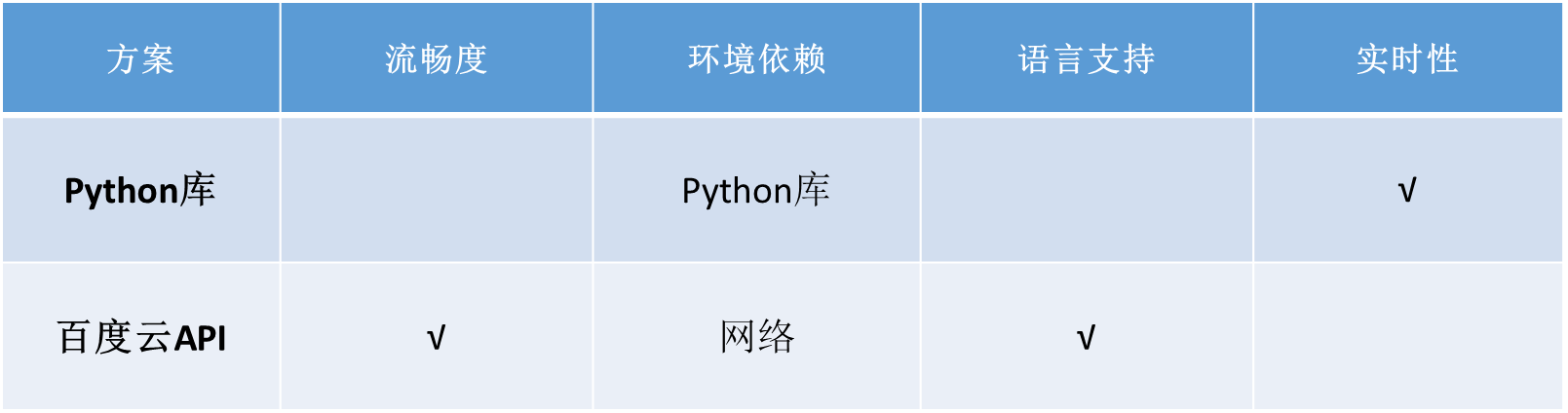
**方案二 利用文字识别API**

* 1. **语音合成**

语音合成模块同样分为两种方案供选择，一个是使用本地的python库，pyttsx3模块可以把文本转换成语音并播放出来，也可以保存为文件。但保存的文件为aiff格式，需要用AudioSegment库将其转换为mp3编码的文件。示例：

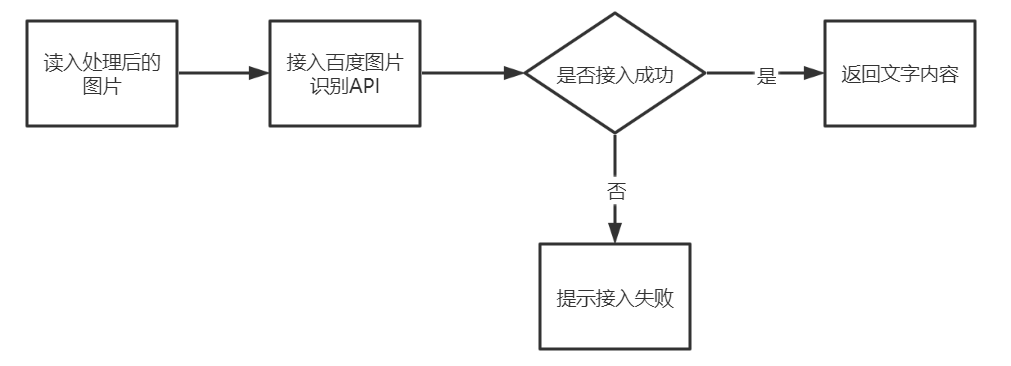


另一个是调用百度语音合成API来实现，其对比如下表。



* 1. **文字识别**

我们计划调用百度API实现文字识别的功能。系统框图如下所示：首先读入处理后的图片，接入百度通用图片识别API，若接入成功，即可返回识别的文字内容，若接入失败，则向系统提示接入失败。

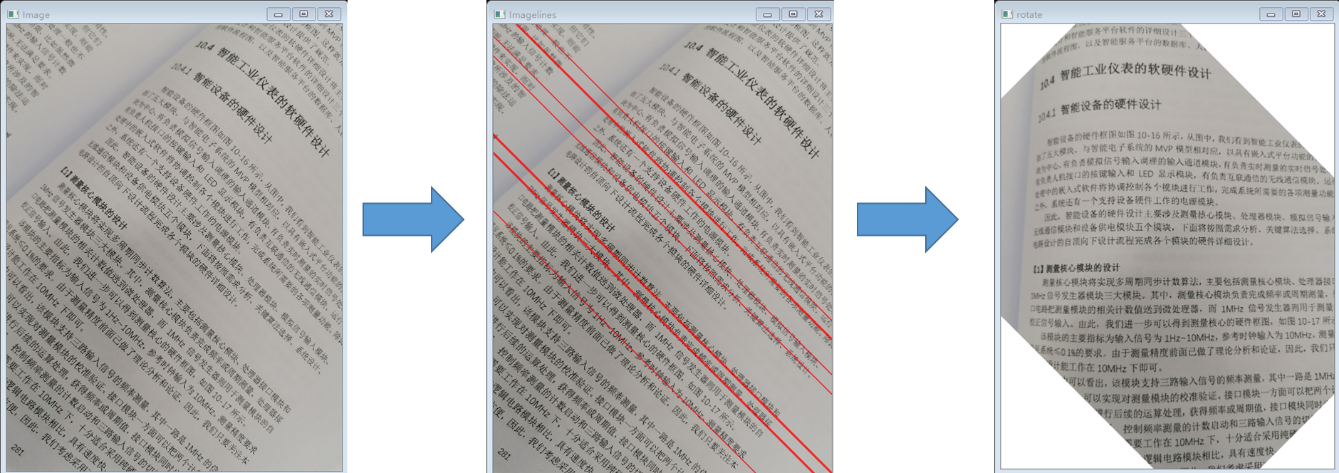


为帮助百度API提高识别准确率，需要使文字方向尽可能水平、图像分辨率尽可能高、文字尽可能清晰（文本字号不能过小）。在受限于摄像头分辨率的情况下，需要对图片进行预处理，使文本清晰，可以采用伽马变换、线性变换、二值化、倾斜校正等方法。

这些软件方案还需要后续的衡量来进一步确定具体使用哪一个方案。

1. **关键算法验证**

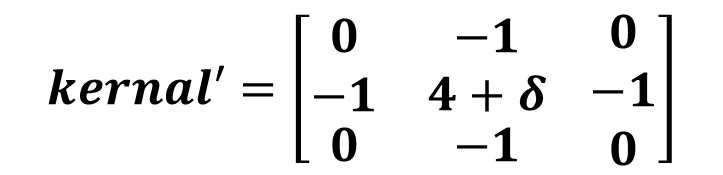
我们对图像预处理做了算法验证，针对的情况是书籍倾斜时，可见根据霍夫线检测的原理可以进行倾斜校正。其中关键参数在于霍夫线检测的阈值，阈值的含义是算法认为满足直线要求的点的数目，数目过大检测不到直线，数目过低直线数目过多不够准确，因此需要确定一个相对合适的阈值。



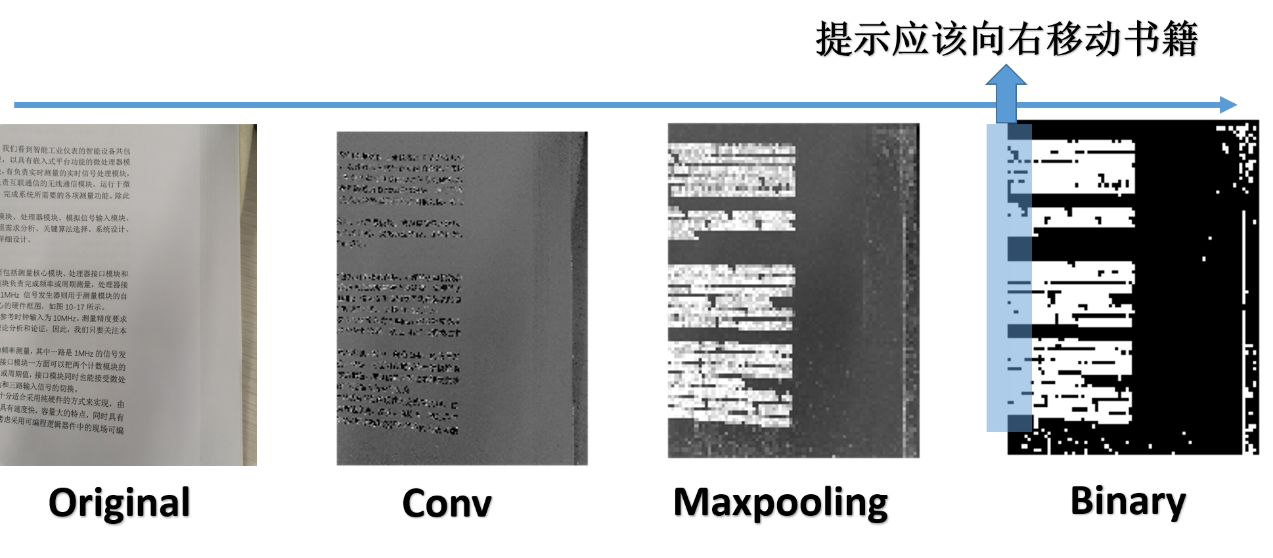
我们对Opencv本地处理图片从而进行书籍位置移动提示做了算法验证。该算法的详细流程如下：灰度化处理🡪过卷积核🡪最大池化操作🡪二值化处理🡪判断某一边上为255的点是否达到阈值🡪提示移动方向。卷积核的作用是边缘提取，池化操作的作用是更关注区域，即提取文字块，二值化处理则是起到效果增强的作用。

为了增强文字与背景的对比度，在卷积这一步使用的边缘提取卷积核其和大于0，这样就相当于对整体的值都进行了放大。

其卷积核如下图



其中与最后一步二值化处理的阈值有关，其与阈值呈线性关系，即越大，代表卷积核对图片矩阵整体值的放大越多，那么在最后一步进行二值化处理时，判断为文字的阈值也应该越大。本算法为系统的原创算法，其灵活性更高，也可以进行后续更改与调整以获得更好的效果。



1. **智能服务设计**
   * + 1. **数据采集接口设计**

为便于获取图像数据，设计数据采集接口：首先使用JSON打包拍照获得的图片信息，然后再通过API上传至百度百度云对JSON进行解析，并对数据进行分析，最后将分析后的数据输出，上传到云端服务器

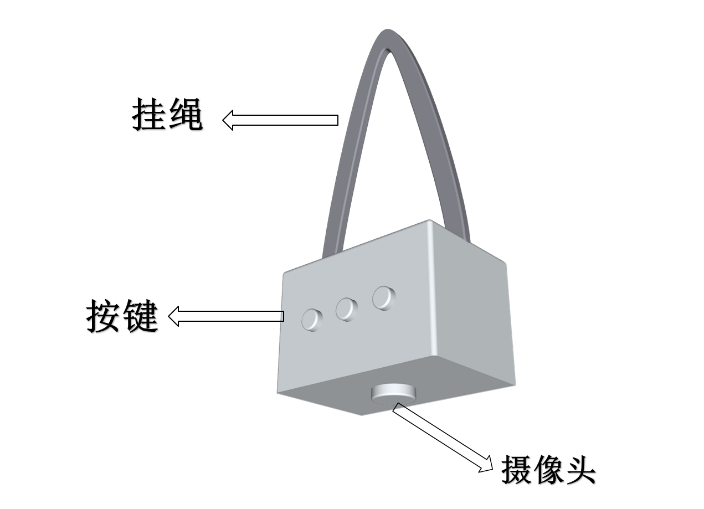
* + - 1. **人机交互界面**

为提高用户体验、提供多方面统计服务，我们将设计人机交互界面。我们将借助于云服务器搭建网站，显示相关信息，例如最热图书、阅读时长排行等。

* + - 1. **数据库设计**

为便于数据存储与管理，我们设计数据库。数据库引擎为数据的查询和操作提供高效接口，使用MySQL数据库管理系统，包含三个项目表，读书信息表、书籍内容表、系统用户表，其中读书信息表包含书名、阅读时长两个项目。

1. **工业设计**
   * + 1. **外壳示意图**

****

* + - 1. **制作方案**
         1. **PCB打版制作**
  1. **系统整体安装：**将PCB及所有器件安装成为一个系统，确保互相之间的连接
  2. **外壳制作与安装：**外壳大小基本与树莓派大小相同(85mm×56mm)，将外壳组装到系统外部。

1. **系统调试/测试方案**
   * + 1. **测试场景设计**

使用真人测试，测试环境为WiFi覆盖的室内环境。使测试者戴上眼罩模拟盲人，并随机选择一本书开启系统进行阅读。

* + - 1. **指标设计方案**
         1. **文字识别的准确率**

将提取文字结果与人工校正后的结果相比较，得到文字识别准确率，要求该准确率大于等于98%。

* + 1. **语音合成的流畅度**

测试者对语音信号的流畅性和可理解性进行评估打分（差、一般、良好、优秀），要求该结果为优秀。

* + 1. **辅助阅读的有效性**

特殊的图书位置情况下，系统能否给出正确的移动提示。要求最终能够获得正确的图书页面，且要求对书籍页面提取的正确性（即在特殊情况下能否正确辅助阅读）大于等于98%。

* + 1. **云端数据的正确性**

将云端保存的阅读信息与真实情况相比较，得到云端数据的正确性，要求该正确性大于等于98%。

* + 1. **系统功耗指标**

从电池充满电开始正常运行系统，直到系统电量耗尽，得到系统续航时间，要求该时间大于等于1小时。

1. https://zhuanlan.zhihu.com/p/492253579 [↑](#footnote-ref-1)