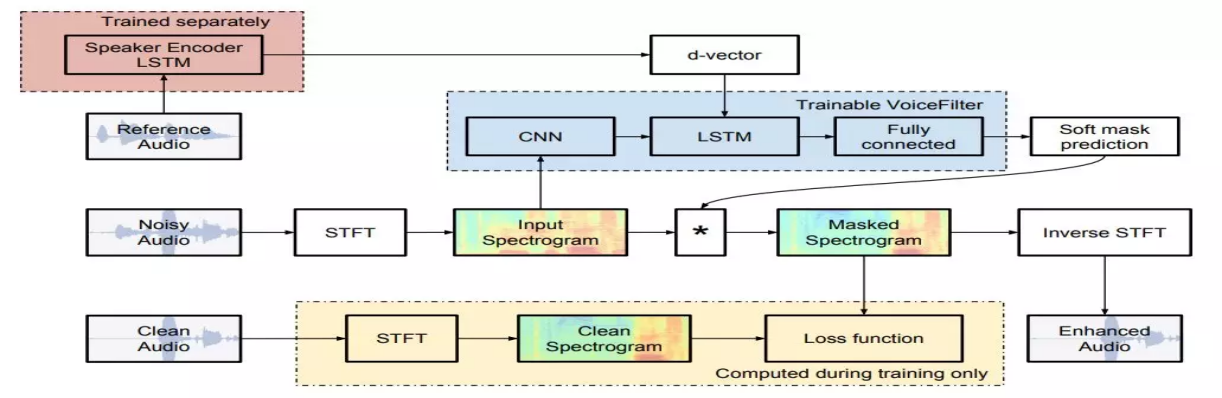
1. 话者分离



主要由两部分组成:声纹识别模块(红色) 声音过滤模块(蓝色)

声纹识别模块：它使用的网络是三层的递归神经网络加上GEE损失，输入1600ms的mel话谱，最后将三个d-vectrorL2正则化后取平均值得出一个具有说话者声纹特征的向量

声音过滤模块:输入d-vector向量和混合有噪声的音频，由八个卷积神经网络和一层递归神经网络组成，这其中除了最后的输出层的激活函数使用sigmoid函数之外，其余的隐藏层全部使用ReLu函数。将d-vector和时间轴重复之后与卷积层的结果合并输入到递归神经网络中最后得出单纯的参考人的说话音频

2、语音转写

**语音识别技术采用开源ASRT**

**仓库地址**：https://github.com/nl8590687/ASRT\_SpeechRecognition

ASRT 是一套基于深度学习实现的[语音识别](https://www.jiqizhixin.com/articles/021102)系统，全称为 Auto Speech Recognition Tool，由 AI 柠檬博主开发并在 GitHub 上开源（GPL 3.0 协议）。本项目声学模型通过采用[卷积神经网络](https://www.jiqizhixin.com/articles/021102)（CNN）和连接性时序分类（CTC）方法，使用大量中文语音数据集进行训练，将声音转录为中文拼音，并通过[语言模型](https://www.jiqizhixin.com/articles/021102)，将拼音序列转换为中文文本。基于该模型，作者在 Windows 平台上实现了一个基于 ASRT 的语音识别应用软件它同样也在 GitHub 上开源了。

**系统流程：**

特征提取：将普通的 wav 语音信号通过分帧加窗等操作转换为神经网络需要的二维频谱图像信号，即语谱图。

声学模型：基于 Keras 和 TensorFlow 框架，使用这种参考了 VGG 的深层的卷积神经网络作为网络模型，并训练。

CTC 解码：在语音识别系统的声学模型输出中，往往包含了大量连续重复的符号，因此，我们需要将连续相同的符号合并为同一个符号，然后再去除静音分隔标记符，得到最终实际的语音拼音符号序列。

语言模型：使用统计语言模型，将拼音转换为最终的识别文本并输出。拼音转文本本质被建模为一条隐含马尔可夫链，这种模型有着很高的准确率。

**模型使用：**

test\_mspeech.py

3、关键词检索系统的研究方法

* 关键词检索根据应用场景不同分为两种
* 模式固定，文本不定

场景：事先定义一些模式(比如说黄色关键词)。对每个给定的文本，确定是否含有

这些关键词

* 文本固定，模式不定

场景：已知网络上的网页库，给定一个字符串，寻找含有这些字符串的网页。

* 我们的需求是在客服与用户的对话中检索一些业务用语，所以我们的检索是“模式固定，文本不定”类型。
* “模式固定，文本不定”检索的算法研究：
* 单模式匹配(仅仅有单一模式文本)

KMP算法

* 多模式匹配

AC(Aho-Corasick, KMP)的升级版本