1. 特征提取

不同声纹模型往往需要不同的特征作为输入。Davis与Mermlestein提出的梅尔倒频谱系数代表了一般人耳对频率的感受度，因而被广泛运用于相对早期的说话人识别模型中。研究者在此前提出的线性预测编码、线谱对、感知线性预测系数也较为泛用。N. Dehak提出的i-vector提高了计算效率，常用于文本无关的说话人模型中，谷歌研究者提出的d-vector取自深度神经网络，研究者在此基础上进一步提出了x-vector，在大规模数据上正确率高于i-vector。

1. 声纹模型

文本相关的说话人识别主要运用动态时间规整（DTW）算法或隐马尔可夫模型（HMM），DTW算法解决了孤立词语音长度不同的问题，但时间复杂度为二次方级，Salvador, Stan, Chan与Philip提出了 FastDTW算法，将时间复杂度降至线性。HMM为经典的机器学习模型，效果相对更优秀，且有多种优化与改进方式。文本无关的说话人识别主要运用混合高斯模型(GMM)，GMM对数据的需求量较大，D. A. Reynolds等研究者建立了通用背景模型以对非目标用户的声音加以利用。后续的研究者通过运用因子分析算法框架实现了高斯分量均值集合的降维。Patric Kenny在此基础上进一步提出了联合因子分析，N. Dehak又对该框架进行了改进，对噪声部分实现了相似的降维。后续的研究者多在此基础上提出改进。

进度安排

【 . . - . . 】项目需求确认

【 . . - . . 】windows平台下文本相关的说话人识别模型与唇形识别程序测试

【 . . - . . 】树莓派平台下文本相关的说话人识别模型与唇形识别程序测试

【 . . - . . 】声纹锁系统框架构建

【 . . - . . 】接口完善

【 . . - . . 】完整声纹锁系统测试