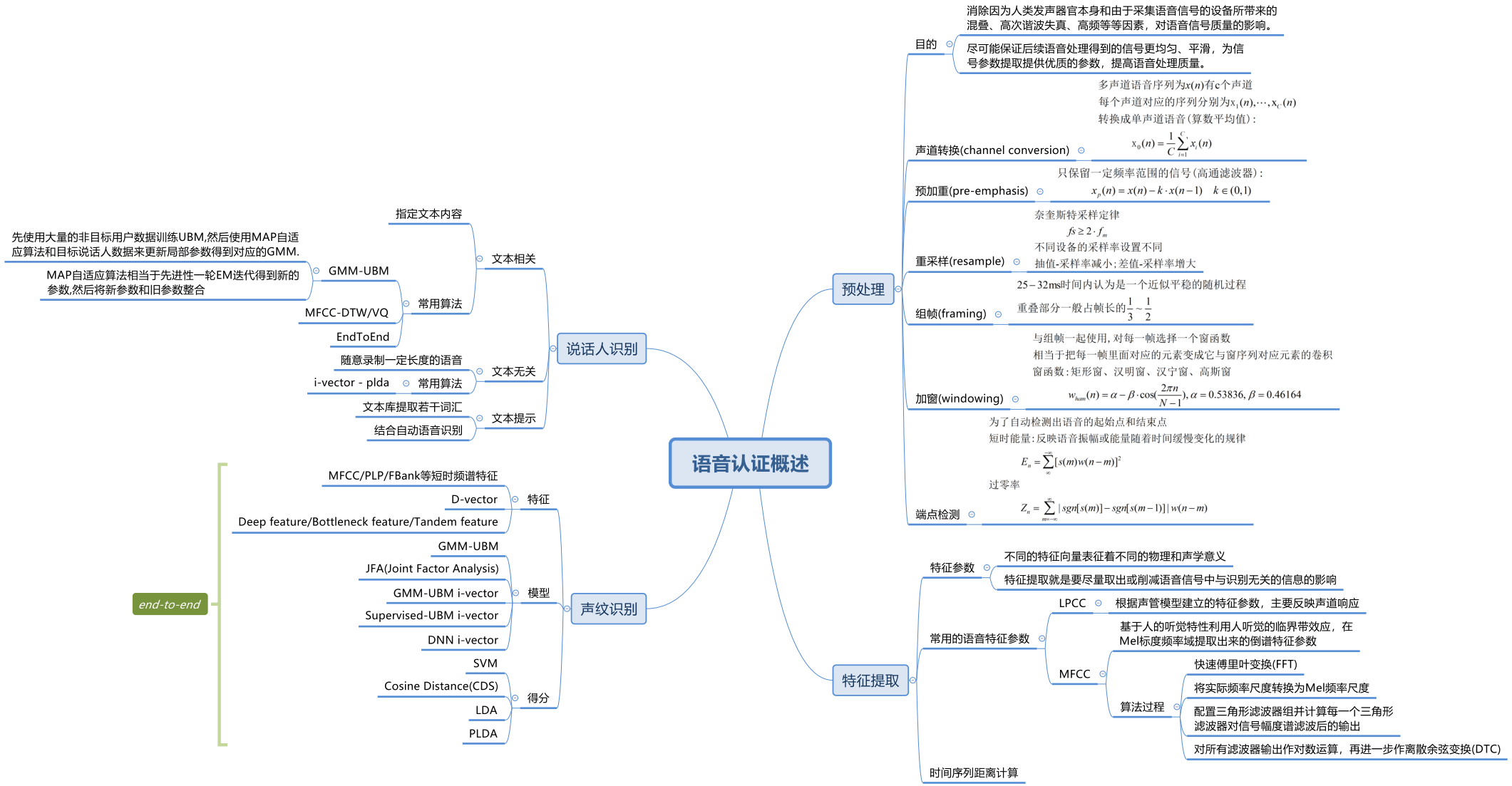
语音认证概述
预处理
目的
消除因为人类发声器官本身和由于采集语音信号的设备所带来的混叠、高次谐波失真、高频等等因素，对语音信号质量的影响。
尽可能保证后续语音处理得到的信号更均匀、平滑，为信号参数提取提供优质的参数，提高语音处理质量。
声道转换(channel conversion)
多声道语音序列为$x(n)$有c个声道
每个声道对应的序列分别为$x_1(n),\cdots,x_c(n)$
转换成单声道语音(算数平均值):
$x_0(n)=\frac{1}{C}\sum_{i=1}^{C}x_i(n)$
预加重(pre-emphasis)
只保留一定频率范围的信号(高通滤波器):
$x_p(n)=x(n)-k\cdot x(n-1)\quad k\in(0,1)$
重采样(resample)
奈奎斯特采样定律
$fs\geq 2\cdot f_m$
不同设备的采样率设置不同
抽值-采样率减小;差值-采样率增大
组帧(framing)
25－32ms时间内认为是一个近似平稳的随机过程
重叠部分一般占帧长的$\frac{1}{3}\sim\frac{1}{2}$
加窗(windowing)
与组帧一起使用,对每一帧选择一个窗函数
相当于把每一帧里面对应的元素变成它与窗序列对应元素的卷积
窗函数:矩形窗、汉明窗、汉宁窗、高斯窗
$w_{ham}(n)=\alpha-\beta\cdot\cos(\frac{2\pi n}{N-1}),\alpha=0.53836,\beta=0.46164$
端点检测
为了自动检测出语音的起始点和结束点
短时能量:反映语音振幅或能量随着时间缓慢变化的规律
$E_n=\sum_{m=-\infty}^{\infty}[s(m)w(n-m)]^2$
过零率
$Z_n=\sum_{m=-\infty}^{\infty}|sgn[s(m)]-sgn[s(m-1)]|w(n-m)$
特征提取
特征参数
不同的特征向量表征着不同的物理和声学意义
特征提取就是要尽量取出或削减语音信号中与识别无关的信息的影响
常用的语音特征参数
LPCC
根据声管模型建立的特征参数，主要反映声道响应
MFCC
基于人的听觉特性利用人听觉的临界带效应，在Mel标度频率域提取出来的倒谱特征参数
算法过程
快速傅里叶变换(FFT)
将实际频率尺度转换为Mel频率尺度
配置三角形滤波器组并计算每一个三角形滤波器对信号幅度谱滤波后的输出
对所有滤波器输出作对数运算，再进一步作离散余弦变换(DTC)
时间序列距离计算
声纹识别
特征
MFCC/PLP/FBank等短时频谱特征
D-vector
Deep feature/Bottleneck feature/Tandem feature
模型
GMM-UBM
JFA(Joint Factor Analysis)
GMM-UBM i-vector
Supervised-UBM i-vector
DNN i-vector
得分
SVM
Cosine Distance(CDS)
LDA
PLDA
end-to-end
说话人识别
文本相关
指定文本内容
常用算法
GMM-UBM
先使用大量的非目标用户数据训练UBM,然后使用MAP自适应算法和目标说话人数据来更新局部参数得到对应的GMM.
MAP自适应算法相当于先进性一轮EM迭代得到新的参数,然后将新参数和旧参数整合
MFCC-DTW/VQ
EndToEnd
文本无关
随意录制一定长度的语音
常用算法
i-vector - plda
文本提示
文本库提取若干词汇
结合自动语音识别