基于端点检测和高斯滤波器组的MFCC说话人识别①

王　萌, 王福龙

(广东工业大学 应用数学学院, 广州 510520)

摘　要: 在上下文无关的说话人识别应用中, 针对传统 MFCC 特征参数在语音预处理方面不足以及三角滤波器组的缺陷, 提出一种改进的MFCC特征参数提取方法. 一方面在传统算法上加入端点检测, 去除与说话人语音特征无关的静音段; 另一方面用高斯滤波器组(Gaussian shaped filters GF)代替三角滤波器组进行频率到 Mel 频率的转换, 提高识别准确率. 说话人识别模型使用流行的高斯混合模型(GMM). 实验结果显示, 高斯滤波器组的引入相比于传统三角滤波器组识别率有 4.45%的提升, 本文改进后的 MFCC 特征参数相比于传统方法识别率也提升了 6.43%, 能更好的代表说话人的语音特征.

关键词: MFCC 特征参数; 端点检测; 高斯滤波器组(GF); 高斯混合模型(GMM); 说话人识别

Speaker Identification with Improved MFCC Based on Endpoint Detection and Gaussian Shaped Filters

WANG Meng, WANG Fu-Long

(Applied Mathematics, Guangdong University of Technology, Guangzhou 510520, China)

Abstract: In the application of text-independent speaker recognition, this paper puts forward an improved feature extraction of MFCC parameters to supply the inefficient traditional MFCC. Endpoint detection is added in traditional algorithm to remove silence parts. Gaussian shaped filters are used to replace triangular filter banks to improve the accuracy of speaker identification. Gauss mixed model is for speaker recognition. Experiments show that Gaussian shaped filters gain 9.63% performance improvement while proposed MFCC can significantly improve recognition rate by 11.07%. The result indicates that the new method is an effective feature extraction algorithm.

Key words: MFCC parameters; endpoint detection; Gaussian shaped filters(GF); Gauss mixed model(GMM); speaker identification

1　引言

说话人识别系统是自动识别说话人的过程, 它包括说话人识别和说话人确认. 说话人识别可以分为上下文相关的说话人识别和上下文无关的说话人识别[1,2]; 前一种要求说话人在训练和测试阶段说相同的内容, 后一种对训练和测试阶段的说话内容没有要求, 或者说识别是与内容无关的. 本文主要针对上下文无关的说话人进行识别, 对说话人的说话内容没有特别要求, 推广应用价值比较大.

说话人识别系统的主要组成部分有特征提取[3]、模型的训练和识别过程. 目前, 在说话人识别领域中使用较多的特征提取方法有 Mel 频率倒谱系数(MFCC)[4-6]、线性倒谱系数(LPCC)、线性倒谱对(LSP)等. 这些方法中 MFCC 特征参数能够有效描述语音信号在频率域上的能量分布, 较好的模拟人耳听觉系统的感知能力, 自提出以来一直被用在说话人的语音特征提取中. 本文同样以 MFCC 作为特征提取方法, 并在传统 MFCC 的基础上进行改进, 一方面是在语音特征提取之前加入端点检测算法[7], 将语音中的静音段除去, 通过缩短语音时长来降低 MFCC 的计算量; 另

① 基金项目:广东省自然科学基金(S2011040004273)

收稿时间:2016-02-17;收到修改稿时间:2016-04-05　[doi:10.15888/j.cnki.csa.005425]

一方面是将频率到 Mel 频率转化过程中的三角滤波器组改为高斯滤波器组(GF)[8], 与三角滤波器组比较起来, GF 的滤波参数能够更有效地提供相邻子带之间的平滑过渡, 提高识别的准确率[9]. 在说话人识别模型的建立方面使用最流行的高斯混合模型(GMM)[10-12]对提取的特征进行分析, 检验改进方法的优劣.

本文总共分为六个部分, 其余部分的内容概括如下, 第二部分介绍传统 MFCC 方法, 第三部分重点讲解本文在语音端点检测处理和针对传统 MFCC 中三角滤波器组的改进算法, 说话人识别系统在第四部分进行介绍, 第五部分是实验结果和分析, 第六部分对全文进行总结.

2 传统MFCC特征参数

MFCC 特征参数基于人耳听觉感知系统来对语音信号进行处理, 在语音特征提取方面有独特的优势. 传统 MFCC 特征提取[13]过程是: 首先将语音信号进行预处理; 接下来对处理后的信号进行 FFT 变换; 然后通过 Mel 尺度滤波器组, 将频率信号转化到 Mel 频率; 对 Mel 频率上的信号取对数计算能量谱; 最后对输出信号作离散余弦变换(DCT), 得到 MFCC 特征参数.

传统 MFCC 方法的流程图如图 1 所示. 具体操作如下:

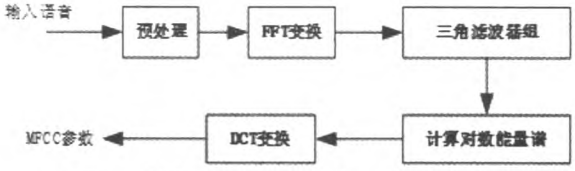


图 1 传统 MFCC 特征提取的流程图

1) 预加重、分帧、加窗

语音信号首先经过预加重滤波器进行信号的平坦化处理, 在时间域上, 预加重的输入输出关系如下:

$$S_n = S_n - \alpha S_{n-1} \tag{1}$$

其中 α 通常取为 0.97. 之后对语音信号进行分帧, 帧长一般设定在 20-40ms 来保证帧内的稳定性, 同时相邻帧之间有 10-30ms 的重叠来保证帧间的稳定性. 为了减少语音帧的边缘效应, 需要对预加重之后的语音帧进行加窗处理. 加窗处理通常是将语音帧乘以一个窗函数. 常用韩明窗的表达式如下:

$$\omega(n) = 0.54 - 0.46\cos\left(2\pi\frac{n-1}{N-1}\right) \tag{2}$$

$$n = 1,2...,N$$

经过处理之后得到的加窗语音信号表达式为:

$$S_\omega(n) = S_n \times \omega(n) \tag{3}$$

以上公式中, N 是当前信号帧中采样点的数量, $\omega(n)$ 是窗函数, $S_\omega(n)$ 是加窗后的语音信号.

2) FFT

为了更方便的观察到语音信号内包含的信息, 需要对加窗后的语音信号进行 FFT,将时域上的信号变化到频率域, 从而得到语音信号的线性频谱, 其计算公式如下:

$$X(k) = \sum_{n=0}^{N-1} S_\omega(n) e^{-\frac{j2\pi nk}{N}} \quad (0 \le n,k \le N-1) \tag{4}$$

3) 三角滤波器组

经过 FFT 之后, 为了使信号频率更接近人耳的听觉感知系统, 需要将声音信号从线性频率转换到 Mel 频率[14,15], 公式如下:

$$Mel(f) = 2595\log_{10}\left(1+\frac{f}{700}\right) \tag{5}$$

上述变化过程的实现是将每帧语音信号通过一系列滤波器组进行滤波处理, 通常选取三角滤波器组作为处理工具. 三角滤波器组如图 2 所示.

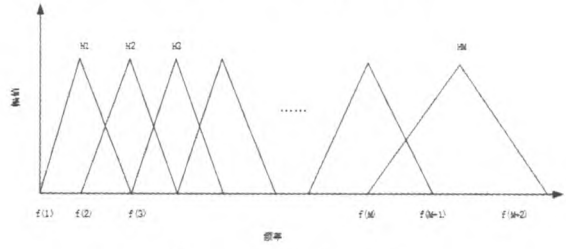


图 2 三角滤波器组

在这个过程中, Mel 滤波器组在语音信号的频谱范围内设置若干带通滤波器 $H_m(k), 0 \le m \le M$, M 为滤波器组的个数. 三角滤波器组的中心频率为 $f(m)$, 当 m 值较小时, 相邻 $f(m)$ 之间的间隔小, 随着 m 的增加, 相邻 $f(m)$ 的间隔逐渐变大. 三角滤波器组的传递函数为:

4) 计算对数能量谱

对滤波器组处理得到的 Mel 频率谱进行取对数的操作, 得到语音信号的能量谱. 由线性频谱 $X(k)$ 到

对数频谱 $S(m)$ 的传递函数为:

$$H_m(k)=\begin{cases}0,(k<f(m-1))\\ \dfrac{k-f(m-1)}{f(m)-f(m-1)},(f(m-1)\le k\le f(m))\\ \dfrac{f(m+1)-k}{f(m+1)-f(m)},(f(m)<k\le f(m+1))\\ 0,(k>f(m+1))\end{cases}\quad(0\le m\le M)\tag{6}$$

$$S(m)=\ln\left(\sum_{k=0}^{N-1}|X(k)|^2H_m(k)\right),(0\le m<M)\tag{7}$$

5) DCT, 得到 MFCC

将步骤 4)中的对数能量谱 $S(m)$ 经过 DCT(离散余弦变换)转换到时域得到倒频谱, 即可得到 Mel 频率倒谱系数 $c(n)$:

$$c(n)=\sum_{m=1}^{M-1}S(m)\cos\left(\frac{\pi n(m+1/2)}{M}\right),\quad(0\le m<M)\tag{8}$$

3 改进MFCC算法

传统 MFCC 在预处理阶段只对声音信号进行分帧、加窗操作, 但是说话人语音信号中会含有录音过程中的停顿静音部分, 这部分会对语音特征提取产生干扰, 也会降低识别准确率; 而且静音部分会造成语音时长变长从而增加特征提取时的计算量. 本文的改进算法, 一方面是对原始语音信号在提取 MFCC 特征之前加入端点检测, 进行静音部分的检索和删除; 另一方面是将传统MFCC中的三角滤波器组改为高斯滤波器组(GF). 将这两部分增加到传统 MFCC 特征提取算法中, 得到新的 MFCC 特征参数. 虽然端点检测在语音识别算法中已经有过应用, 但是将端点检测和高斯滤波器组两方面同时应用在语音信号的特征提取中, 之前的文章都没有这样使用过, 经本文实验验证, 发现这样得到特征参数更具有代表性, 识别效果更好. 下面针对这两点改进的具体操作进行详细介绍.

3.1 端点检测算法

语音录制过程中, 开头、结尾以及中间部分词与词之间很可能有静音段出现, 这些段没有包含说话人的声学特征, 但是在 MFCC 特征参数提取中还会进行计算和分析, 因此如果能够在提取 MFCC 特征之前先将这些静音部分除去, 不仅可以降低静音部分对整体语音特征的影响, 还能够在不破坏语音关键特征部分的同时减少数据的储存量、处理时间和计算量, 提高系统的识别效率. 本文使用的端点检测算法计算量小、操作简单、易于实现. 算法基于语音信号的频域和时域两方面进行处理, 分别采用短时能量和短时过零率来进行语音端点检测[7].

1) 短时能量(Short Time Energy-STE)

语音信号的能量与每帧信号 $x(m)$ 的平方和有关, 从而语音信号的短时能量定义为:

$$E=\sum_{m=-\infty}^{\infty}x^2(m)\tag{9}$$

2) 短时过零率(Zero Crossing Rates-ZCR)

短时过零率是指每帧语音信号通过零幅度线的次数, 其计算定义如下:

$$Z_n=\sum_{m=-\infty}^{\infty}|\mathrm{sgn}[x(m)]-\mathrm{sgn}[x(m-1)]|\omega(n-m)\tag{10}$$

其中 $\mathrm{sgn}(x)$ 是符号函数, 即:

$$\mathrm{sgn}(x)=\begin{cases}1, & (x\ge 0)\\ -1, & (x<0)\end{cases}\tag{11}$$

在语音信号处理过程中, 静音段语音的短时能量为零, 理想情况下, 静音段的过零率也为零. 在实际中, 假设有一段语音, 其中某一部分的短时能量和短时过零率都为零或者是很小的值, 就可以认为这部分为静音段; 如果某部分短时能量很小, 过零率很大, 则认为此段语音为清音段; 或者某部分短时能量很大, 过零率很小, 则认为此段语音为浊音段.

在语音端点检测中, 需要对短时能量和短时过零率两方面进行分析, 通常采用双门限判定法[14]来检测语音端点, 即利用过零率检测清音段, 短时能量检测浊音段, 两者进行配合. 具体操作如下: 为短时能量和短时过零率分别确定两个门限值, 一个较低的门限数值设定较小, 这样容易很敏感的检测到语音信号的变化; 另一个较高的门限数值设定较大. 如果低门限值被超过, 可能是语音的开始, 或者也可能是短时噪

声引起的; 如果高门限值被超过且在随后短时间段内的语音信号值超过低门限, 这就意味着信号的开始, 从而判断出静音段部分语音.

通过上述端点检测算法找出语音中的静音段, 将这部分语音段删除来缩短语音时长, 进而缩短语音信号处理的时间.

3.2 GF

传统 MFCC 特征参数提取中, 为了得到 Mel 频率范围的语音谱, 计算时选取了三角滤波器组对频率上的语音信号进行处理. 在能量谱中, 三角滤波器组能提供语音帧之间的明确分区, 对分帧后语音的重叠部分和非重叠部分阈值分别设定; 但是三角滤波器组的阈值设定会造成相邻子带输出之间联系的丢失, 从而对识别率产生影响.

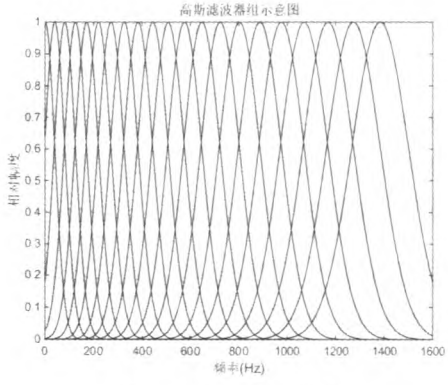


图 3 高斯滤波器组

本文改进算法使用高斯类型的滤波器组[8,15](如图 3 所示)代替三角滤波器组, 这样改进产生的效果有: GF 相比于三角滤波器组能够更好地提供一个子带到另一个子带的平滑过渡, 保留相邻子带间的大部分联系; 而且 GF 的控制方差可以根据实验独立选择, 从而控制相邻子带间重叠部分的程度. 下面具体介绍 GF:

GF 的滤波响应为:

$$GF_{MFCC} = e^{-\frac{(k-k_{bi})^2}{2\sigma_i^2}} \tag{12}$$

其中 k_{bi} 是第 i 个滤波器的边界点, σ_i 是标准偏差, 且定义如下:

$$\sigma_i = \frac{k_{b_{i+1}} - k_{b_i}}{\alpha} \tag{13}$$

其中, α 控制方差(根据参考文献[8]中对控制方差 α 的分析和实验测试, 本文在实验中同样选取 $\alpha = 2$). 通过横向乘以 FFT 幅度值(包含在滤波器组中), 就可以获得 Mel 频率倒谱系数:

$$c^{g_{MFCC}}{}_m = \sqrt{\frac{2}{Q}} \times \sum_{i=0}^{Q-1} \log\left[e^{g_{MFCC}}(i+1)\right] \cdot \cos\left[m \cdot \left(\frac{2l-1}{2}\right) \cdot \frac{\pi}{Q}\right] \tag{14}$$

其中, Q 为高斯滤波器组的个数(通常来说 $Q = 20$ [8], 本文为了与传统 MFCC 方法中的三角滤波器组个数设置保持一致, 在实验过程中选择取 $Q = 23$ 作为高斯滤波器组的个数值); $e^{g_{MFCC}}$ 为高斯滤波器组计算得到的倒谱, 表达式如下:

$$e^{g_{MFCC}}(i) = \sum_{k=1}^{\frac{Ms}{2}} |Y(k)|^2 \cdot \Psi^{g_{MFCC}}{}_i(k) \tag{15}$$

将以上两点改进加入到传统 MFCC 特征参数提取中, 就得到本文提出的改进方法. 改进算法不仅能够将不含有语音特征的静音部分除去, GF 的引入又提升了不同语音特征之间识别的准确率, 在提升识别率的同时又保留了传统方法在说话人识别中的优势. 改进特征参数提取流程图如 4 所示(其中边框加粗为改进部分).

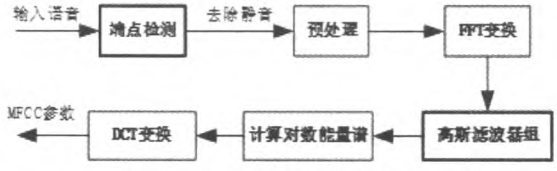


图 4 改进特征参数提取流程图

4 基于GMM 的说话人识别系统

GMM 是概率分布模型, 其分布函数的维度与声学特征维度一致. GMM 由多个高斯概率分布函数经过加权和构成, 在声学模型建立中, GMM 的分布函数能最大程度接近实际声学特征. 当下, GMM 是说话人识别系统中最常用的模型之一[16-18].

4.1 基于 GMM 的识别系统

说话人识别系统中, 通过 MFCC 得到不同人的语音特征, 对特征进行训练确定模型参数, 之后利用得到的模型进行说话人识别. 对某个说话人 s 来说, GMM 由 M 个单高斯分布模型 $b_i^s(x)$ 通过线性加权构成, p_i^s 是混合权重, 对均值 μ_i^s 和方差 $\sum_i^s$ 进行参数化处理, 将提取出的 D 维的特征矢量用 x 表示, 则说话人 s 的混合概率密度定义为:

$$p(x|\lambda_s) = \sum_{i=1}^{M} p_i^s b_i^s(x) \tag{16}$$

其中,

$$\sum_{i=1}^{M} p_i^s = 1 \tag{17}$$

$$b_i^s(x) = \frac{1}{(2\pi)^{D/2}\left|\sum_i^s\right|^{1/2}} \times \exp\left\{-\frac{1}{2}(x-\mu_i^s)'\left(\sum_i^s\right)^{-1}(x-\mu_i^s)\right\} \tag{18}$$

由此可以得到说话人 s 的参数模型定义为:

$$\lambda_s = \left\{p_i^s, \mu_i^s, \sum_i^s\right\}, i = 1,...,M \tag{19}$$

GMM 的结构可以用图 5 来表示.

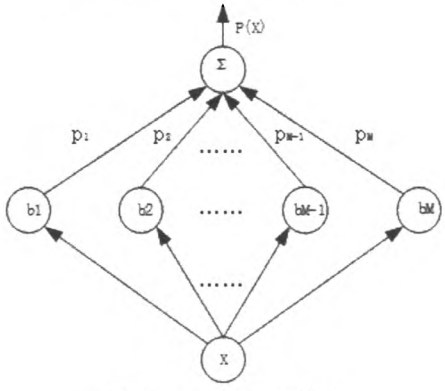


图 5　GMM 结构图

4.2 模型参数的训练过程和结果输出

利用 GMM 模型中的参数 $\lambda_1, \lambda_2, ..., \lambda_S$ 来代表 S 个说话人, 其中 $S = \{1, 2, ..., s, ..., S\}$. 说话人识别系统的目标是找到最能代表某个说话人的模型参数 λ, 令 $p_c, \mu_c, \sum_c$ 分别代表第 c 个高斯分量的权重、均值和方差, 经过计算得到 MFCC 观察特征矢量序列假设为 $X(x_1, x_2, ..., x_d, .., x_D)$, 那么 GMM 的似然函数如下:

$$p(X|\lambda) = \prod_{d=1}^{D} p(x_d|\lambda) \tag{20}$$

对于给定的观察特征矢量序列, 需要选用优化准则来确定模型参数 λ 的取值, 使其更好地符合给定矢量序列的真实分布. 优化准则一般选用最大似然估计准则, 它能在给定训练特征矢量的情况下, 寻找合适的模型参数 λ, 使得 GMM 的似然函数最大, 即:

$$\hat{\lambda} = \arg\max_{\lambda}\{p(x|\lambda)\} \tag{21}$$

由于上述公式在求解参数 λ 时非常困难, 因此, 通常利用期望最大(EM)迭代算法来对参数进行估计, EM 算法的基本思想是, 给定 GMM 模型参数一个初始值 λ_0, 从 λ_0 开始进行循环迭代运算, 在迭代中不断修改模型参数直到模型参数 λ 达到指定意义上的收敛, 即每次迭代都有 $p(x|\lambda_i) \geq p(x|\lambda_{i+1})$, 其中, λ_{i+1} 是在前一次模型参数 λ_i 的基础上计算出的优化模型参数; 到此训练完成, 即可得到 GMM 参数. 其中 GMM 各参数的更新过程满足如下公式:

$$p_c = \frac{1}{D}\sum_{d=1}^{D} P_c(x_d) \tag{22}$$

$$\mu_c = \frac{1}{\sum_{d=1}^{D}\lambda_j^d}\sum_{d=1}^{D} P_c(x_d)x_d \tag{23}$$

$$\sum_c = \frac{1}{\sum_{d=1}^{D}\lambda_c^d} \times \sum_{d=1}^{D} P_c(x_d)(x_d-\mu_c)^D(x_d-\mu_c) \tag{24}$$

其中, $P_c(x_d)$ 为每个 x_d 在高斯分量 c 上的隐含类别概率. 第 c 个混合分量的后验概率为:

$$p(c|x_d, \mu_c, \sum_c) = \frac{\omega_c p[x_d|\mu_c, \sum_c]}{\sum_{k=1}^{M}\omega_k p[x_d|\mu_k, \sum_k]} \tag{25}$$

则 GMM 识别的目标是找到有最大后验概率的模型所对应的说话人, 即

$$S = \arg\max_{1\leq s\leq S} p(\lambda_s|x) \tag{26}$$

最终, 根据贝叶斯准则, 测试语音数据中说话人的统计选择根据以下公式进行:

$$\begin{aligned} &S = \arg\max_{1<s<S} P(x|\lambda_s) \xrightarrow{take\log} \\ &S = \arg\max_{1<s<S}\sum_{d=1}^{D}\log P(\overline{x}_d|\lambda_s) \end{aligned} \tag{27}$$

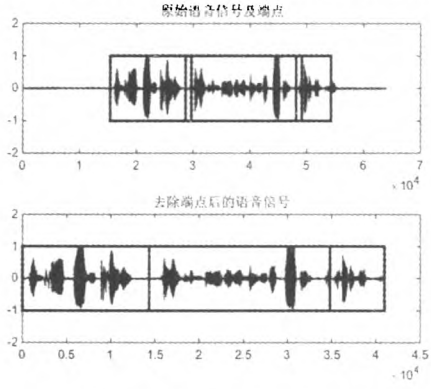
到此构成整个说话人的语音识别系统.

5 实验结果和分析

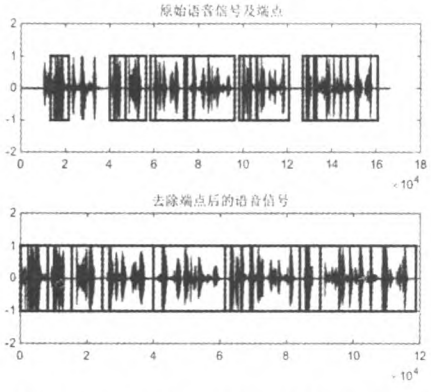
实验在 MatlabR2014a 环境下操作, 采用 8KHz 采样率、8bits 量化、单声道录音. 整个数据库是在普通实验室环境下录制 18 个说话人, 每人录制 150 段语音, 录制过程中每个人的说话内容各不相同, 从中抽取 80 段语音作为训练样本, 余下 70 段用于测试. 特征提取阶段每帧语音长度设定为 25ms, 帧移量为 10ms, MFCC 阶数设定为 13, 窗函数选用韩明窗, 说话人识别模型选用高斯混合模型(GMM)来完成整个说话人识别系统.

利用不同人的 MFCC 特征参数在 GMM 中分别进行训练(其中高斯混合度选为 8), 最终确定 GMM 中的

三组参数. 测试阶段, 分为四组对比实验对需要测试的1260段语音进行识别. 第一组使用传统MFCC进行语音的特征提取; 第二组使用加入端点检测算法的MFCC, 端点检测的实现效果如图6所示; 第三组将传统MFCC中的三角滤波器组改为高斯滤波器组进行语音特征提取; 第四组是本文提出的改进 MFCC 特征参数提取方法. 识别阶段统一采用GMM, 以识别正确来计算识别率, 四组实验的识别率统计结果见表 1.



(a) 保留语音特征完好的端点检测



(b) 部分语音特征被删除的端点检测

图 6 端点检测

表 1 不同特征提取方法识别率的比较

说话人序号	识别率(%) 传统MFCC	端点检测+MFCC	高斯滤波器组+MFCC	本文提出的MFCC
1	94.29	97.14	97.14	95.71
2	95.71	94.29	95.71	97.14
3	98.57	98.57	98.57	98.57
4	80.00	80.00	84.29	85.71
5	87.14	91.43	91.43	94.29
6	77.14	77.14	84.29	84.29
7	94.29	94.29	95.71	97.14
8	78.57	82.86	84.29	92.86
9	35.71	37.14	41.43	44.29
10	80.00	71.43	85.71	88.57
11	87.14	88.57	90.00	90.00
12	78.57	71.43	85.71	88.57
13	84.29	87.14	87.14	85.71
14	92.86	92.86	97.14	98.57
15	68.57	71.43	85.71	97.14
16	45.71	45.71	54.29	54.29
17	95.71	95.71	97.14	97.14
18	98.57	98.57	97.14	98.57
平均	81.82	81.98	86.27	88.25

由实验结果(见表 1)的对比分析可知, 本文提出的改进算法与传统 MFCC 比较起来识别率有 6.43%的提升. 只使用端点检测时某些说话人语音识别率相比于传统方法有三个降低和六个持平的情况出现(如表 1 中绿色字体为与传统方法比较识别率降低的情况, 蓝色字体为持平情况). 本文采用的端点检测算法为最基础算法, 具有计算简单, 寻找端点速度快的特点; 但是对部分语音帧的处理中, 会去除包含有说话人特征的语音帧(如图 6(b)中所示), 从而造成少量特征丢失的情况, 这一缺陷在后期的研究中会着重考虑和进一步改进、优化基础算法性能的不足. 只考虑 GF 时, 整体识别率相比于传统 MFCC 有 4.45%的提升, 说明高斯滤波器组在说话人识别中具有独特的优势. 本文将端点检测和 GF 结合提出一种改进的 MFCC 特征参数, 这样的改进在之前的文章中都没有使用过, 且与传统方法比较起来, 改进方法的整体识别率有较大提高, 因而具有很大的实用价值和继续深入研究意义.

6 结论

说话人识别领域中, 传统 MFCC 特征参数是最常使用的说话人特征提取算法. 本文针对传统算法不能处理语音中的静音段, 以及三角滤波器组会破坏相邻语音帧之间联系的问题, 提出在特征提取之前加入端点检测算法, 并使用 GF 代替三角滤波器组进行频率到 Mel 频率的转化. 这两方面结合进行说话人识别的方法, 之前文章中都没有使用过. 本文采用这种改进的 MFCC 特征参数对说话人语音进行特征提取. 实验结果表明, 改进后的MFCC特征参数与传统方法相比, 在相同实验环境下, 识别率有很大程度提高, 能够在说话人识别系统中得到较准确的识别结果. 下一步的工作, 将继续对改进方法进行优化, 特别是对 GF 中控

制方差的自适应选取进行优化, 在此基础上提升识别精度和速度.

参考文献

1 Furui S. 40 years of progress in automatic speaker recognition. Advances in Biometrics. Springer Berlin Heidelberg, 2009: 1050–1059.

2 Nidhyananthan SS, Shantha SKR. Language and text-independent speaker identification system using GMM. Wseas Trans. on Signal Processing, 2013: 185–194.

3 Kinnunen T, Li H. An overview of text-independent speaker recognition: From features to supervectors. Speech Communication,2010, 52(1): 12–40.

4 Vergin R, O'Shaughnessy D, Farhat A. Generalized mel frequency cepstral coefficients for large-vocabulary speaker-independent continuous-speech recognition. IEEE Trans. on Speech & Audio Processing, 1999, 7(5): 525–532.

5 Ahmad KS, Thosar AS, Nirmal JH, et al. A unique approach in text independent speaker recognition using MFCC feature sets and probabilistic neural network. 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR). IEEE. 2015. 1–6.

6 Metzger RA, Doherty JF, Jenkins DM. Analysis of compressed speech signals in an automatic speaker recognition system. 2015 49th Annual Conference on Information Sciences and Systems (CISS). IEEE. 2015. 1–5.

7 Zhu JW, Sun SF, Liu XL, et al. Pitch in speaker recognition. Ninth International Conference on Hybrid Intelligent Systems, 2009, HIS'09. IEEE. 2009. 33–36.

8 Chakroborty S, Saha G. Improved text-independent speaker identification using fused MFCC & IMFCC feature sets based on Gaussian filter. International Journal of Signal Processing, 2009, (1): 11–19.

9 Reynolds DA. A Gaussian Mixture Modeling Approach to Text Independent Speaker Identification [Ph.D Thesis]. Atlanta, Ga, USA: Georgia Institute of Technology, September 1992.

10 Larson HJ, Shubert BO. Random Variables and Stochastic Processes. Wiley, 1991: 957–958.

11 Preti A, Bonastre JF, Matrouf D, et al. Confidence measure based unsupervised target model adaptation for speaker verification. Proc. Interspeech, 2007, (12): 754–757.

12 Nirmal J, Kachare P, Patnaik S, et al. Cepstrum liftering based voice conversion using RBF and GMM. 2013 International Conference on Communications and Signal Processing (ICCSP). IEEE. 2013. 570–575.

13 Sivan S, Gopakumar C. An MFCC based speaker recognition using ANN with improved recognition rate. International Journal of Emerging Technologies in Computational and Applied Sciences, 2014, 8(4): 365–369.

14 Kinnunen T, Rajan P. A practical, self-adaptive voice activity detector for speaker verification with noisy telephone and microphone data. ICASSP. 2013. 7229–7233.

15 Sahidullah M, Saha G. Design. Analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Speech Communication. 2012, 54(4): 543–565.

16 Reynolds DA, Rose RC. Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans. on Speech & Audio Processing, 1995, 3(1): 72–83.

17 Maesa A, Garzia F, Scarpiniti M, et al. Text independent automatic speaker recognition system using mel-frequency cepstrum coefficient and Gaussian mixture models. Journal of Information Security, 2012, 3(4): 335–340.

18 Dhameliya K, Bhatt N. Feature extraction and classification techniques for speaker recognition: A review. 2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO). IEEE. 2015. 76–79.