

差分和加权 Mel 倒谱混合参数应用于说话人识别

柯晶晶¹, 周 萍¹, 景新幸², 杨 青¹

(1 桂林电子科技大学 电子工程与自动化学院, 广西 桂林 541004;

2 桂林电子科技大学 信息与通信学院, 广西 桂林 541004)

摘 要: 说话人识别是信息技术和生物学的新一代身份验证方式, 在说话人识别的研究中, 特征参数的提取直接影响到识别系统最终的识别效率. 通过对 Mel 频率倒谱系数特征参数进行分析研究, 基于 Mel 频率倒谱系数改进加权函数, 将体现个人语音特性的加权特征参数与反映语音帧间变化的差分 Mel 频率倒谱系数进行维度筛选, 再进行参数混合. 实验结果表明, 通过改进加权函数提取得到的特征参数与差分 Mel 频率倒谱系数的混合参数在矢量量化的说话人识别系统中, 码本容量为 16 和 32 时可以达到 100% 的识别率.

关键词: 说话人识别; 加权 Mel 频率倒谱系数; 混合参数; 矢量量化

中图分类号: TN912.34

文献标识码: A

文章编号: 1000-7180(2014)09-0088-04

Mixed Parameters of Differential and Weighted Mel Cepstrum Used in Speaker Recognition

KE Jing-jing¹, ZHOU Ping¹, JING Xin-xing², YANG Qing¹

(1 School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China;

2 School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: In the speaker recognition system, the extraction of characteristic parameters directly affects the final recognition efficiency. The study analyzed Mel frequency cepstrum coefficient characteristic parameter, improved the function of weighted Mel frequency cepstrum coefficient, and mixed the improved weighted Mel frequency cepstrum coefficient reflecting individual voice characteristics and differential Mel frequency cepstrum coefficient reflecting the voice changes between frames. The experimental results show that the mixed parameters of weighted Mel frequency cepstrum coefficient from the improved function and differential Mel frequency cepstrum coefficient makes the recognition rate of 100% when codebook-size of vector quantization is 16 or 32.

Key words: speaker recognition; weighted Mel frequency cepstrum coefficient; mixed parameters; vector quantization

1 引言

说话人识别也被称为“声纹”识别. 说话人识别技术在互联网应用及通信领域、司法鉴定、军队安全系统、生物医学方面等有着广泛的应用前景, 是国内外的研究热点并取得了相当多的研究成果. 提高说话人识别效率可以从两个方面着手: (1) 改善特征参数, 即降低噪声对参数的影响, 放大更具个人特性的

语音特征; (2) 补偿模型, 即对模型进行修正, 补偿在训练与实际应用时存在的差异.

目前常用的说话人识别特征参数有基因频率、线性预测参数 LPC (Linear Prediction Coefficient)、Mel 频率倒谱系数 MFCC (Mel Frequency Cepstrum Coefficient) 等^[1]. 其中 MFCC 能够较充分模拟人耳的特殊感知特性, 得到大量研究者的青睐, 基于 MFCC 特征参数的改善与优化是说话人识别研

收稿日期: 2014-01-05; 修回日期: 2014-02-23

基金项目: 广西自然科学基金(2012GXNSFAA053221); 国家自然科学基金(61363005)

究的重点.大量研究表明,基于静态 MFCC 提取的动态系数差分 Mel 频率倒谱系数 Δ MFCC (Differential Mel Frequency Cepstrum Coefficient)、加权 Mel 频率倒谱系数 WMFCC (Weighted Mel Frequency Cepstrum Coefficient),相比 MFCC 在识别率上有所提高.为提取更具鲁棒稳健的语音特征参数,就必须大幅降低噪声的影响,挖掘语音信号中更能体现个人特性的语音参数^[2].由于每个说话人固有的说话习惯,由其特征参数分布可以得到表征说话人特性的信息.本文基于 MFCC 特征参数进行研究,对现有加权系数进行改进,获得改进的加权 Mel 频率倒谱系数 IWMFCC (Improved Weighted Mel Frequency Cepstrum Coefficient),通过特征参数的维度筛选,保留最能表征个人语音特征的部分,再将反映语音帧间变化的 Δ MFCC 和更具个性化的 WMFCC 或 IWMFCC 进行维度扩张,得到混合参数,基于矢量量化的说话人识别模型进行实验,结果表明,说话人识别的鲁棒性得到了提高.

2 特征参数的提取

文献[2]提出的 Mel 频率倒谱系数^[3],能很好地体现人耳的听觉特性,鲁棒性较好.为达到模拟人耳听觉感知特性的目的,在 MFCC 提取过程中,利用 Mel 频率尺度对实际频率进行弯折, Mel 频率 f_{mel} 与实际频率 f 的转换公式如下:

$$f_{\text{mel}} = 1127 \ln(1 + f/700) \quad (1)$$

参数的提取过程如图 1 所示

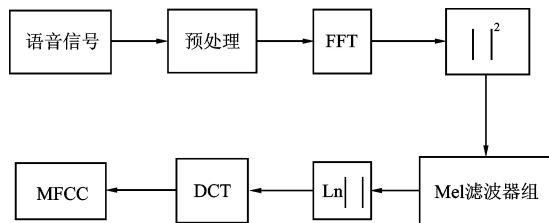


图 1 MFCC 特征参数的提取过程

在预处理过程中,为减少信号受外界影响而损失的高频频谱,对信号进行预加重滤波处理,函数如下:

$$H(z) = 1 - 0.95z^{-1} \quad (2)$$

经过预加重处理后,对信号进行加窗分帧处理,本文采用汉明窗,窗宽(帧长)512 个采样点,帧移 256 个采样点, N 为采样点数,汉明窗表达式如下:

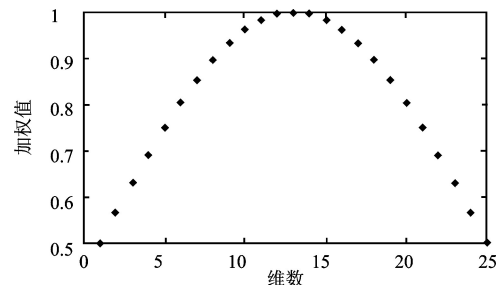
$$\omega(n) = \begin{cases} 0.54 - 0.46 * \cos[2\pi n / (N-1)], & 0 \leq n \leq N-1 \\ 0, & n = \text{else} \end{cases} \quad (3)$$

语音信号经过预处理得到短时数字信号,进行离散 FFT,并取其模的平方得到离散功率谱,对离散功率谱进行 24 维 Mel 滤波器组滤波,求取对数能量后再进行离散余弦变换(DCT),得到 24 维 MFCC 特征参数矩阵.图 3(a)为 MFCC 特征参数的条线示例图.

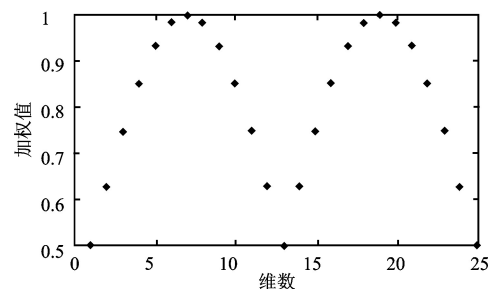
与传统静态 MFCC 相比,动态特征参数可以更加完整地表征说话者的语音特征^[4].差分和加权是两种获取动态特征参数的方法.将提取 MFCC 静态频谱换位动态频谱,并从动态频谱中提取 MFCC 的一阶差分倒谱系数 Δ MFCC, Δ MFCC 能够描述语音信号帧间瞬时变化的动态特征,图 3(b)为 24 维 Δ MFCC 的条形示例图;由于各维参数分量对说话人识别的表征能力不同,为使所得特征参数更具个性化信息,更具区分性,往往在特征向量内部采用加权,加权特征参数 WMFCC 能够突出说话人的个性特征.加权一般采用升半正弦函数^[4],如下式:

$$r = 0.5 + 0.5 * \sin(\pi * (i-1)/n) \quad (4)$$

式中, $i=1,2,\dots,n$, n 为特征参数的维数,本文 $n=24$,由升半正弦函数加权所得的加权特征参数 WMFCC 的条形图如图 3(c)所示.传统采用的升半正弦函数加权分布图 2(a)所示.



(a)升半正弦函数加权分布



(b)改进加权函数加权分布

图 2 $n=24$ 时加权函数的加权分布

由图 2(a)可知升半正弦加权函数值在 13 时最大,而通过图 3(a)可知, MFCC 特征参数在第 13 维左右的特征参量值不是最大的,为了更精确地反映不同说话人的区别,加权最大值应该与 MFCC 的参数最大值位置相符,本文提出另一种加权函数,如

下式:

$$r_1 = 0.5 + 0.5 * |\sin(2\pi * (i-1)/n)| \quad (5)$$

加权分布如图 2(b)所示,得到的加权特征参数 IWMFCC 条形图如图 3(d)所示.

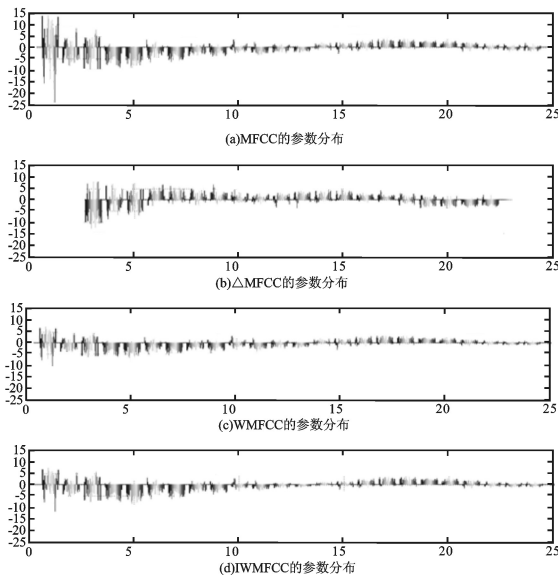


图 3 参数分布条形图

(注:示例为 24×92 的特征参数矩阵,横坐标为特征参数的维数,纵坐标为各维参数量值)

由于语音中不可避免的包含了噪声,大量实验表明,在特征参数的提取过程中,噪声影响主要表现在特征矩阵的低维部分.由图 3(a)可知,MFCC 特征参数的第 1、2、3 维参数值受噪声影响最大,这使得在说话人识别系统的欧式距离测度中,MFCC 特征参数矩阵高维参数的绝对距离容易被低维参数的绝对距离所掩盖.由 MFCC 得到 Δ MFCC、WMFCC,以及 IWMFCC 的低维特征参数也受到噪声的干扰.为了得到较好的识别效果,采用维度筛选,选取受噪声影响小的且能表征语音特征的部分.由于差分倒谱的低二维和高二维都为零,通过分析图 3(b)可知,差分倒谱的第 3、4、5 维受到的噪声影响,与 MFCC、WMFCC、IWMFCC 的第 1、2、3 维受到的影响相对应.在进行参数组合时,选取 6-22 维的 Δ MFCC,分别与 4-20 维的 WMFCC 和 4-20 维的 IWMFCC 进行同维度参数分量扩张,得到 17 维的混合特征参数^[5].

3 矢量量化模型

矢量量化(Vector Quantization, VQ)是一种高效的数据压缩和编码技术,使得提取特征参数的数据量小,存储空间小且计算效率高.在说话人识别

中,矢量量化起着双重作用^[6].在训练阶段,把每一个说话人的特征参数运用 LBG 算法进行聚类分析,从而训练生成一个码本.识别时,对于输入的说话人语音特征序列,计算该序列与每一个码本的平均失真测度,失真度最小的码本所对应的说话人即为识别结果.基于矢量量化技术的说话人识别系统原理框图如图 4 所示.

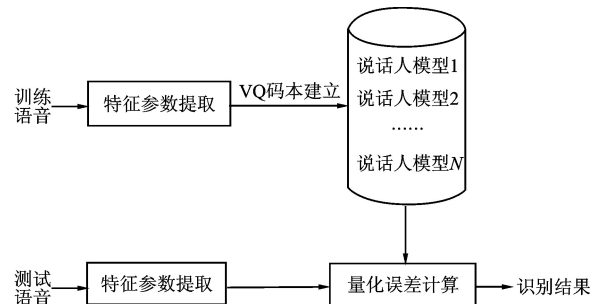


图 4 基于矢量量化的说话人识别系统原理框图

在计算平均失真测度时,需要用到一种常用的距离测度:欧式距离(Euclidean distance)也称欧几里德距离.设码本维数为 $M \times N$ 的矩阵 $\mathbf{X}$ 、 $\mathbf{Y}$,定义它们之间的欧氏距离为:

$$D = \sqrt{\sum_{m=1}^M \sum_{n=1}^N (\mathbf{X}_{mn} - \mathbf{Y}_{mn})^2} \quad (6)$$

欧氏距离 D 的值越小,失真度就越小,即训练语音与测试语音越接近;反之,欧氏距离 D 的值越大,失真度就越大,表明训练语音与测试语音出于同一人的可能性就越小.

4 识别实验

硬件环境:PC 个人计算机(Intel(R) Core(TM) i5-2450M CPU @ 2.50 GHz);麦克风+声卡+普通扬声器.

软件环境:操作系统(Windows 7);开发平台(MATLAB R2010a);录音软件(CoolEditpro-v2.1).

实验环境:普通实验研究室.

本文实验采用自建的小型汉语普通话语音数据库.为适应实验的需要,结合实际情况,语音文件在普通研究室环境下录制,录音时尽量保持安静的环境,避免不必要的噪声干扰,采样频率为 16 kHz,量化精度为 16 bit.录音者(包括 11 个男性、14 个女性,共 23 人)进行约 3 分钟长度的随机朗读(与文本无关).截取训练语音与测试语音均为 3 s,进行截取时,避免所截取的语音段为同一部分语音数据,以提高本文后续实验的准确性.

选取 6-22 维的 Δ MFCC, 4-20 维的 WMFCC, 4-20 维的 IWMFCC, 均为 17 维的特征参数, 进行同维度参数分量扩张^[7]. 矢量量化码本容量为 16, 随机抽取一录音者的测试语音, 将抽取的测试语音与 23 人的训练语音进行识别, 图 5 为 Δ MFCC 分别与 WMFCC、IWMFCC 混合参数的欧氏距离计算结果.

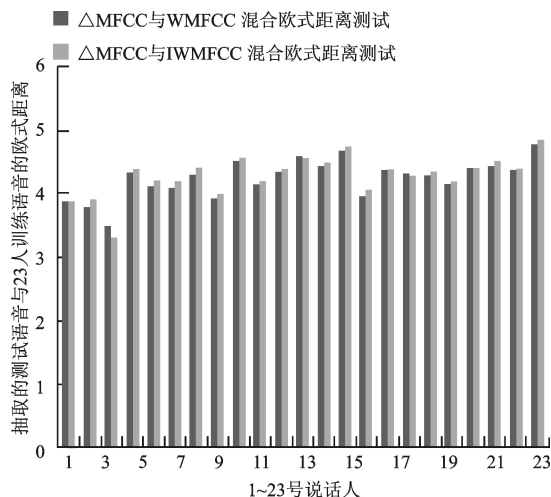


图 5 3 号录音者两种混合特征参数的欧式距离测度

由图 5 可以看出两种混合参数都能正确识别抽取的 3 号测试语音与 3 号训练语音是同一说话人, 通过图示欧氏距离可以看出, 对于正确识别的 3 号录音者, Δ MFCC 与 IWMFCC 混合参数的欧式距离比 Δ MFCC 与 WMFCC 混合参数的距离要低; 而 3 号测试语音与其他训练语音相识别时, Δ MFCC 与 IWMFCC 混合参数的欧式距离普遍高于 Δ MFCC 与 WMFCC 混合参数的欧式距离, 这表明 Δ MFCC 与 IWMFCC 混合特征参数的识别鲁棒性更好.

接下来取不同码本容量进行说话人识别实验^[8], 码本容量分别取 8、16、32, 表 1 列出了 MFCC、 Δ MFCC、WMFCC、IWMFCC、 Δ MFCC 与 WMFCC 组合以及 Δ MFCC 与 IWMFCC 组合的识别率.

表 1 不同码本容量下各特征参数的识别 %

	8	16	32
MFCC	78.26	82.61	86.96
Δ MFCC	82.61	86.96	86.96
WMFCC	82.61	86.96	91.30
IWMFCC	86.96	91.30	91.30
Δ MFCC+WMFCC	91.30	95.65	95.65
Δ MFCC+IWMFCC	95.65	100	100

表 1 表明, 本文提出的加权函数获得的加权特征参数 IWMFCC 要优于 Δ MFCC 和传统加权所得

特征参数 WMFCC, 而 Δ MFCC 与加权特征参数组合的识别率比单一特征参数的识别率要高. 由于本文的语音库是在普通实验室环境下录制, 带有不同因素的环境噪音, 如: 桌椅声, 对着麦克风吹气的声音等, 噪声不可避免地影响了系统的识别性能. 在此环境下 Δ MFCC 与 IWMFCC 组合特征参数在码本容量为 16 和 32 时可以达到 100% 的识别率.

5 结束语

本文通过对 MFCC 特征参数分布的分析, 改进传统的加权函数, 基于差分 Mel 倒谱和加权 Mel 倒谱对于语音信号特征的不同表征, 将二者维度选择后进行组合获得混合参数. 选用 23 人小型语音库, 基于 VQ 的说话人识别系统进行说话人识别, 实验结果表明, 改进的加权函数所得的特征参数及其混合参数能够得到更高的识别率, 鲁棒性更好.

参考文献:

- [1] 赵力. 语音信号处理[M]. 北京: 机械工业出版社, 2012.
- [2] 李建文, 张晋平. 基于改进语音特征提取方法的语音识别[J]. 微电子学与计算机, 2009, 26(7): 230-233.
- [3] 王伟, 邓辉文. 基于 MFCC 参数和 VQ 的说话人识别系统[J]. 仪器仪表学报, 2006, 27(6): 2253-2155.
- [4] 吴迪, 曹洁, 王进花. 基于自适应高斯混合模型与静态听觉特征融合的说话人识别[J]. 光学精密工程, 2013, 21(06): 1598-1604.
- [5] 陈明义, 余伶俐, 朱晗, 等. 基于特征参数融合的语音情感识别方法[J]. 微电子学与计算机, 2006, 23(12): 168-171.
- [6] 吴朝晖, 杨莹春. 说话人识别模型与方法[M]. 北京: 清华大学出版社, 2008.
- [7] Zhou Ping, Li Xiaopan, Li Jie, et al. Speech emotion recognition based on mixed MFCC[J]. Computer Measurement & Control, 2013, 21(7): 108-123.
- [8] 周昆湘. 基于矢量量化的与文本无关的说话人确认系统的研究[D]. 长沙: 中南大学, 2007.

作者简介:

柯晶晶 女, (1988-), 硕士研究生. 研究方向为语音信号处理和说话人识别技术语音识别.

周萍 女, (1961-), 教授, 硕士. 研究方向为语音识别与智能控制研究.

景新幸 男, (1960-), 博士, 教授. 研究方向为语音识别及其混合集成电路设计等研究.

杨青 女, (1976), 硕士, 副教授. 研究方向为语音识别算法与仿真.