

改进型 Mel 混合参数应用于说话人识别

黄 婷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 541001)

摘 要: 在实验室录音设备条件下, 提取基于 Mel 的特征参数及其差分倒谱系数 (DCC) 和差分频谱倒谱系数 (DSCC), 利用增减分量法进行有效融合, 建立基于 VQ 的说话人识别模型, 采用 LBG 算法设计不同码本容量进行说话人识别实验, 并在各类噪声环境下验证混合参数性能. 实验结果表明, 与单一参数及其他融合参数相比该混合特征参数具有更高的识别率和鲁棒性.

关键词: 说话人识别; VQ; 混合参数; 鲁棒性

中图分类号: TP3

文献标识码: A

文章编号: 1000-7180(2016)04-0060-04

Speaker Recognition Based on Improved Mel Hybrid Parameters

HUANG Ting¹, 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: With the recording voice library in the laboratory, we extract Mel frequency cepstral coefficient, its delta-cepstral coefficient and delta-spectral cepstral coefficient, and fuse them with change component method effectively. And a speaker recognition model is established based on VQ model using LBG algorithm to design different codebook capacity for the experiment and we evaluate the hybrid feature parameters in noise environments. Experimental results show that by applying the fusion hybrid feature parameters the recognition system rate and robustness are obviously improved compared with other parameters.

Key words: speaker recognition; VQ; hybrid feature parameters; robustness

1 引言

随着现代生活信息交流的多元化发展, 身份鉴定成为了人们参与社会活动的重要环节^[1]. 高精度的科技数码产品对身份验证技术有了更高的要求, 不仅要求高安全性, 同时要求验证便捷经济. 说话人识别技术相比较其他生物认证技术如虹膜识别、指纹识别、掌形识别等, 具有很多应用上的优势, 诸如使用方便、成本低等, 在计算机科学技术领域受到越来越广泛的关注^[2].

说话人识别技术发展到现在, 虽然已经达到很

高的精度, 但仍有许多不完善的地方有待发展, 实际应用中也存在许多难点^[3-4]. 例如语音信号受到通信环境的噪声干扰, 尤其是训练和测试语音易受到干扰而造成失配, 从而降低识别性能. 通过抑制通道畸变可改进识别系统的性能, 典型的方法有基于 Mel 参数的一阶差分倒谱系数 (Delta-cepstral Coefficients). DCC 在 ASR 系统中的应用已经较为广泛, 虽然在识别语音动态信息时有较高准确性, 但在噪声及混响条件下鲁棒性相对较差, 文献^[5]针对此问题提出一种基于 Mel 参数的差分频谱倒谱系数 (Delta-spectral Cepstral Coefficients). 本文将

DCC、DSCC 与 MFCC 基于增减分量法有效地相融合,组成一种新的混合参数以克服单一参数的不足,在 VQ 说话人识别模型下进行实验.实验结果表明,该混合参数相比于其他参数有更高的识别率,且改进后的混合参数鲁棒性及噪声补偿性能较好.

2 基于 Mel 的一阶差分倒谱系数

语音是一种连续的信号,常用的特征参数特征如 MFCC 是从一帧帧语音中提取的瞬时特征,虽然充分考虑了人耳的听觉感知特性,但其只能反映语音特征参数的静态特性,无法描述帧间的变化情况,尤其在低信噪比情况下容易导致参数与训练模板失配,从而影响识别率^[6].然而提取其一阶差分倒谱系数 DCC,就可以很好描述帧间变化的动态特征^[7].DCC 能够描述语音信号帧间瞬时变化的动态特征,在求 DCC 前先计算短时谱对数

$$\log |S(\omega, t)| = \sum_{m=-\infty}^{\infty} c_m(t) e^{-j\omega m} \quad (1)$$

式中, $S(\omega, t)$ 表示语音的短时频谱,其中 $C_m(t)$ 表示 t 时刻第 m 阶语音特征倒谱系数,语音短时谱的变化率如下:

$$\partial \frac{\log |S(\omega, t)|}{\partial t} = \sum_{m=-\infty}^{\infty} \frac{\partial c_m(t)}{\partial t} e^{-j\omega m} \quad (2)$$

通常采用差分倒谱来描述这种动态特征,用有限窗函数内的多项式来求倒谱系数的微分,一般使用汉明窗.差分系数表示如下:

$$d_{\text{cep}}(i, t) = C_{\text{cep}}(i, t-1) - C_{\text{cep}}(i, t) \quad (3)$$

式中, d_{cep} 表示动态特征, C_{cep} 表示倒谱特征, $i = 0, 1, 2, \dots, P, t = 1, 2, \dots, K$ (P 为特征阶数, K 为特征向量个数).差分倒谱变化可以对连续多帧语音进行多项式拟合,然后使用该多项式导数的参数作为差分特征.变换形式的差分倒谱比简单的连续帧差分倒谱效果更好,它是在良好的时间分辨率和和表示时间变化的动态特征的折中.如下式:

$$\begin{aligned} d_{\text{cep}}(i, t) &= \alpha \times \sum_{k=-K}^K k \times [C_{\text{cep}}(i, t-k) - C_{\text{cep}}(i, t+k)] \\ &= \alpha \times \sum_{k=-K}^K k \times C_{\text{cep}}(i, t-k) \end{aligned} \quad (4)$$

3 基于 Mel 的差分频谱倒谱系数

如图 1(a)所示为一阶差分倒谱系数(DCC)框图,图 1(b)所示为频谱差分倒谱系数(DSCC)框图,与图(a)相比,图 1(b)在系统线性化之前进行时间差分操作,并且应用了多维高斯非线性化.显然,在频谱域进

行差分可以增强快速变化的语音信号,补偿噪声的影响,降低因噪声信号干扰而造成的测试语音的失配.

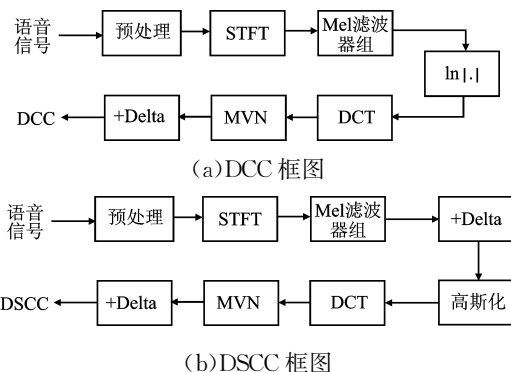


图 1 一阶差分倒谱系数及差分频谱倒谱系数框图

本文采用差分频谱倒谱系数特征参数来提高白噪声性噪比(SNR),均值为 0,方差为 σ ,自由度为 N ,高斯白噪声能量为 P ,则有

$$P = \frac{1}{N} \sum_{i=1}^N |\omega_i|^2 \quad (5)$$

$$E[P] = \frac{1}{N} E\left[\sum_{i=1}^N \omega_i^2\right] = \sigma^2 \quad (6)$$

可得方差(Variance)如下

$$\begin{aligned} \text{Var}[P] &= E[P^2] - E[P]^2 = \frac{E\left[\sum_{i,j} \omega_i^2 \omega_j^2\right]}{N^2} - \sigma^4 = \\ &= \frac{1}{N^2} \left(\sum_i E[\omega_i^4] + \sum_{i,j, i \neq j} E[\omega_i^2 \omega_j^2] \right) - \sigma^4 = \frac{2\sigma^4}{N} \end{aligned} \quad (7)$$

由此可知,能量 P 近似服从于正态分布 $N(\sigma^2, \frac{2\sigma^4}{N})$,即 P 的均值为 σ^2 ,方差为 $\frac{2\sigma^4}{N}$.如图 2(a)所示为原始信道语音样本在进行一阶差分非高斯化的短时能量谱图,图 2(b)所示为进行一阶差分后高斯化的短时能量谱图.由此可见,经过高斯化的语谱图信号更为平稳,可以更好地描述原始的信号分布.

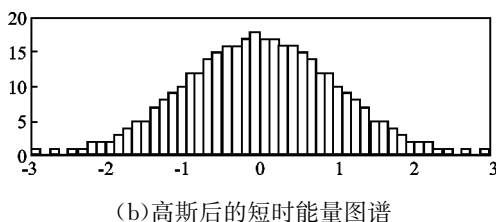
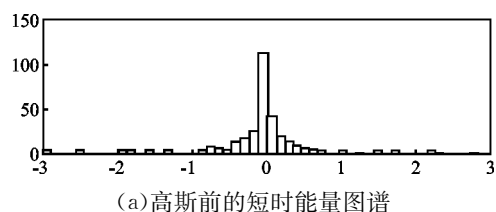


图 2 短时能量语谱图

4 增减分量法

不同类型的特征参数表征着说话人不同的特性,这是由于不同特征参数由不同模型提取出.本文提出的动、静态特征融合是将两种或两种以上的特征参数进行维度扩张,从不同角度展现说话人个性,利用混合得到的新的特征参数进行说话人识别,由此得到更高的识别率,提高识别系统的性能.为了有效地将特征参数融合,获得更高的识别率,引用增减分量法的概念.通过在特征矢量中增加或去掉某一分量,考核识别率的变化来确定该分量的贡献^[8].增减分量法是一种有效计算每阶倒谱分量平均贡献的算法,具体公式如下:

$$R(i)=\frac{1}{n}\left[\sum_{j>i}(p(i,j)-p(i+1,j))+\sum_{j>i}(p(j,i)-p(j,i-1))\right] \tag{8}$$

式中, $R(i)$ 表示第*i*阶倒谱分量的平均贡献值, n 为倒谱阶数, $p(i,j)$ 为从第*i*阶到第*j*阶倒谱系数特征的识别率.

本文使用增减分量法计算 MFCC、DCC 和 DSCC 各倒谱分量对语音识别的贡献,将三种特征参数中对识别率贡献高的 n 维倒谱分量组合到一起,按照 MFCC + DCC, MFCC + DSCC, MFCC + DCC + DSCC 三种方式构成新的混合参数进行说话人识别实验.

5 基于 VQ 的说话人识别系统

5.1 VQ 模型

矢量量化 VQ(Vector Quantization)是一种高效的数据压缩和编码技术^[9],能够有效地利用矢量中各元素间的相关性保存信号基本特征,并去除冗余信息.设维数为 k ,尺寸为的矢量量化器 Q 定义映射 $Q:R^k \rightarrow C$,即 k 维空间 R^k 到包含 N 个输出(重构)点的有限集合 C 的映射,其中 $C = \{y_0, y_1, \dots, y_{N-1}\}$, $y_i \in R^k, i \in \tau \equiv \{0, 1, \dots, N-1\}$. 集合 C 称作码书或码本,容量大小为 N . 码本 C 中的元素 y_i 称作码字或码矢量. 当任意空间的矢量 x_v 输入 Q 时,首先需要判断 x_v 属于 Q 的哪个子空间 R_w ,在确定后即可知道该空间对应的代表矢量 y_w ,然后对 x_v 进行量化成 y_w .

将 VQ 应用于说话人识别时,矢量量化的码本设计是关键.由于 LBG 算法压缩比大且失真较小,因此本文选取目前码本训练性能比较好的 LBG

(LINE-BUZO-GRAY) 算法在 VQ 训练中进行矢量量化码本设计.失真测度 $d(x,y)$ 表征输入矢量 x 被重构矢量(即码字) y 量化而产生的非负失真,均方误差欧式距离失真测度是矢量量化的语音信号处理最常用的方法,如式(9):

$$d_2(X,Y)=\frac{1}{N}\sum_{i=0}^{N-1}(x_i-y_i)^2 \\ =\frac{(X-Y)^T(X-Y)}{N} \tag{9}$$

VQ 应用于说话人识别的过程时,在训练阶段通过聚类算法将训练语音的特征参数构成一个由向量表示的码本,每个说话人对对应一个码本.

5.2 说话人识别原理

说话人识别主要有训练阶段和识别阶段两个部分.训练阶段主要包括预处理和特征提取,预处理主要是去除语音信号中的噪声并提升高频部分.本文在进行特征提取后,建立 VQ 码本,得到训练模型.识别阶段用待测参数与所有特征参数码本逐一进行失真测度计算,选择失真距离最小的参数作为识别结果.系统框图如图 3 所示.

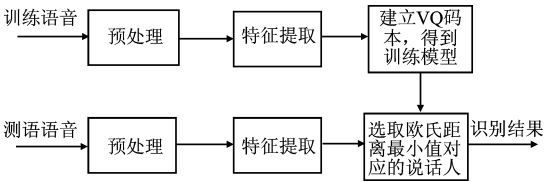


图 3 基于 VQ 的说话人识别系统原理框图

6 实验结果分析

6.1 语音库

本文在普通研究室环境下录制语音文件,采用 CoolEdit 软件,采样频率为 8 kHz,量化精度为 16 bit,录音者(包括 15 个男性、16 个女性,共 31 人)进行约 3 分钟的随机朗读(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.

6.2 实验

实验一:进行与文本无关的说话人辨认实验.实验预处理过程,预加重系数为 0.95,选用汉明窗处理.使用 MFCC、MFCC + DCC、MFCC + DSCC 和 MFCC + DCC + DSCC 四种特征进行说话人识别实验,实验中 VQ 的码本容量 $M=8, 16, 32, 64$,实验结果如表 1 所示.

表 1 不同码本容量各特征的识别率 %

M	MFCC	MFCC+DCC	MFCC+DSCC	MFCC+DCC+DSCC
8	67.88	68.93	70.38	74.56
16	80.57	81.35	80.31	83.87
32	83.79	81.92	80.49	90.42
64	81.03	81.53	82.87	84.87

由表 1 可以看出,系统识别率随着码本增加而提高,而当码本到达一定数量时识别率趋于平缓,且码本增加会增大计算量,故去除各单一参数贡献率低的 6 维分量,保留 18 维特征分量,按照 MFCC+DCC、MFCC+DSCC 和 MFCC+DCC+DSCC 三种特征混合方式,得到混合特征参数,进行说话人识别实验,表 2 为说话人识别结果.

表 2 改进后不同码本容量各特征的识别率 %

M	MFCC	MFCC+DCC	MFCC+DSCC	MFCC+DCC+DSCC
8	73.85	74.19	77.69	77.69
16	76.34	83.42	83.87	87.17
32	84.45	85.87	90.32	93.55
64	85.57	90.32	87.10	90.46

实验二:不同噪声下基于混合参数 MFCC+DCC+DSCC 的识别率. 为了进一步验证改进后的混合参数的鲁棒性及有效性,本文利用白噪声、音乐噪声、市场噪声、回声环境,在不同信噪比(SNR)及回声情况下对改进前及改进后的混合参数进行识别实验,实验结果如图 4 所示.

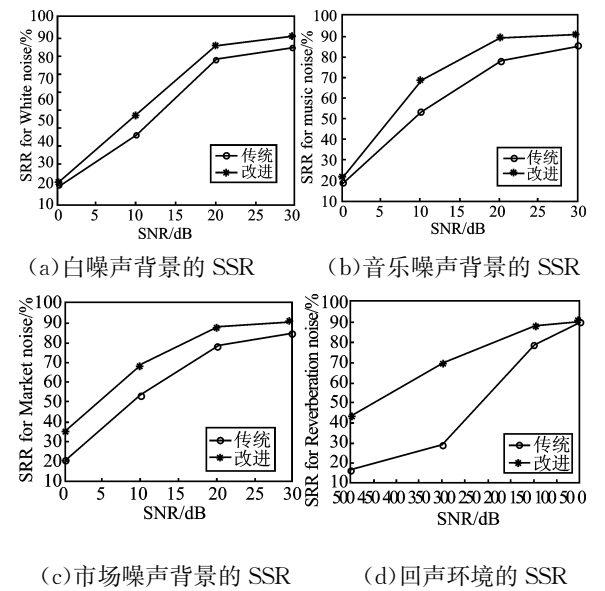


图 4 不同噪声下改进前后的 SRR 对比

由图 4 中(a)、(b)、(c)可以看出,基于改进的混合参数在 SNR 为 10 dB 时识别率都至少有 10%左右的提升,在高信噪比环境下也有 5%以上的提升,

尤其在音乐噪声下,10~20 dB 时有 15%左右的提升;由(d)可看出,在回声为 300~500 ms 时识别率有 26%~40%的提升. 由此可见,改进的混合算法应用于说话人识别系统对于噪声补偿具有良好的性能.

7 实验结果分析

本文使用实验室环境下的语音库,建立基于 VQ 的说话人识别模型,在不同码本容量下,采用单一特征参数和混合特征参数进行实验,并在白噪声及音乐噪声下验证改进后混合参数的鲁棒性. 对于混合特征参数的提取,本文采用增减分量法分析特征参数中各倒谱分量对识别系统识别率的影响,选取倒谱分量中对识别率贡献高的分量组成混合参数. 实验结果表明,本文获取的混合特征参数具有较高的识别率和鲁棒性,并具有良好的噪声补偿性能.

本文的识别结果都是在稳定的实验室环境下得到的,在实际应用中可能会受到其他噪声及信道的干扰从而影响识别率. 因此,结合现已有技术的优势,继续研究混合参数在说话人识别中的应用,如系统的稳定性,都将是今后工作的重点.

参考文献:

[1] 蒋晔. 基于短语音和信道变化的说话人识别研究[D]. 南京:南京理工大学,2013.

[2] 宋乐. 基于说话人识别中改进特征提取算法的研究[D]. 太原:太原理工大学,2014.

[3] Chanwoo Kim, Richard M Stern. Power-normalized cepstral coefficients (PNCC) for robust speech recognition[C]// IEEE International Conference on Acoustics, Speech and Signal Processing- Proceedings. [s. l.]: IEEE, 2012: 4101-4104.

[4] 王书诏,邱天爽. 说话人识别研究综述[J]. 电声技术, 2007,31(1):51-55.

[5] Kshitiz Kumar, Chanwoo Kim, Richard M Stern. Delta-spectral cepstral coefficients robust speech recognition[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. [s. l.]: IEEE, 2011:4784-4787.

[6] 胡政权,曾毓敏,宗原,等. 说话人识别中 MFCC 参数提取的改进[J]. 计算机工程与应用, 2014, 50(7): 217-220.

[7] 张宇. 基于倒谱特征的说话人识别方法研究[D]. 大连:大连海事大学,2010.

功能点进行充分的测试.

表 2 各个模块的 sequence 重用

顶层序列	模块级序列	描述
u2a_incr_payload	uart_ctrl_config_reg_seq	UART 的寄存器配置序列
	uart_incr_payload_seq	UART UVC 向 UART 发送数据
	intrpt_seq	AHB UVC 读取 UART 接收缓存器中的数据
	ahb_to_uart_wr	AHB UVC 向 UART UVC 发送数据
apb_spi_incr_payload	spi_cfg_reg_seq	SPI 的寄存器配置序列
	spi_incr_payload	SPI UVC 向 SPI 发送数据
	read_spi_rx_reg	AHB UVC 读取 SPI 接收缓存中的数据
	ahb_to_spi_wr	AHB UVC 向 SPI UVC 发送数据
	spi_en_tx_reg_seq	SPI 数据发送使能
apb_gpio_simple_vseq	gpio_cfg_reg_seq	GPIO 的配置寄存器序列
	gpio_simple_trans_seq	GPIO UVC 向 GPIO 发送数据
	read_gpio_rx_reg	AHB UVC 读取 GPIO 接收缓存器中的数据
	ahb_to_gpio_wr	AHB UVC 向 GPIO UVC 发送数据

5 结束语

UVM 验证方法学统一了各大验证方法学,不仅功能强大,而且架构层次清晰,在可重用、自动化、可扩展、规范化等方面具有很大的优势.本文阐述了UVM 验证方法学中的纵向可重用方法,并以 APB 总线的验证为例搭建了一个基于模块重用的验证平台.这种纵向的重用能够很好的利用前期的资源,大大减少了验证的时间和精力,显著提高了验证效率,在实际应用中具有很高的利用价值.

参考文献:

[1] SalahK. A UVM-based smart functional verification platform;Concept, pros, cons, and opportunities[C]// Design &. Test Symposium (IDT), 9th International. Dead Sea; IEEE,2014;94-99.

[2] 李博. 基于 system verilog-vmm 的仿真环境设计及其应用[D]. 哈尔滨:哈尔滨工业大学,2009;3.

[3] 张强. UVM 实战[M]. 北京:机械工业出版社. 2014;4.

[4] Drechsler R, Chevallaz C, Fummi F, et al, Panel: Future SoC verification methodology: UVM evolution or revolution[C]// Design, Automation and Test in Europe Conference and Exhibition (DATE). Dresden; DATE,2014;1-5.

[5] Yun Young-Nam. Beyond UVM for practical SoC verification[C]//SoC Design Conference (ISOC), 2011. Jeju: ISOC,2011;158-162.

作者简介:

熊 涛 男,(1989-),硕士研究生,研究方向为数字集成电路物理设计与验证. E-mail:xiongtao@ime. ac. cn.

蒋见花 女,(1976-),副研究员. 硕士生导师,研究方向为数字集成电路低功耗设计.

(上接第 63 页)

[8] 姜占才,孙燕,姚刚. 模糊聚类与 LBG 级联的 VQ 算法[J]. 计算机工程与科学,2011, 33(5) : 155-159.

[9] 毛旭. 说话人检索系统的研究与实现[D]. 大连: 大连理工大学,2009.

作者简介:

黄 婷 女,(1990-),硕士研究生. 研究方向为说话人识别.

E-mail:tilitha@163. com.

周 萍 女,(1961-),硕士,教授. 研究方向为语音识别、智能控制.

景新幸 男,(1990-),博士,教授. 研究方向为信号与信息处理.

杨 青 女,(1976-),硕士,副教授. 研究方向为智能处理与信息处理.