

基于多特征有效组合的说话人识别

谢迎春¹, 于湘珍², 刘建平³, 张卫华⁴

- (1. 武警工程学院 研究生队 陕西 西安 710086; 2. 武警工程学院 电子技术基础实验室 陕西 西安 710086;
3. 武警工程学院 通信工程系 无线通信工程教研室 陕西 西安 710086;
4. 武警工程学院 通信工程系 信息工程教研室 陕西 西安 710086)

摘要: 通过分析当今说话人识别系统中常用的一些特征参数, 以提高说话人识别的识别率为目的, 在 Matlab 6.5 软件环境下提出了将 Mel 频率倒谱 (MFCC)、线性预测倒谱 (LPCC) 及他们的一阶差分 and 基音周期等多种特征有效结合进行说话人识别的方法。采用短时自相关法提取基音周期, 在识别过程中采用改进的动态规整算法, 将模板的匹配过程与检验量的计算分离开, 每帧给出一个说话人辨认结果, 最后综合各帧的辨认结果, 得出最佳匹配结果。经过多次实验证明, 采用以上方法使用多特征有效结合比单个使用各种特征效果要好, 能在一定程度上提高系统区分说话人的能力。

关键词: 说话人识别; 动态规整; MFCC; LPCC; 基音周期

中图分类号: TN 912

文献标识码: B

文章编号: 1004-373X (2005) 09-068-03

Speaker Identification Based on Efficiently Combining Manifold Features

XIE Yingchun¹, YU Xiangzhen², LIU Jianping³, ZHANG Weihua⁴

(1. Graduate Student Team, College of Armed Police Force, Xi'an, 710086, China;

2. Electron Technique Basic Lab, College of Armed Police Force, Xi'an, 710086, China;

3. Teaching Chamber of Wireless Communication Engineering in Communication Engineering Department, College of Armed Police Force, Xi'an, 710086, China;

4. Teaching Chamber of Computer Information Project in Communication Engineering Department, College of Armed Police Force, Xi'an, 710086, China)

Abstract: Through analyzing some features that be used usually in speaker identification system nowadays, in order to improve the rate of identification, this paper puts forward a method that combining efficiently more features such as MFCC and LPCC and their one ranks coefficients and keynote period and so on to do speaker verification under Matlab 6.5. We pick up keynote periods by self-correlation method and use a new Dynamic Time Warping (DTW) method to do identification. This new DTW method is a way that dividing template matching and calculation of test measure and calculating identification results of all frames after every frame getting out a identification result. At last, we can make out the best matching result. Through series of experiments, it proves that the method of using manifold features is better than the method of using single feature and the ability of speaker identification can be improved by using this way.

Key words: speaker verification; DTW; MFCC; LPCC; keynote period

1 引言

说话人识别是语音识别的一个分支, 在公安侦察、声控系统、医疗诊断、电子金融业务等方面有着广泛的应用前景。他和语音识别的区别在于, 他并不注意语音信号中的语义内容, 而是希望从语音信号中提取出个人的信息特征。从这点上说, 说话人识别是谋求挖掘出包含在语音信号中的个性因素, 而语音识别是谋求从不同人的语音信号中寻找共同因素。

通过分析前人对说话人识别的工作总结, 可以得出不同人的发音特征可以用基音周期、Mel 频率倒谱 (MFCC)、线性预测倒谱 (LPCC) 及其一阶差分系数等多

种特征来描述。实验表明, 这样的特征组合对提高说话人识别率是有效的。为了进一步提高识别率, 本文采用了改进的动态规整 (DTW) 方法, 在单帧确认的基础上结合置信度估计进行整词确认^[1]。在增加少量运算代价的情况下, 新方法改善了辨认系统的性能。

2 特征参数的提取

在提取特征之前, 所采集的信号必须经过预处理, 一般包括预加重、加窗和分帧。当然, 为减少计算量提高计算精度, 在预处理后要进行端点检测。本文利用语音短时能频值作为端点检测的参数, 这种方法相当于在传统方法中, 以背景噪声的短时能频值作为基准对绝对门限值做调整。结果表明能频值端点检测的方法适应环境的能力比较强, 准确率较好^[2]。

收稿日期: 2005-01-09

2.1 Mel倒谱系数 (MFCC)

与普通实际频率倒谱分析不同, MFCC (Mel - Frequency Cepstral Coefficients) 的分析着眼于人耳的听觉特性, Mel 频率尺度的值大体上对应于实际频率的对数分布关系, 更符合人耳的听觉特性^[3]. Mel 频率与实际频率的具体关系可表示为:

$$\text{Mel}(f) = 2595 \lg(1 + f/700) \quad (1)$$

这里实际频率 f 的单位是 Hz.

求取MFCC 系数的具体过程如下:

(1) 首先确定每一帧的点数, 本系统采用一帧点数 $N = 256$ 个点, 帧移为 128 点, 对每帧序列进行预加重处理后再经离散 FFT 变换, 取模的平方得离散功率谱 $S(n)$.

(2) 利用临界带通滤波器组技术^[4], 采用滤波器个数为 $M = 24$ 的三角滤波器组 $H_m(n)$, 根据式(1) 将实际频率尺度转换到 Mel 频率尺度, 计算 $S(n)$ 经过此滤波器的功率值, 得到 M 个参数 $p_m, m = 1, 2, \dots, M$.

(3) 计算 p_m 的自然对数, 得到 $L_m, m = 1, 2, \dots, M$.

(4) 对 L_m 进行离散余弦变换, 得到MFCC 参数.

在为每帧计算出MFCC 参数后, 通常要将 M 个 MFCC 参数乘以不同的权系数, 以改善低信噪比时信号的特征性能. 由于标准的MFCC 参数只反映了语音参数的静态特征, 而人耳对动态的语音特征又比较敏感, 所以计算能描述语音动态特性的参数 MFCC 的一阶差分 (MFCC), 与MFCC 参数共同组成一个特征矢量, 作为一帧语音信号的特征参数^[2].

2.2 线性预测倒谱 (LPCC)

线性预测倒谱系数是一种非常重要的特征参数. 他的主要优点是比较彻底地去掉了语音产生过程中的激励信息^[2], 主要反映声道相应, 而且往往只要十几个倒谱系数就能较好地描述语音信号的共振峰特性, 因此在语音识别中取得了较好的效果. 在实际计算中, LPCC 参数不是由信号直接得到的, 而是由LPC 系数得到的. 关系式如下:

$$\begin{cases} c_0 = \log G^2 \\ c_m = a_m + \sum_{k=1}^{m-1} \frac{k}{m} c_k a_{m-k} & 1 \leq m \leq p \\ c_m = \sum_{k=1}^{m-1} \frac{k}{m} c_k a_{m-k} & m > p \end{cases} \quad (2)$$

这里 c_0 实际上是直流分量, 反映频谱能量, 其值的大小不影响谱形, 在识别中通常不用, 也不去计算. 当LPCC 系数不大于LPC 系数时用第二式, 当LPCC 系数大于LPC 系数时, 用第三式进行计算. 本系统的线性预测模型阶数 $p = 12$, LPCC 的阶数为 $m = 16$.

由于标准的LPCC 只能反映声道参数的静态特征, 而同一个人声道参数的变化比不同人声道参数的变化要敏感, 因此采用反映声道参数的动态特征的特征参数: LPCC 一阶差分 (LPCC), 作为语音特征矢量的一个分量.

2.3 基音周期

在说话人识别中, 基音周期是一个重要的特征参数. 基音频率取决于声带的大小、厚薄、松紧程度以及声门上下之间的气压差的效应等, 其范围约为 60~ 450 Hz. 实验表明, 基音周期特征独立用于说话人识别效果并不理想. 当与其他特征组合之后, 才对说话人有较强的区分性. 本文将基音周期 (T) 作为特征矢量的另一个分量. 采用的是短时自相关方法^[3] 提取基音周期, 分别与MFCC 和LPCC 参数序列组合成为两个特征矢量.

3 识别方法

3.1 DTW 动态时间规整匹配法

DTW (Dynamic Time Warping) 动态时间规整匹配, 基于动态规划的思想, 解决了发音长短不一的匹配问题, 是语音识别中出现较早、较为经典的一种算法. 本文的系统把采用DTW 的双模板特定人孤立词识别器^[3] 作为原型. 在训练阶段, 从两遍训练语音中逐帧提取的特征以矢量序列的形式存放为模板. 在辨认阶段, 测试语音的特征矢量序列通过DTW 方法与所声明用户的两个口令模板分别匹配. 设测试模板的特征矢量序列是 $X = (x_1, x_2, \dots, x_M)$, 参考模板的特征矢量序列是 $Y = (y_1, y_2, \dots, y_N)$. 由于语音信号各段在不同的情况下长短不一定相同, 因此实际中更多采用动态规划(DP) 方法^[3].

将测试模板的各帧 $n = 1, 2, \dots, N$ 作为二维直角坐标系的横轴, 参考模板的各帧号 $m = 1, 2, \dots, M$ 作为纵轴, 这样形成一个网络, 网络中的每个交叉点 (n, m) 就是测试模板与参考模板某一帧的交汇点. DP 算法可以归结为寻找一条通过此网络若干格点的路径, 路径经过的格点就是参考模板和测试模板进行距离计算的帧号. 由于语音发音的快慢可以改变, 但各部分的先后顺序不可改变, 因此所选路径必须从左下角出发到右上角结束, 如图 1 所示. 帧间匹配距离采用欧氏距离的平方:

$$d(X_i, Y_j) = \sum_{n=1}^N (x_{i,n} - y_{j,n})^2 \quad (3)$$

其中: $X_i = (x_{i,1}, x_{i,2}, x_{i,3}, \dots, x_{i,N})$, $Y_j = (y_{j,1}, y_{j,2}, y_{j,3}, \dots, y_{j,N})$, N 是特征矢量维数.

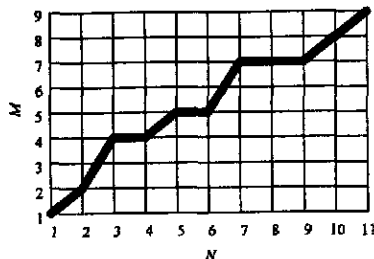


图 1 DTW 算法搜索路径

如果路径通过了格点 (n_{i-1}, m_{i-1}) , 那么下一个路过的格点 (n_i, m_i) 只能是以下 3 种情况之一:

$$(n_i, m_i) = (n_{i-1} + 1, m_{i-1} + 2)$$

$$(n_i, m_i) = (n_{i-1} + 1, m_{i-1} + 1) \quad (4)$$

$$(n_i, m_i) = (n_{i-1} + 1, m_{i-1})$$

搜索最佳路径的方法如下: 搜索从 (n_1, m_1) 点出发, 展开若干条满足条件式(4)的路径, 假设可以计算每一条到达 (n_N, m_N) 的累积距离, 具有最小累积距离的路径则为最佳路径。这套DP算法就是DTW算法。

3.2 改进的DTW 动态时间规整

标准DTW方法在匹配矢量序列的同时计算匹配距离作为判决检验量。现在将模板的匹配过程与检验量的计算分离开, 仍通过DTW进行语音的匹配, 然后通过新的方法计算检验量进行判决。新方法采用在帧和词条两个层次对说话人进行辨认的结构: 首先从每帧给出一个说话人辨认结果, 然后将各帧的辨认结果进行综合, 给出一个整体的结果^[1]。

3.3 MFCC与LPC及基音周期的组合判决

MFCC美尔倒谱反映的是说话人语音的听觉频率非线性特性, LPC线性预测倒谱反映的是说话人声道生理结构的差异, 而他们的一阶差分系数 ($\Delta MFCC$, ΔLPC) 都描述了各自的动态特性。同时基音周期 (T) 包含大量的说话人固有特征信息, 把基音周期分别加入前两者特征序列构成两个特征矢量, 可以提高识别率^[3]。利用MFCC参数识别时, 如果有人盗用密码, 系统容易判该人为合法人员。因此引入反映声道响应的LPC及其一阶差分同基音周期共同作为另一特征参数, 提高系统的安全性。识别过程中, 首先将组合的两个特征矢量分别用DTW动态规整算法进行模板匹配, 为减少误判设置模板匹配距离阈值 p , 若该语音模板匹配距离 $d > p$, 则认为该语音人员为非法人员, 反之为合法人员。对比两个特征矢量匹配得出的最小距离脚码 i, j : 若 $i = j$, 则判该语音人员为合法人员, 反之则为非法人员。

4 识别实验

4.1 数据来源及数据的预处理

由于国内目前尚无说话人识别标准语音数据库, 本实验所用的语音数据均在普通实验室环境下, 采用 Sound-blaster16 声霸卡和耳机自带话筒录音采集, 录音软件为 Sonic Foundry Sound Forge 6.0。为克服采样率与识别计算量的矛盾, 根据多次实验对比, 本实验采样频率为 11 025 Hz。录音数据按帧长 256 点, 帧移为 128 点, 预加重为 $1 - 0.95Z^{-1}$, 加汉宁窗逐帧计算基音周期, 16 阶 MFCC 系数和 $\Delta MFCC$ 、12 阶 LPC 系数和 ΔLPC 。

4.2 文本相关说话人辨认系统的实现

本实验采用 Matlab 6.5 作为开发环境, 分别针对不同的特征矢量用DTW动态规整识别法做了三类实验。本文的测试以正识率^[5,6]为评价系统识别性能的标准。

实验一 以 16 阶 MFCC 系数及其一阶差分 $\Delta MFCC$

和基音周期 (T) 合成的 33 维特征序列为特征矢量。

方式 1: 采样 27 个人的两组发音 (发音内容为 4~15 个字的汉语语句, 每人的两组发音内容相同, 不同人内容不同。

方式 2: 采样 27 个人的两组发音 (发音内容为 4~15 个字的汉语语句), 每人的两组发音内容相同, 不同人有说相同内容的。

方式 3: 采样 27 个人的两组发音 (发音内容为 4~15 个字的汉语语句), 每人的两组发音内容不同, 不同人有说相同内容的。

实验二 以 12 阶 LPC 系数及其一阶差分 ΔLPC 和基音周期 (T) 合成的 25 维特征序列为特征矢量。实验方式与实验一相同。

实验三 组合使用实验一、二的两类特征矢量, 用改进的 DTW 动态规整法做出判决。实验方式与实验一相同。

4.3 实验结果

实验结果如表 1 所示, 单独使用 MFCC 及 $\Delta MFCC$ 、基音周期和 LPC 及 ΔLPC 、基音周期不如实验三中组合使用时的效果好。LPC 弥补了 MFCC 不能描述的声道特性, 而 $\Delta MFCC$ 和 ΔLPC 反映了语音及声道的动态特性, 组合这些特征矢量可以较好地反映说话人的个性特征。另外, 改进的 DTW 动态时间规整法, 也在一定程度上提高了说话人识别系统的识别率。

表 1 文本相关说话人辨认实验结果

	MFCC, $\Delta MFCC, T$	LPC, $\Delta LPC, T$	MFCC, $\Delta MFCC$, LPC, $\Delta LPC, T$
不同说话人内容 互不相同(识别率)	96.30%	92.52%	97.12%
不同说话人内容 部分相同(识别率)	89.82%	87.65%	91.43%
同一说话人两组 语音内容不同(识别率)	81.48%	79.74%	82.54%

虽然采用多种特征参数, 增加了少量的计算量和运算时间, 却提高了文本相关说话人辨认系统的识别性能, 在实际应用中获得较好的效果。实验证明, 利用多个特征参数的有效组合, 采用改进的 DTW 动态规整识别方法在提高文本相关说话人辨认系统的识别性能方面是有效的。

参 考 文 献

- [1] 文学, 刘加, 刘润生. 一种改进的新型说话人确认算法 [J]. 清华大学学报 (自然科学版), 2003, 43 (1): 51-54.
- [2] 刘永红. 说话人识别系统的研究 [D]. 成都: 西南交通大学, 2003, (4).
- [3] 赵力. 语音信号处理 [M]. 北京: 机械工业出版社, 2003.
- [4] 杨行峻, 迟惠生. 语音信号数字处理 [M]. 北京: 电子工业出版社, 1995.
- [5] Atal B S. Automatic Recognition of Speakers from Their Voices [J]. Proceedings of IEEE, 1976, 64 (4): 460-475.

(下转第 73 页)

看作是极具开创性和发展前景的研究方向, 而该研究方向的主要内容正是专家系统在数据融合技术中的应用。专家系统在工程应用上经过二十多年的发展, 就其理论基础、系统设计和开发工具而言, 已经取得了较为全面而丰硕的成果, 正因为如此, 运用专家系统求解数据融合中的态势和威胁估计问题是非常适宜的。

下面以某指挥自动化系统为例, 给出利用专家系统实现态势和威胁估计中数据融合的具体实现模型如图3所示。

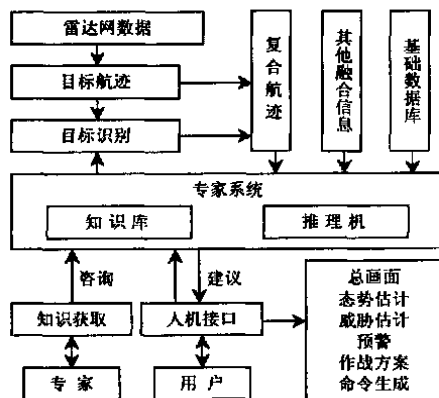


图3 专家系统实现 STA 模型

此模型充分发挥了专家系统以知识推理形式解决定性分析问题的特点, 又发挥了指挥自动化系统以数据计算解决定量分析问题的特点, 充分做到定性分析和定量分析的有机结合, 使得解决问题的能力 and 范围得到了很大的发展。系统既可以精确输出各种目标航迹和属性, 又可以输出态势和威胁的定性分析结果, 最终形成多个作战备选方案, 供指挥机关决策选择。

3 结 语

数据融合作为一门信息处理技术, 实际是涉及到决策论、认识论、模糊理论、估值论、通讯、数字信号处理、计

算机科学及人工智能等多学科理论知识, 是一门新兴的边缘交叉学科。军队指挥自动化系统采用了先进的电子技术、计算机技术、传感器技术, 目标环境信息的综合程度越来越高。只有通过有效的数据融合技术才能把从各个传感器和其他方面获得的信息准确自动合成, 减少信息损失, 提高环境态势合成度, 给指挥员提供准确的辅助决策。可以说, 是军队指挥自动化系统的发展需求牵引着数据融合技术理论的研究和应用, 数据融合技术的深化又推动着军队指挥自动化系统的进一步发展和提高。随着传感器、数据处理、计算机、网络通讯、人工智能、并行计算等技术的发展, 数据融合必将成为未来军事指挥控制系统智能检测与数据处理的重要技术。

参 考 文 献

- [1] 竺南直, 朱德成. 指挥自动化系统 [M]. 北京: 电子工业出版社, 2001.
- [2] 康耀红. 数据融合理论与应用 [M]. 西安: 西安电子科技大学出版社, 1997.
- [3] 徐涛, 杨国庆, 陈松灿. 数据融合的概念、方法及应用 [J]. 南京航空航天大学学报, 1995, 27 (2): 258-265.
- [4] Taur J S, Kung S Y. Fuzzy Decision Networks and Application to Data Fusion [J]. 0-7803-0928-6/93 IEEE.
- [5] 程岳, 王宝树, 李伟生. 实现态势估计的一种方法 [J]. 计算机科学, 2002, 29 (6): 111-113.
- [6] 张华生. 一种体系作战雷达网络的数据融合 [J]. 现代雷达, 2004, 26 (1): 1-4.
- [7] 朱敏, 游志胜, 聂健荪. 基于数据融合的雷达主监控系统的设计与实现 [J]. 计算机应用, 2003, 23 (2): 79-81.
- [8] 蔡自兴, 徐光佑. 人工智能及其应用 [M]. 北京: 清华大学出版社, 1996.

作者简介 杨福平 男, 1975年出生, 山西太原人, 硕士生。研究方向为人工智能, 指挥自动化, 神经网络, 数据融合。
白振兴 男, 1954年出生, 山西太原人, 教授。研究方向为人工智能, 软件理论。

(上接第70页)

- [6] Naik J M. Speaker verification: A tutorial [J]. IEEE Commucation Magazine, 1990, 28 (1): 42-48.
- [7] 李蕴华. 将倒谱参数与基音信息有效结合进行说话人辨认 [J]. 信号处理, 2000, 16 (3): 85-89.
- [8] ZHANG Wanfeng, WU Zhaozhui, YANG Yingchun, et al. Feature Combination For Speaker Identification [J], 2003, 21 (3): 10-15.
- [9] 范新伟, 申瑞民, 杜彦蕊. 用LPC及DTW进行语音模式比较的设计与实现 [J]. 计算机工程, 2004, 30, (1): 126-128.

作者简介 谢迎春 女, 1980年出生, 硕士研究生。研究方向为信号与信息处理, 智能信息处理。
于湘珍 女, 1964年出生, 工程师。研究方向为电子信息处理。
刘建平 男, 1967年出生, 教授。研究方向为生物医电, 数字信号处理。
张卫华 男, 1977年出生, 讲师。研究方向为指挥自动化。