

通信技术

说话人识别中改进的 MFCC 参数提取方法

何朝霞 潘 平

(贵州大学计算机科学与信息学院, 贵阳 550023)

摘 要 在说话人识别技术中,特征参数的提取对语音训练和识别有着非常重要的作用。而 Mel 频标倒谱系数 MFCC 是一种常用的特征,它能对语音信号进行分析处理,去除对语音识别无关紧要的冗余信息,获得影响语音识别的重要信息。同时由于语音信号具有时变和混沌特性,以非线性随机共振理论和人类对听觉的理解为基础,提出了一种基于随机共振的 MFCC 特征参数提取方法。通过实验比较两种方法的结果,论证了改进方法的可行性以及优越性,为说话人识别技术中特征参数提取提供了一条新的研究方向。

关键词 说话人识别 特征提取 MFCC 参数 随机共振

中图分类号 TN912.34;

文献标志码 A

语音特征提取是指从语音信号中获取一组能够描述语音信号特征参数的过程。在理想情况下,这些特征应该具有以下特点:相对保持稳定;不易被模仿;尽量不随时间和空间变化等。MFCC 参数具有良好的识别性能和抗噪能力,在目前的语音识别和说话人识别中极为常用。同时由于语音信号具有类混沌的特性^[1],因而,部分学者利用随机共振原理对语音信号进行处理^[2]。本文尝试了一种根据人耳听觉机理和随机共振理论^[3,4]相结合的方法,将随机共振理论运用到提取 MFCC 参数的过程中。对标准 MFCC 参数提取过程和改进的 MFCC 参数提取过程进行研究,并对其在说话人识别中的效果做出了实验比较。

1 标准的 MFCC 参数提取方法^[5,6]

在语音识别和说话人识别中,MFCC (Mel 频标倒谱系数)参数是将人耳的听觉感知特性和语音的

产生的机理相结合,是目前大多数语音识别中广泛使用的特征参数。人耳具有一些特殊的功能,这些功能使得人耳在嘈杂的环境中,以及各种变异情况下仍能够正常地分辨出各种语音,其中耳蜗起了关键的作用。耳蜗实质上相当于一个滤波器组,耳蜗的滤波作用是在对数频率尺度上进行的,在 1 000 Hz 以下为线性尺度,而在 1 000 Hz 以上为对数尺度,这使得人耳对低频信号比对高频信号更加敏感。根据这一原则,研究者根据心理实验得到了类似于耳蜗作用的一组滤波器组,这就是 Mel 滤波器组。

MFCC 参数的提取流程框图如图 1 所示,具体计算步骤如下。

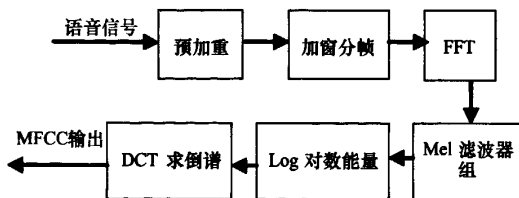


图1 标准 MFCC 参数提取流程图

(1) 语音信号经过预加重、加窗分帧处理后变为短时信号,用 FFT 将这些时域信号转化为频域信号。

2011 年 3 月 24 日收到

国家科技计划基金资助项目

(2008RR0003)、贵州省国际科技合作计划基金资助项目

([2009]700109,[2009]700125)资助

第一作者简介:何朝霞(1984—),女,湖北人,硕士研究生,研究方向:语音信号处理。

(2) 求出频谱平方,即能量谱,并通过 Mel 频率滤波器组得到 Mel 频谱,并通过对数能量的处理得到对数频谱。

(3) 将上述对数频谱经过离散余弦变换(DCT)得到 L 个 Mel 频率倒谱系数。MFCC 系数为

$$C_n = \sum_{k=0}^M \lg x'(k) \cos[\pi(k-0.5)], n = 1, 2, \dots, L。$$

(4) 将这种直接得到的 MFCC 特征作为静态特征,再将这种静态特征做一阶和二阶差分,得到相应的动态特征。

2 改进的 MFCC 参数提取方法

2.1 基本原理

世界上所有的声音都是物体振动产生的声波在介质中传递的结果,因此声音的产生离不开振动的动力、振动的源头和振动的共鸣腔。语音说到底也是一种声音,因而也离不开振动的动力、源头和共鸣腔,只不过语音是由人类的发音器官各部分协同动作所产生的,比如说由声带颤动而产生的声带音是通过喉腔、咽腔、口腔、唇腔和鼻腔这五个共振腔才传到人的耳朵里。

同时,人的听觉器官内耳中的耳蜗也相当于一个共振器。耳蜗是一螺旋形骨管,绕蜗轴卷曲两周半。整个耳蜗由耳蜗隔膜隔成三个区域,中间的隔膜叫基底膜,上部为瑞士膜,中间区域称为耳蜗导管,上下两个区域分别称为前庭阶和鼓阶,前庭阶和鼓阶在尖端部分相通。1960 年 Von Bekesy 用正弦信号对基底膜进行了详细的研究,原来,基底膜的听觉响应与刺激的频率有关,频率较低时,靠近耳蜗底部的基底膜产生响应,反之,频率较高时,靠近圆形窗的窄而紧的基底膜产生响应^[7]。

根据人的发音机理和听觉机理,我们认为可以用随机共振的原理对语音信号进行处理。从信号处理的角度来讲,在非线性的系统中,当输入带噪信号时,以适宜的物理量来衡量系统特性,如信噪比等,通过调节输入噪声强度或系统参数^[8],使系统特性达到一个最大值,此时,我们称信号、噪声和非线性随机系统产生的协同现象为随机共振。由于

说话人语音信号自身的特性——语音信号中包含有准周期性信号和类噪声信号。因而,从理论上利用随机共振理论,是可以获得说话人语音信号的基本特征。

2.2 改进的 MFCC 参数提取过程

从上述基本原理,我们考虑将随机共振的原理运用到 MFCC 参数的提取过程中,这样就更加符合人耳的听觉特性,以期能够分辨多种语音,它的提取过程如图 2 所示,具体步骤如下。

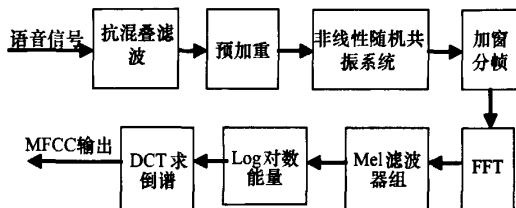


图 2 改进的 MFCC 参数提取过程

(1) 抗混叠滤波指滤除高于 $1/2$ 采样频率的信号成分或噪声,一般用低通滤波器实现。典型的抗混叠滤波器的技术指标是:通带内波动绝对值小于 1 dB,通带带宽 3 400 Hz,阻带衰减 50 dB 以上。

(2) 抗混叠滤波后,为了提升高频部分,使信号的频谱变得平坦,将滤波后的信号经过预加重数字滤波器 $H(z) = 1 - 0.9375z^{-1}$ 。

(3) 将预加重后的信号,经过非线性随机共振系统(如图 3),然后进行加窗分帧,这里使用的是汉明窗,再经过 FFT 变换得到频谱,并计算其短时功率谱。

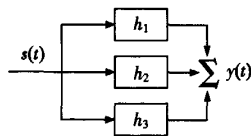


图 3 非线性随机共振系统

(4) Mel 滤波器组的设计。将短时功率谱从频域映射到 Mel 域,这里采用的 24 阶 Mel 滤波器组。

(5) 将短时功率谱经过滤波器后再作对数运算和 DCT,就可以得到 MFCC 参数了。

3 实验和分析

实验采用自己录制的语音,录音人数为 10 人,其中 6 男 4 女。录音内容为贵州大学校训“明德至善,博学笃行”,长度为 7 s。将标准的和改进的 MFCC 参数提取方法在 MATLAB2010b 环境下进行仿真设计。图 4 为非线性随机共振系统的传输函数的特性。

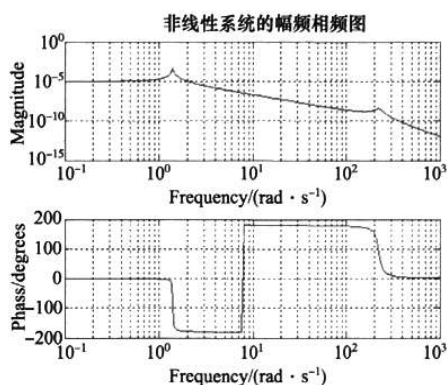


图 4 非线性随机共振系统的幅频相频特性

以 1 号女生的语音信号作为测试语音,其他语音信号处理方法同 1 号。其测试结果如图 5,图 6,图 7,图 8。

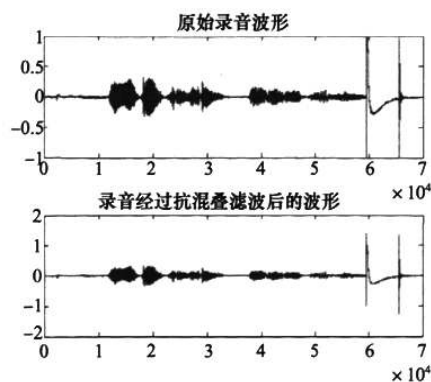


图 5 抗混叠滤波前后的波形

从图 7 和图 8 可以看出,改进的 MFCC 参数输出线条更加清晰,更容易读取其特性。

同时,实验还对相同说话人和不同说话人进行

仿真对比,其结果如图 9 和图 10。

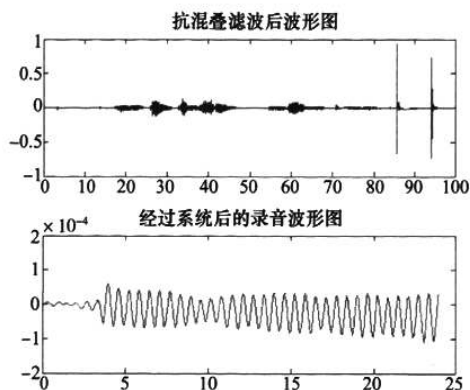


图 6 经过随机共振系统前后的波形

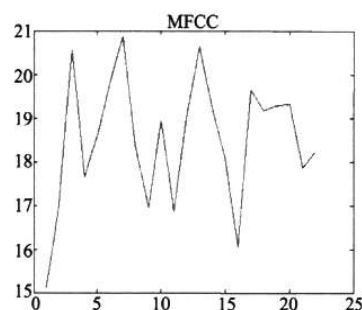


图 7 改进的 MFCC 输出

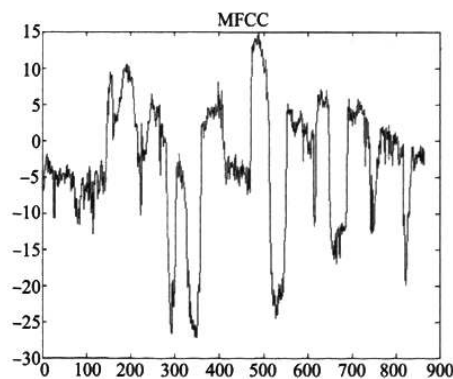


图 8 标准的 MFCC 输出

分别将标准的 MFCC 参数和改进的 MFCC 参数作为特征矢量通过识别系统,通过实验得到两种参数下系统的识别率如表 1 所示。

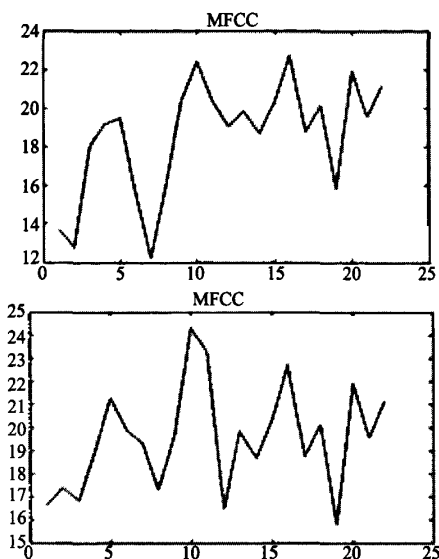


图9 同一说话人,同一语言 MFCC 输出

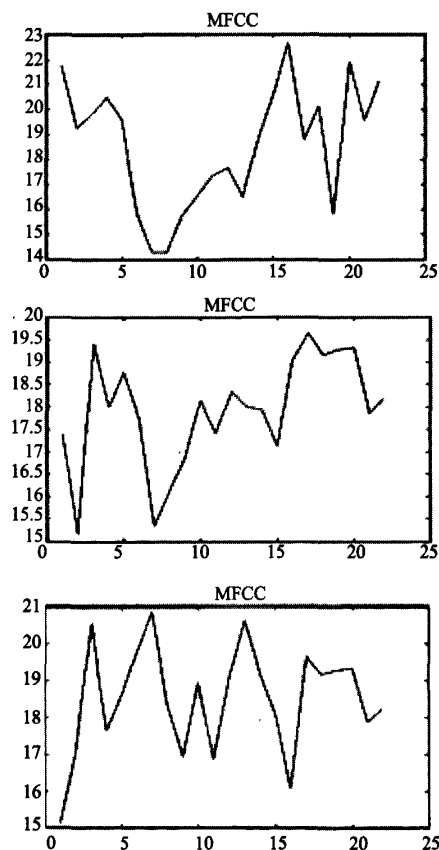


图10 不同说话人,同一语言 MFCC 输出

表1 两种方法的识别准确率比较

待识别语音	标准 MFCC	改进 MFCC
2 s	82	86
3 s	86	90
4 s	90	94

实验结果表明,改进的 MFCC 参数提取方法所得到的特征矢量提高了系统的识别率,说明基于随机共振的 MFCC 参数提取方法可行且非常有效。

4 总结和展望

本文介绍了说话人识别中 MFCC 参数的提取方法及其基于随机共振的改进方法,并对两种方法进行了比较,得出改进的 MFCC 参数提取方法在说话人识别中起到了更好的识别效果。在以后的实验和仿真过程中,将不断改进提取过程中的算法,可以进一步提高说话人识别的识别率。

参考文献

- 1 Kumar A, Mullick S K. Nonlinear dynamical analysis of speech. J Acoust Soc Am, 1996: 615—628
- 2 潘平,杨平,何朝霞. 基于随机共振的说话人识别方法. 电信科学, 2010; (52):
- 3 聂春燕. 混沌系统与弱信号检测. 北京: 清华大学出版社, 2009
- 4 王俊国,周建中,付波,等. 基于 Duffing 振子的微弱信号混沌检测. 电子器件, 2007: 1380—1383
- 5 韩纪庆,张磊,郑浩然. 语音信号处理. 北京: 清华大学出版社, 2008: 84—85
- 6 王炳锡,屈丹,彭煌. 实用语音识别基础. 北京: 国防工业出版社, 2004: 148—149
- 7 陈伟兵,周凌宏,肖中举. 耳蜗基底膜振动模型的建立与应用. 中国医学物理杂志, 2007: 221—223
- 8 陈晓霞,王辅忠. 利用参数调节随机共振检测大参数信号. 天津工业大学学报, 2008: 60—64

(下转第 4227 页)

Research of Data Integration Based on Data Elements

SHI Gui-ying, WEN Bi-long, WANG Zhi-bao

(School of Computer&Information Technology, Northeast Petroleum University, Daqing 163318, P. R. China)

[**Abstract**] For the shortcomings of the attribute-level mapping in the current data integration approach, a new data integration approach based on data elements is presented, which uses semantic tree of data elements to describe the connotation of the properties's concept, and the semantic of the properties can be understood by computer, and then the instance-level semantic mapping can be realized by semantic computation, so that the data integration can be completed.

[**Key words**] data integration data elements semantic tree semantic mapping



(上接第 4218 页)

An Improved Method of MFCC Parameter Extraction in Speaker Recognition

HE Zhao-xia, PAN Ping

(College of Computer Science & Information, Guizhou University, Guiyang 550023, P. R. China)

[**Abstract**] Speech feature parameter extraction is an very important part of the speech recognition system, especially in speech training and recognition. Mel frequency cepstrum coefficient (MFCC) is a common feature, It can analysis and process speech signal, remove redundant information in speech recognition, and gain important information which influence speech recognition. Owing to time-varying and chaotic characteristic of voice signal, a improved MFCC feature extraction method based on nonlinear stochastic resonance theory is proposed. By comparison results of two methods, it is proved that the improved one is practicable and more superior which provides a new direction of speech feature parameter extraction in speech recognition.

[**Key words**] speech recognition feature extraction MFCC stochastic resonance