

基于耳蜗基底膜特性的语音特征提取方法 及在智能轮椅上的应用^{*}

罗元, 陈君, 张毅, 童开国

(重庆邮电大学智能系统及机器人研究所, 重庆400065)

摘要: 从人耳听觉特性出发, 对能模拟耳蜗基底膜滤波特性的伽马通滤波器组进行了研究、修正, 并以修正的滤波器组为基础, 提出了一种基于耳蜗基底膜特性的 GT-ERBCC (equivalent rectangular bandwidth cepstrum coefficient based on GammaTone filters) 语音特征提取方法。该方法能准确地表征出语音信号的特征, 降低语音识别系统的难度, 并将该方法应用于智能轮椅人机交互实验中。结果表明, 基于耳蜗基底膜特性的特征提取方法能有效提高语音识别系统的识别率。

关键词: 伽马通; 耳蜗基底膜特性; 特征提取; 智能轮椅; 语音识别

中图分类号: TP311 **文献标志码:** A **文章编号:** 1001-3695(2012)10-3765-04

doi:10.3969/j.issn.1001-3695.2012.10.041

Speech feature extraction method based on cochlear basilar membrane properties and its application on intelligent wheelchair

LUO Yuan, CHEN Jun, ZHANG Yi, TONG Kai-guo

(Research Center of Intelligent System & Robot, Chongqing University of Posts & Telecommunications, Chongqing 400065, China)

Abstract: This paper studied and modified the GammaTone filters, which could simulate the filter characters of cochlear basilar membrane based on auditory system. It proposed a GT-ERBCC (equivalent rectangular bandwidth cepstrum coefficient based on GammaTone filters) feature extraction method based on the novel GammaTone-4 filters. Such method can precisely show the characters of speech signals, reduce the difficulty of speech recognition system. Furthermore, this method was applied on human-computer interaction of the intelligent wheelchair. Experiment results show that it can reflect the characters of speech signals precisely and increase the speed recognition rate effectively.

Key words: GammaTone; cochlear basilar membrane characters; feature extraction; intelligent wheelchair; speech recognition

0 引言

人类听觉系统在噪声环境下能够很好地工作, 所以如果语音识别系统能模拟人类听觉感知的处理特点, 噪声环境下识别率就会有所提高。近年来, 基于听觉模型的语音特征提取方法在语音识别领域日益受到重视^[1,2], 就是因为听觉模型最接近人耳对声音信号的处理过程, 提取的特征能反映声音的本质, 具有很好的鲁棒性^[3]。

GT-ERBCC (equivalent rectangular bandwidth cepstrum coefficient based on GammaTone filters) 就是基于人类听觉特性的一种特征提取方法。该方法对 GammaTone 滤波器进行了研究与改进, 使得改进后的滤波器 (GammaTone-4) 具有更好的人耳听觉特性, 能有效提高语音识别系统的识别率。同时, 也改善了传统的 MFCC 特征提取方法中采用三角滤波器组带来的相邻频带之间频谱能量相互泄漏严重不利于反映共振特性的问题, 为整个语音识别系统的优化提供了

基础。

1 耳蜗基底膜特性

耳蜗是人类听觉系统的重要组成部分, 对声音的处理类似于一族带通滤波器对信号的滤波。基底膜是一个贯穿耳蜗底部自顶部的膜结构, 耳蜗基底膜具有很重要的频率选择特性和频谱分析特性^[3,4]。

当声音传入中耳时, 镫骨的运动引起耳蜗内流体压强的变化, 从而引起行波沿基底膜的传播, 图1为流体波的简单表示。在耳蜗底部基底膜的硬度很高, 流体波传播得很快; 随着波的传播, 基底膜的硬度变得越来越小。而这种耳蜗基底膜硬度的变化决定了它在不同位置有不同的特征频率。图2为基底膜特征频率与基底膜位置的关系曲线, 由图可以看出, 基底膜特征频率变化是非线性的, 且基底膜底部的特征频率最高, 顶部的特征频率最低, 并且特征频率与基底膜位置呈指数关系。

收稿日期: 2012-03-19; **修回日期:** 2012-04-27 **基金项目:** 国家科技部国际合作项目(2010DF12160); 国家自然科学基金资助项目(51075420); 重庆市攻关项目(CSTC, 2010AA2055)

作者简介: 罗元(1972-), 女, 湖北人, 教授, 博士, 主要研究方向为机器视觉、智能信号处理、数字图像处理(luoyuan@cqupt.edu.cn); 陈君(1987-), 女, 四川遂宁人, 硕士研究生, 主要研究方向为语音定位与识别; 张毅(1966-), 男, 重庆人, 教授, 博导, 博士(后), 主要研究方向为机器人及应用、数据融合、信息无障碍技术; 童开国(1986-), 男, 湖北人, 硕士研究生, 主要研究方向为语音识别。

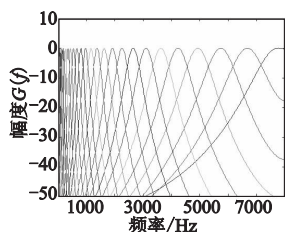


图6 GammaTone-4滤波器组幅频特性曲线

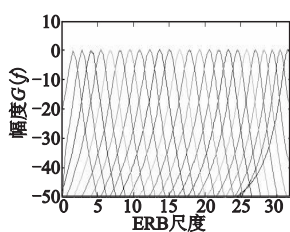


图7 GammaTone-4滤波器组ERB尺度特性曲线

3 基于 GammaTone-4 的语音特征提取

本文以改进的 GammaTone-4 滤波器组为基础,提出了一种基于耳蜗基底膜特性的语音特征提取方法,即 GT-ERBCC 特征提取方法。该方法从耳蜗基底膜的频率选择特性和频谱分析特性入手,体现了人类听觉感知的处理特点。GT-ERBCC 特征提取和计算过程如图 8 所示。

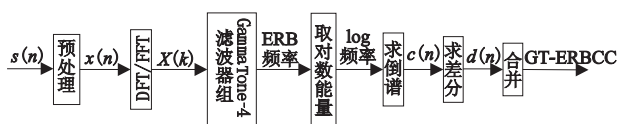


图8 GT-ERBCC特征提取过程

a) 原始信号 $s(n)$ 经过预加重、分帧、加窗等预处理后,得到每个语音帧的时域信号 $x(n)$ 。

b) 将信号进行离散傅里叶变换 (DCT) 得到其频谱 $X(k)$, 转换公式为

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N} \quad (2)$$

其中: $0 \leq n, k \leq N-1$ 。

在实际应用中,常常用快速傅里叶变换 (FFT) 以减少计算时长。

c) 求能量谱。将线性频谱 $X(k)$ 通过 GammaTone-4 滤波器组得到 ERB 频谱。其中 GammaTone-4 滤波器组为在语言的频谱范围内设置的若干个带通滤波器 $H_m(k)$, $0 \leq m < M$, M 为滤波器个数,通常取 $24 \sim 40$ ^[5],本文取 24。每个滤波器都具有人耳基底膜的滤波特性, GammaTone-4 滤波器组在线性频率上,当 m 较小时,相邻中心频率的带宽很小,随着 m 的增大,相邻中心频率的带宽逐渐增大,这与人耳对不同频率的语音具有不同的感知能力是一致的。MFCC 中使用的三角滤波器是在 Mel 刻度上均匀分布的,这不符合听觉特性中临界带宽的概念,而 GammaTone-4 是在 ERB 尺度上均匀分布的,ERB 与临界带宽紧密相关,但不是基于传统掩蔽实验得到的,所以不容易被信号和掩蔽信号之间的互调所影响,且比 bark 尺度表现得更平滑。

d) 对滤波器组输出的 Mel 频谱取对数能量,得到对数频谱 $S(m)$,由线性频谱 $X(k)$ 到对数频谱 $S(m)$ 的总传递函数为

$$S(m) = \ln \left(\sum_{k=0}^{N-1} |X(k)|^2 H_m(k) \right) \quad (3)$$

其中: $0 \leq m < M$ 。

e) 将对数频谱 $S(m)$ 经过离散余弦变换转换到倒频谱域获得 GT-ERBCC 系数 $c(n)$ 。转换公式如下:

$$c(n) = \sum_{m=1}^{M-1} S(m) \cos \left[\frac{\pi(m+0.5)n}{M} \right] \quad (4)$$

其中: $0 \leq m < M$ 。

f) GT-ERBCC 参数值反映了语音参数的静态特征,而人耳

对语音的动态特征更敏感,通常用差分倒谱参数来描述这种动态特征,计算公式如下:

$$d(n) = \frac{1}{\sqrt{\sum_{i=-k}^k i^2}} \sum_{i=-k}^k i \times c(n+i) \quad (5)$$

其中: c 和 d 都表示一帧语音参数, k 为常数,通常取 2,这时差分参数就称为当前帧的前两帧和后两帧参数的线性组合。

g) 合并。将 GT-ERBCC 参数和一阶差分参数合并后一起构成语音的特征参数。

4 实验与结果

4.1 实验设计

实验所用的数据库为 Cooledit 软件采集的相对纯净的语音,并用谱减法进行降噪后混入不同种类和强度噪声的一个汉语单词语音库。该数据库由 20 人 (10 男 10 女) 录制,每个录音人对汉语单词重复录制 10 次 (从 9:00 到 18:00 每小时录一遍),考虑了时间的遍历性。录音设置采样率为 16 kHz,采样精度为 16 位,单声道,Windows PCM 编码,语音文件为标准 Windows 的 WAV 音频格式。噪声来自于 Noise-92 库中的 destroyerengine (简称为 DE)、pink、volvo 和 white 噪声,按照随机的时间偏移混入干净语音,形成 $0 \sim 30$ dB 不同信噪比的数据库。

本文用交叠分段的方法对输入的语音信号进行分帧,每 70 个采样点分为一帧,叠接步长为 35 个采样点,利用语音的短时平稳特性,对信号进行特征提取。

4.2 实验结果与分析

将本文的特征与当前应用最多的 MFCC 特征在 DTW 识别模式下进行语音识别实验对比,结果如图 9、10 所示。由图可以看出,噪声较小时, MFCC 特征与 GT-ERBCC 特征的识别率基本相同,但随着噪声的增加, GT-ERBCC 特征的语音识别率明显提高。

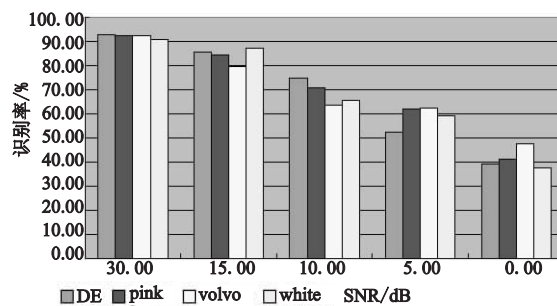


图9 MFCC的语音识别结果

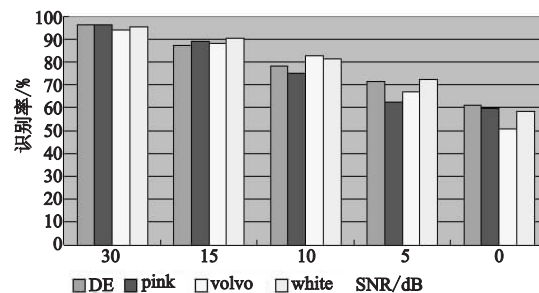


图10 GT-ERBCC的语音识别结果

4.3 GT-ERBCC 在智能轮椅人机交互上的应用

在实验室环境下,由 20 人 (10 男 10 女) 分别在特征参数

为 MFCC 与 GT-ERBCC 的语音智能轮椅人机交互系统中说前进、后退、左转、右转、停止五个命令,控制轮椅行走固定的“8”字路线。

输入语音格式设置如下:

```
m_wfx.wFormatTag = WAVE_FORMAT_PCM; // 录音格式
m_wfx.nChannels = 1; // 声道数
m_wfx.nSamplesPerSec = 16000; // 采样率
m_wfx.wBitsPerSample = 16; // 采样位
m_wfx.nAvgBytesPerSec = 16000 * 2;
// 数据率,即每秒采样的字节数
m_wfx.nBlockAlign = 2; // 块对齐大小
m_wfx.cbSize = 0;
```

实验证明,智能轮椅在各个命令下能成功执行相应的行动,完成整个路线的行走。

5 结束语

本文结合耳蜗基底膜的特性,提出了一种符合人耳听觉特性的语音特征提取方法 GT-ERBCC,并在智能轮椅人机交互系统上实现了语音控制。实验证明,在一般噪声环境下,所提算法能准确表征出语音段信号,有效提高语音识别率,为信息无障碍做出了贡献。

将来随着研究的深入,基于耳蜗基底膜特性的语音特征提取算法有很广泛的应用前景。它不仅可以应用于机器人的语音识别方面,还可以提高语音系统识别率;它不仅可以应用于声音信号的检索和辅助目前的文字检索,还可以应用于船舶的辐射噪声测量等。

参考文献:

- [1] CHATTERJEE S, KLEIJN W B. Auditory model-based design and optimization of feature vectors for automatic speech recognition[J]. IEEE Trans on Audio, Speech, and Language Processing, 2011, 19(6): 1813-1825.
- [2] PENG Yuan, MU Lin, KONG Xiang-li, *et al.* A study on echo feature extraction based on the modified relative spectra (RASTA) and perception linear prediction (PLP) auditory model[C]//Proc of IEEE International Conference on Intelligent Computing and Intelligent Systems. 2010: 29-31.
- [3] 陈世雄, 宫琴, 金慧君. 用 GammaTone 滤波器组仿真人耳基底膜的特性[J]. 清华大学学报: 自然科学版, 2008, 48(6): 1044-1048.
- [4] 张雪英. 数字语言处理及 MATLAB 仿真[M]. 北京: 电子工业出版社, 2010: 190-195.
- [5] LYON R F, MEAD C. An analog electronic cochlea[J]. IEEE Trans on Acoustics, Speech, and Signal Processing, 1988, 36(7): 1119-1134.
- [6] LYON R. A computational model of filtering, detection, and compression in the Cochlea[C]//Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. 1982: 1282-1285.
- [7] PATTERSON R D, MOORE B C J. Auditory filters and excitation patterns as representation of frequency resolution[C]//Frequency Selectivity in Hearing. London: Academic Press, 1986: 123-177.
- [8] JOHANNESMA P I M. The pre-response stimulus ensemble of neurons in the cochlear nucleus[C]//Proc of IPO Symposium on Hearing Theory. 1972: 58-69.