

基于共振峰曲线的语音信号动态特征提取方法

韩志艳,王 健

(渤海大学 工学院,辽宁 锦州 121000)

摘 要:为了提高噪声环境下语音识别的鲁棒性,提出了一种基于共振峰曲线的语音信号动态特征提取方法。采用基于 Hilbert-Huang 变换的方法来估算预处理后的语音信号共振峰频率特征,然后按照从第一帧到最后一帧的帧序,将预处理后的每帧语音信号的第一共振峰频率特征值进行组合获得第一共振峰曲线,依此类推,获得第二共振峰曲线、第三共振峰曲线及第四共振峰曲线。对获得的每条共振峰曲线进行快速傅里叶变换获得线性频谱,然后再求取能量谱,计算对数能量和离散余弦变换。与 MFCC 方法相比,提取的语音信号动态特征具有时间相关性,揭示了语音信号前后以及相邻之间存在的密切关联,提高了语音识别的性能。

关键词:语音信号;动态特征;语音识别;特征提取;共振峰曲线

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2017)06-0072-04

doi:10.3969/j.issn.1673-629X.2017.06.015

Dynamic Feature Extraction for Speech Signal Based on Formant Curve

HAN Zhi-yan, WANG Jian

(College of Engineering, Bohai University, Jinzhou 121000, China)

Abstract: In order to improve the robustness of speech recognition in noise environment, a dynamic feature extraction for speech signal based on formant curve is put forward. It uses Hilbert-Huang transform to estimate speech signal formant frequency characteristics after preprocessing, and then gets the first formant curve by combining the first formant frequency characteristics of each frame from the first frame to the last frame, and so forth, gets the second, the third and the fourth formant curve. And then takes Fast Fourier Transform for each formant curve to obtain linear spectrum, and calculates the energy spectrum, logarithmic energy and discrete cosine transform. Compared with the method of MFCC, the proposed dynamic feature of speech signal has the time correlation, revealing the close correlation between the speech signal frames, improving the performance of speech recognition.

Key words: speech signal; dynamic feature; speech recognition; feature extraction; formant curve

0 引 言

语音识别最基础最重要的开发环节是语音信号特征参数的提取。语音信号特征参数提取,即利用数学理论提取语音信号中所携带的有用信息,获得一个矢量序列。R. K. Potter 等^[1]早在二十世纪四十年代就提出了“可视语音”的概念,指出语谱图对语音信号有很强的描述能力,而且用语谱图进行了语音识别,即形成了最早的语音特征。到了五十年代,人们发现要想减少模板数目、运算量、存储量及提高识别率,就必须提取语音信号中能够反映语音特性的某些参数,滤除语音信号中的冗余信息,于是就出现了幅度特征、短时帧平均能量特征、短时帧过零率特征、短时自相关系数特征、平均幅度差函数特征等。但随着语音识别技术的

发展,发现无论从稳定性还是区分能力,上述时域特征参数的表现都不是很好,于是开始利用频域特征参数进行识别,比如基音周期^[2]、共振峰频率特征^[3]、线性预测系数(LPC)特征^[4]、线谱对(LSP)特征^[5-6]、倒谱系数特征等^[7]。目前基于全声道全极点模型的线性预测倒谱系数(LPCC)^[8-10]和基于人耳听觉模型的梅尔倒谱系数(MFCC)^[11-14]应用最为广泛。

但上面所述的特征参数反映的都是语音信号的静态特征,要使提取出的特征参数能更好地表达语音信号,就必须提取动态特征参数,语音信号的动态特性即为从连续几帧语音信号中提取的特征参数。动态特性是语音多样性的一部分,它不同于平稳的随机过程,具有时间相关性,比如可以通过静态特征的差分参数和

收稿日期:2016-07-29

修回日期:2016-11-03

网络出版时间:2017-04-28

基金项目:国家自然科学基金资助项目(61403042,61503038);辽宁省教育科研项目(L2013423)

作者简介:韩志艳(1982-),女,博士,副教授,研究方向为语音识别、情感识别。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20170428.1704.084.html>

加速度参数来获取。但它们并不能将动态信息挖掘得很充分,所以尚不能很好地反映语音信号的动态特性。

因此,提出了一种基于共振峰曲线的语音信号动态特征提取方法,构成的共振峰曲线具有时间相关性,揭示了语音信号前后以及相邻之间存在的密切关联。其中采用基于 Hilbert-Huang 变换方法来估算预处理后的语音信号共振峰频率特征,其中用经验模态分解法(EMD)将信号分解成一组含有不同尺度的固有模态函数(IMF)分量,经分解得到的每一个 IMF 分量都代表了一个频率成分,这些频率成分可以有效突出信号的局部特性和细节变化,有助于快速有效地掌握信号的动态特征。

因此,语音特征的动态变化,可以通过动态特性来描述,而研究语音信号的动态特性,也是匹配新的语音动态模型、提高语音辅助工程性能的必然趋势。

1 共振峰特征提取

在语音识别技术应用领域,共振峰特征参数是重要的声学特征参数之一。长期以来该参数的提取都是基于人的发声系统是线性的和语音信号是短时平稳的两个基本假设。随着对语音发声机理的深入研究,发现在语音产生过程中存在着非线性,因此传统的线性共振峰特征参数估计方法的准确性就会受到影响^[15]。另一方面,由于传统分析方法建立在短时平稳的假设上,对快速变化的共振峰特征参数的提取无能为力。所以研究者们越来越重视对随时间快速变化的动态信息的提取。

近年来,尽管也提出了一些新的参数提取方法,如逆滤波器法^[16]和频域线性预测算法等^[17],但这些方法都只是在算法和处理方法上进行改进,本质上仍属于线性分析方法的范畴,而且分析计算过程复杂,需要根据主观经验来调整参数。文中采用一种基于 Hilbert-Huang 变换(Hilbert-Huang Transform, HHT)的适用于非平稳、非线性信号处理,具有自适应特性的时间-频率分析新方法。

HHT 包括 2 个基本步骤:第一步是经验模态分解(Empirical Mode Decomposition, EMD),它的核心是“筛选”,即从被分析信号中提取一族固有模态函数(Intrinsic Mode Function, IMF);第二步是计算信号的 Hilbert 谱(Hilbert Spectrum),将每个 IMF 与它的 Hilbert 变换构成一个复解析函数,并由此导出作为时域函数的瞬时幅值(能量)和瞬时频率。

通过 EMD 得到的每个 IMF 满足两个条件:

(1)在整个序列上,极值点个数和过零点个数相等或至多相差一个;

(2)分别构造其各局部极大值和局部极小值所形

成的上、下 2 条包络线的均值在任一点处为零。

分解后得到信号 $x(t)$ 的 n 个 IMF 分量 $c_1(t)$, $c_2(t)$, \dots , $c_n(t)$ 和剩余项 $r_n(t)$, 即有:

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (1)$$

对每个 $c_i(t)$, $i = 1, 2, \dots, n$, 求其 Hilbert 变换 $d_i(t)$, 然后计算相应的瞬时频率 $\omega_i(t)$ 和幅值 $a_i(t)$:

$$\omega_i(t) = d\theta_i(t)/dt \quad (2)$$

$$a_i(t) = [c_i^2(t) + d_i^2(t)]^{1/2} \quad (3)$$

其中, $\theta_i(t)$ 为瞬时相位。

$$\theta_i(t) = \arctan[d_i(t)/c_i(t)] \quad (4)$$

根据每个 IMF 的瞬时频率和幅值,可将信号表示为:

$$x(t) = \sum_{i=1}^n a_i(t) \exp[j\int \omega_i(t) dt] \quad (5)$$

由于 $r_n(t)$ 不是一个常数就是一个单调函数,对信号分析和信息提取没有实质性的影响,所以式(5)中略去了式(1)中的剩余项。在时间-频率面上画出每个 IMF 以其幅值加权的瞬时频率曲线,这个时间-频率分布谱图就是 Hilbert 谱,记为 $H(\omega, t)$ 。

当采用 HHT 方法估计语音信号的共振峰频率时,为了避免和抑制各个共振峰分量在 EMD 过程中产生互相干扰,需要事先对各个共振峰分量进行分离,对分离后的各个共振峰分量作 EMD,最后求出相应的共振峰频率及其随时间的变化曲线。

2 动态特征提取

动态特征提取流程如图 1 所示。

其具体步骤如下:

步骤 1:利用麦克风输入语音数据,然后以 11.025 kHz 的采样频率、16 bit 的量化精度进行采样量化,获得相应的语音信号。然后利用一阶数字预加重滤波器对获取的语音信号进行预加重处理,其中预加重滤波器的系数取值范围为 0.93 ~ 0.97。接下来以帧长 256 点的标准进行分帧处理,并对分帧后的语音信号加汉明窗,再利用短时能零积法进行端点检测。短时能零积方法如下:

短时能量与相应的短时过零率之积称为短时能零积,每一帧的短时能量 E_n 和短时过零率 Z_n 以及短时能零积 EZ_n 的定义分别为:

$$E_n = \sum_{k=0}^{N-1} s_w^2(k) \quad (6)$$

$$Z_n = \sum_{k=1}^N |\operatorname{sgn}[s_w(k)] - \operatorname{sgn}[s_w(k-1)]| \quad (7)$$

$$EZ_n = E_n * Z_n \quad (8)$$

其中, n 为语音信号的第 n 帧; N 为每一帧的长

度; $s_w(k)$ 为加窗语音信号。

用短时能零积法进行语音端点检测的步骤如下:

(1) 确定噪声的门限阈值。

无音片段主要包括的是背景噪声, 由于录音开始阶段往往有一段无音区, 所以在实验室环境下通常取最开始的 5 帧信号作为背景噪声的分析, 对这 5 帧信号按式(6)和式(7)分别按帧计算 E_n 和 Z_n , 并按式(8)计算 EZ_n , 通过多帧平均, 就得到了平均短时能零积 EZ , 并按照式(9)确定噪声的门限阈值 TH 。

$$TH = k \times EZ \quad (9)$$

其中, k 为经验值, 通常取 1.2。

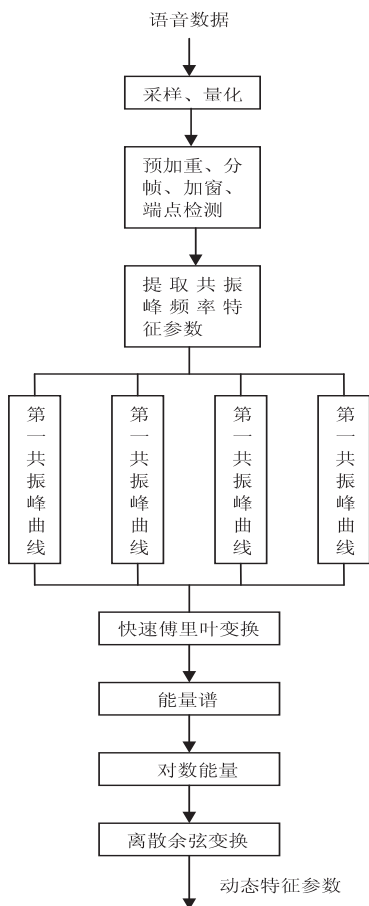


图 1 动态特征提取流程图

(2) 利用短时能零积进行语音端点检测。

计算每帧录音信号的短时能零积 EZ_n , 与噪声的门限阈值 TH 做比较。 EZ_n 大于 TH , 就以该帧的帧号作为有音片段的起点 N_1 , 表明进入了有音片段。如果由过去帧已经得到了 N_1 , 那么当 EZ_n 小于 TH 时, 就以该帧的帧号作为有音片段的终点 N_1 。相反, 如果 N_1 还未得到, 那么当 EZ_n 小于 TH 时, 表明当前帧仍处于无音片段。

步骤 2: 计算共振峰频率特征参数, 其中获得的每帧语音信号的第一共振峰特征值为 F_1 、第二共振峰特征值为 F_2 、第三共振峰特征值为 F_3 和第四共振峰特征

值为 F_4 。

步骤 3: 构成共振峰曲线。具体为:

(1) 按照从第一帧到最后一帧的帧序, 将预处理后的每帧语音信号的第一共振峰频率特征值 F_1 进行组合, 获得第一共振峰曲线 $x_1(n)$, $n=0, 1, \dots, N-1$, N 为语音信号的帧数;

(2) 按照从第一帧到最后一帧的帧序, 将预处理后的每帧语音信号的第二共振峰频率特征值 F_2 进行组合, 获得第二共振峰曲线 $x_2(n)$;

(3) 按照从第一帧到最后一帧的帧序, 将预处理后的每帧语音信号的第三共振峰频率特征值 F_3 进行组合, 获得第三共振峰曲线 $x_3(n)$;

(4) 按照从第一帧到最后一帧的帧序, 将预处理后的每帧语音信号的第四共振峰频率特征值 F_4 进行组合, 获得第四共振峰曲线 $x_4(n)$ 。

步骤 4: 对获得的第一、第二、第三和第四共振峰曲线进行快速傅里叶变换, 获得每条共振峰曲线的线性频谱。

$$X_i(k) = \sum_{n=0}^{N-1} x_i(n) e^{-j2\pi nk/N} \quad (10)$$

其中, $X_i(k)$ 表示第 i 条共振峰曲线进行快速傅里叶变换后得到的线性频谱, $i=1, 2, 3, 4$, $k=0, 1, \dots, N-1$, N 为语音信号的帧数; $x_i(n)$ 表示第 i 条共振峰曲线。

步骤 5: 根据线性频谱获得每条共振峰曲线的能量谱。即取上述线性频谱 $X_i(k)$ 模的平方来获得相应的能量谱 $S_i(k)$:

$$S_i(k) = |X_i(k)|^2 \quad (11)$$

步骤 6: 根据能量谱获得每条共振峰曲线的对数能量。即为了使结果对噪声有更好的鲁棒性, 将获得的能量谱 $S_i(k)$ 取对数, 即可获得对数能量 $L_i(k)$:

$$L_i(k) = \log(S_i(k)) \quad (12)$$

步骤 7: 对上述对数能量进行离散余弦变换, 获得倒频谱域, 即获得语音信号动态特征参数:

$$C_i(t) = \sum_{k=0}^{N-1} L_i(k) \cos\left[\frac{\pi t(k+0.5)}{N}\right] \quad (13)$$

其中, $C_i(t)$ 表示第 i 条共振峰曲线的动态特征参数, $i=1, 2, 3, 4$; $t=1, 2, \dots, T$, T 表示设定的倒谱系数个数, 取值范围为 12~16。

3 仿真实验及结果分析

采用 50 个典型的汉语词汇进行实验。由于考虑识别系统容易受环境噪声、信道变化和说话人变化等因素的影响, 因此, 训练集采用安静环境下的语音数据, 而测试集采用含有噪声的数据。

为了验证该特征参数对不同说话人变化的鲁棒

性,训练集数据由前后两次录成,共 50 人,每人每词发音一遍,共获得 5 000 个数据,测试集数据也是分两次录成,共 30 人,每人每词发音一遍,共 3 000 个数据;为了验证该特征参数对不同信道变化的鲁棒性,每次使用不同的麦克风来录音;为了验证该特征参数对不同环境噪声变化的鲁棒性,在测试集的每个语音中手工加入四种噪声,包括:白噪声、粉噪声、街道噪声、坦克噪声,构成信噪比为 15 dB,10 dB,5 dB,0 dB,-5 dB 的含噪语音信号。采用基于遗传算法改进的小波神经网络作为分类器^[18-19]。图 2~5 为采用与文中算法相同条件的 MFCC 方法和文中方法分别在白噪声、粉噪声、街道噪声和坦克噪声干扰下的系统识别性能曲线。

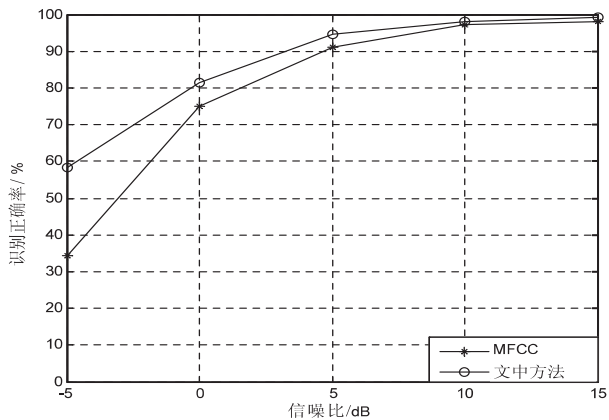


图2 白噪声环境下的系统识别性能曲线

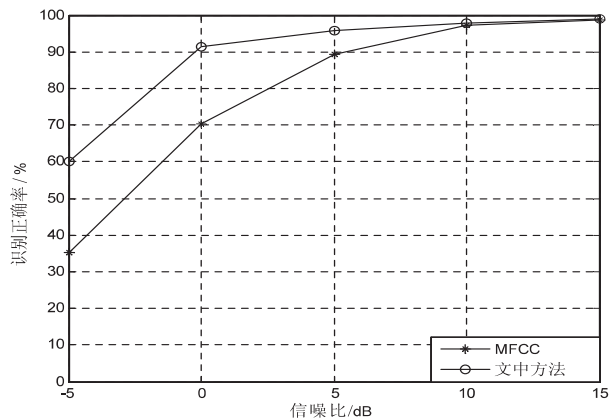


图3 粉噪声环境下的系统识别性能曲线

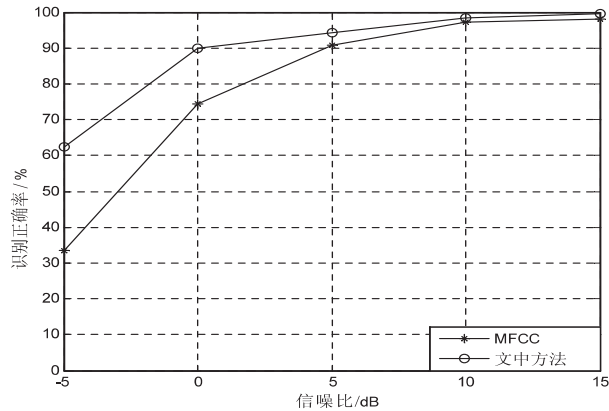


图4 街道噪声环境下的系统识别性能曲线

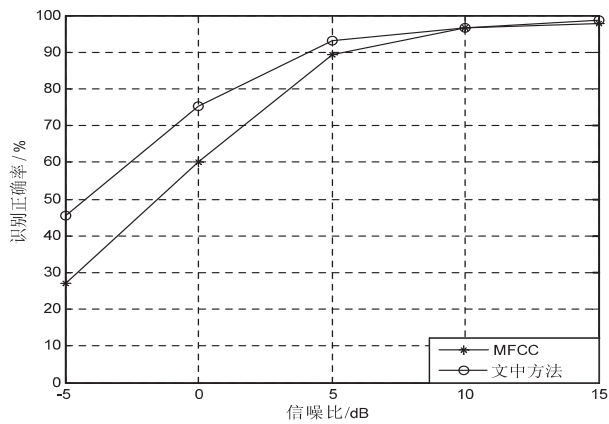


图5 坦克噪声环境下的系统识别性能曲线

从图中可以看出,在信噪比较低时,文中方法与 MFCC 方法相比识别率提高了很多。这是因为文中方法构成的共振峰曲线具有时间相关性,揭示了语音信号前后以及相邻之间存在着密切关联,这一特性,使得在强噪声环境下应用语音识别技术成为了可能。

4 结束语

文中提取的语音信号动态特征,采用基于 Hilbert-Huang 变换的方法来估算预处理后的语音信号共振峰频率特征,其中用 EMD 将信号分解成一组含有不同尺度的 IMF 分量,经分解得到的每一个 IMF 分量都代表了一个频率成分,这些频率成分可以有效突出信号的局部特性和细节变化,有助于快速有效地掌握信号的动态特征。相比于传统的 MFCC 方法,大大提高了语音识别的性能。但是语音信号的某一特征中一般只包含部分语音信息,所以采用动静态特征参数的组合,这样动态信息和静态信息形成了互补,当各组合参数间相关性不大时,会有很好的效果。

参考文献:

[1] Potter R K, Kopp G A, Green H C. Visible speech[M]. New York: Van Nostrand, 1947.

[2] 赵瑞珍,宋国乡. 基音检测的小波快速算法[J]. 电子科技, 1998, 43(1): 16-19.

[3] 黄 海,陈祥献. 基于 Hilbert-Huang 变换的语音信号共振峰频率估计[J]. 浙江大学学报:工学版, 2006, 40(11): 1926-1930.

[4] Christensen R L, Sreong W J, Palmer E P. A comparison of three methods of extracting resonance information from predictor coefficient coded speech[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1976, 24(1): 8-14.

[5] Girin L. Joint matrix quantization of face parameters and LPC coefficients for low bit rate audiovisual speech[J]. IEEE Transactions on Speech and Audio Processing, 2004, 12(3): 265-276.

根据组合学原理,利用 MapReduce 扫描一次数据库从原始事务数据库中完成频繁项集的整体挖掘过程;且在支持度阈值改变的情况下无需重新扫描数据库进行挖掘,提高了频繁项集的挖掘效率。实验结果表明,该算法不受支持度阈值的影响,且对于大量事务数据,运行效率高,适用于智能调度大数据的关联分析。大数据在智能调度中的应用价值不可估量,但是,要加速智能调度化的进程,需要在多源数据融合和全景数据深度分析方面有所突破。

参考文献:

- [1] 刘振亚. 智能电网技术[M]. 北京:中国电力出版社,2010.
- [2] 辛耀中,石俊杰,周京阳,等. 智能电网调度控制系统现状与技术展望[J]. 电力系统自动化,2015,39(1):2-8.
- [3] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases[C]//Proceedings of the ACM SIGMOD conference on management of data. Washington, D C:ACM,1993:207-216.
- [4] Han Jiawei,Pei Jian,Yin Yiwen. Mining frequent patterns without candidate generation [C]//Proceedings of the ACM SIGMOD conference on management of data. Dallas, TX:ACM,2000:1-12.
- [5] Baralis E, Cerquitelli T, Chiusano S, et al. Scalable out-of-core itemset mining[J]. Information Sciences,2015,293(4):146-162.
- [6] Baralis E, Cerquitelli T, Chiusano S. A persistent HY-Tree to efficiently support itemset mining on large datasets[C]//Proceedings of the 2010 ACM symposium on applied computing. New York:ACM,2010:1060-1064.
- [7] Adnan M, Alhajj R. DRFP-tree: disk-resident frequent pattern tree[J]. Applied Intelligence,2009,30(2):84-97.
- [8] Buehrer G, Parthasarathy S, Ghoting A. Out-of-core frequent pattern mining on a commodity PC [C]//Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. New York:ACM,2006:86-95.
- [9] 张东霞,苗新,刘丽平,等. 智能电网大数据技术发展研究[J]. 中国电机工程学报,2015,35(1):2-12.
- [10] 宋亚奇,周国亮,朱永利. 智能电网大数据处理技术现状与挑战[J]. 电网技术,2013,37(4):927-935.
- [11] 彭小圣,邓迪元,程时杰,等. 面向智能电网应用的电力大数据关键技术[J]. 中国电机工程学报,2015,35(3):503-511.
- [12] 李建江,崔健,王聃,等. MapReduce 并行编程模型研究综述[J]. 电子学报,2011,39(11):2635-2642.
- [13] 孟小峰,慈祥. 大数据管理:概念、技术与挑战[J]. 计算机研究与发展,2013,50(1):146-169.
- [14] 吴凯峰,刘万涛,李彦虎,等. 基于云计算的电力大数据分析技术与应用[J]. 中国电力,2015(2):111-116.
- [15] Trentin E, Gori M. Robust combination of neural networks and hidden Markov models for speech recognition[J]. IEEE Transactions on Neural Networks,2003,14(6):1519-1531.
- [16] Hong K K, Rose R C. Cepstrum-domain model combination based on decomposition of speech and noise for noisy speech recognition[C]//IEEE international conference on acoustics, speech, and signal processing. [s. l.]:IEEE,2002:209-212.
- [17] Songhita M, Tusharkanti D, Partha S, et al. Comparison of MFCC and LPCC for a fixed phrase speaker verification system, time complexity and failure analysis[C]//International conference on circuit, power and computing technologies. [s. l.]:[s. n.],2015:1-4.
- [18] Yuan Y J, Zhao P H, Zhou Q. Research of speaker recognition based on combination of LPCC and MFCC[C]//International conference on intelligent computing and intelligent system. [s. l.]:[s. n.],2010:765-767.
- [19] Zhu J C, Liu Z L. Analysis of hybrid feature research based on extraction LPCC and MFCC [C]//10th international conference on computational intelligence and security. [s. l.]:[s. n.],2014:732-735.
- [20] Kopparapu S K, Laxminarayana M. Choice of Mel filter bank in computing MFCC of a resampled speech[C]//10th international conference on information sciences signal processing and their applications. [s. l.]:[s. n.],2010:121-124.
- [21] 周萍,李晓盼,李杰,等. 混合 MFCC 特征参数应用于语音情感识别[J]. 计算机测量与控制,2013,21(7):1966-1968.
- [22] 庞程,李晓飞,刘宏. 基于 MFCC 与基频特征贡献度识别说话人性别[J]. 华中科技大学学报:自然科学版,2013(S1):108-111.
- [23] 沈燕,肖仲喆,李冰洁,等. 采用 GW-MFCC 模型空间参数的语音情感识别[J]. 计算机工程与应用,2015,51(10):219-222.
- [24] 张家騄. 论语音技术的发展[J]. 声学学报,2004,29(3):193-199.
- [25] Watanabe A. Formant estimation method using inverse-filter control[J]. IEEE Transactions on Audio Processing,2001,9(4):317-326.
- [26] Rao P, Barman A D. Speech formant frequency estimation: evaluating a nonstationary analysis method[J]. Signal Processing,2000,80(8):1655-1667.
- [27] 韩志艳,伦淑娟,王健. 基于遗传小波神经网络的语音情感识别[J]. 计算机技术与发展,2013,23(1):75-78.
- [28] 韩志艳,伦淑娟,王健. 语音信号鲁棒特征提取及可视化技术研究[M]. 沈阳:东北大学出版社,2012.