

文章编号: 2095-2163(2021)12-0197-05

中图分类号: TN912.34

文献标志码: A

基于融合特征的汽车鸣笛声识别方法

邓鑫, 王岩松, 杨超, 郭辉

(上海工程技术大学 机械与汽车工程学院, 上海 201620)

摘要: 针对复杂路况和现实环境条件下,传统汽车鸣笛声识别方法的分辨力不足问题,提出了一种基于变分模态分解和融合特征的汽车鸣笛声识别方法。采用变分模态分解(VMD)将鸣笛声信号分解为多个固有模态信号(IMF),基于峭度准则筛选出主IMF分量并重构信号;提取重构信号的MFCC和LPCC特征参数。利用融合算法获得基于上述两种特征的融合特征参数,并将其作为BP神经网络模型的输入特征,实现汽车鸣笛声的准确识别。研究表明,相较于单一特征方法,基于融合特征方法可有效提取汽车鸣笛声特征,提升识别准确率。

关键词: 变分模态分解; 特征提取; 融合算法; BP神经网络

Recognition method of car whistle based on fusion feature

DENG xin, WANG yansong, YANG chao, GUO hui

(College of Automotive Engineering, Shanghai University of Engineering and Science, Shanghai 201620, China)

【Abstract】 Under complex road conditions and real environment conditions, the recognition methods of traditional car whistle usually fails. A car whistle recognition method based on variational mode decomposition and fusion features is proposed. Variational mode decomposition (VMD) is used to decompose the whistle signal into multiple intrinsic mode signals (IMF). Based on kurtosis criterion, the principal IMF components are selected and the signal is reconstructed. The MFCC and LPCC characteristic parameters of the reconstructed signal are extracted. The fusion feature parameters based on the above two features are obtained by using the fusion algorithm, which are used as the input features of BP neural network model to realize the accurate recognition of car whistle. The results show that compared with the single feature method, the fusion feature method can effectively extract the features of car whistle and improve the recognition accuracy.

【Key words】 variational modal decomposition; feature extraction; fusion algorithm; BP neural network

0 引言

语音特征提取方法被广泛应用于各类信号的特征提取。传统的语音特征提取方法包括:梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient, MFCC)、线性预测倒谱系数(Linear Prediction Cepstral Coefficients, LPCC)等。文献[1-3]中采用MFCC特征提取,在相应病理异常分类识别中取得了良好效果。文献[4]通过替换离散余弦变换,提高了对家庭中危险声学事件的检测效果。在MFCC的改进方面,文献[5]将SLCF和SSF分别与MFCC融合成新特征,降低了孤立字识别系统中的总错误率。文献[6]将谱熵梅尔积与MFCC结合,提高了信噪比环境下,语音端点检测的准确率。文献[7]采用逆MFCC变换,在DCASE给定声学场景分类中取得较好的准确率;文献[8]将Teager能量算子引入MFCC,在文本独立扬声器验证任务中,识别效果

明显优于MFCC方法;文献[9]将线性预测-希尔伯特变换与MFCC结合,有效识别不良语音和正常语音。对于非语音信号和非平稳信号,MFCC的特征提取能力不足。LPCC在口译准确度^[10]、情感分类^[11]和文本语音识别^[12]方面都有不同程度的研究及应用,但是LPCC对含噪信号特征提取效果不佳。

综上所述,现实环境中背景噪声大,汽车鸣笛声信号瞬时性强,因此基于单一特征的汽车鸣笛声识别方法效果有待提升。

本文提出的基于融合特征的汽车鸣笛声识别方法,具有识别率高、鲁棒性强、计算快速等特点。该方法首先对鸣笛声信号进行变分模态分解(Variational Modal Decomposition, VMD),获得多个本征模态分量,基于峭度准则筛选出主模态函数(Intrinsic Mode Function, IMF),并重构信号;随后,提取重构信号的MFCC和LPCC特征,并利用ReliefF算法实现特征降维和特征融合;最后将融合特征输入

作者简介: 邓鑫(1996-),男,硕士研究生,主要研究方向:声源识别;王岩松(1971-),男,博士,教授,主要研究方向:车辆声学、振动及其控制。

通讯作者: 王岩松 Email: jzwb@163.com

收稿日期: 2021-06-22

BP 神经网络,实现汽车鸣笛声的准确识别。

1 特征提取及融合算法

1.1 特征提取

MFCC 特征参数是基于人耳对不同频率声音信号的感知能力不同所提出^[13]。标准由 MFCC 参数及其一阶差分和二阶共同组成 MFCC 特征参数。一般取前 12 维 MFCC 特征作为特征参数。

Mel 频率与 Hz 频率的非线性关系近似表示为:

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

线性预测模型是基于最小均方差准则,对声音信号实际值进行预测。当实际值与预测值误差最小时,利用 Durbin 算法求得线性预测系数 LPC。一般将线性预测系数通过倒谱域转化为等效参数,即 LPCC 系数。

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^p a_i s(n-i) \quad (2)$$

式中, $\hat{s}(n)$ 表示预测值; a_k 表示线性预测系数; $s(n-i)$ 表示 $n-i$ 时刻信号采样值。

1.2 融合算法

1.2.1 Fisher 融合算法

Fisher 算法通过寻找最佳投影方向,使得不同类样本间的离散程度最大,同时使得同类样本间的离散程度最小^[14]。声音特征参数有效性 Fisher 定义为:

$$F = \frac{\sigma_{between}}{\sigma_{within}} \quad (3)$$

其中, $\sigma_{between}$ 表示同类样本内的离散程度, σ_{within} 表示不同类样本间的离散程度。

1.2.2 ReliefF 融合算法

Relief 算法^[15]由 Kira 等提出。算法根据样本类别和各个特征的相关性,赋予特征不同权重,仅应用于两类问题的分类。Konoenko 等^[16]在此基础上,提出了 ReliefF 算法,应用于多类问题的分类。本文采用 ReliefF 算法作为融合特征算法。

某一特征 i 的权重更新公式为:

$$W(i) = W(i) - \text{diff}(i,R,H) / m + \text{diff}(i,R,M) / m \quad (4)$$

式中, H 代表与样本 R 在特征 i 上最近邻的同类样本; M 代表与样本 R 在特征 i 上最近邻的不同类样本; m 代表算法迭代次数。

不同样本在某一特征 i 上的距离计算公式为:

$$\text{diff}(i,R_1,R_2) = \frac{|R_1(i) - R_2(i)|}{|\max(i) - \min(i)|} \quad (5)$$

其中, $R_1(i)$ 和 $R_2(i)$ 表示两个样本在特征 i

上的数值。

ReliefF 融合特征算法的具体实现过程如图 1 所示。

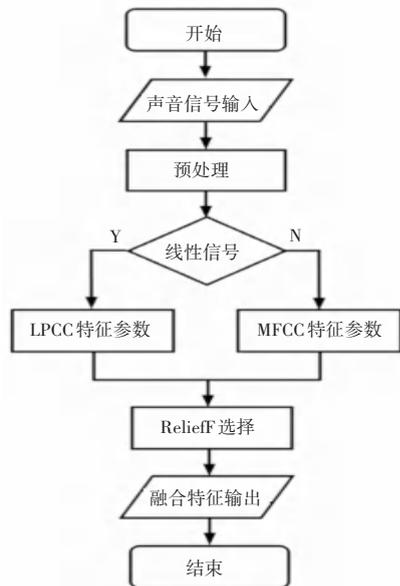


图 1 基于 ReliefF 特征参数融合流程图

Fig. 1 Flow chart of feature parameter fusion based on ReliefF

2 汽车鸣笛声 VMD 处理及融合特征

2.1 VMD 处理

变分模态分解是由 Konstantin Dragomiretskiy 等^[17]提出的一种非递归自适应的模态变分方法,能将复杂信号分解为中心频率 U_k 固定、频率带宽有限的多个模态分量(IMF)。本文利用 VMD 分解汽车鸣笛声信号,取最大分解层数 $k=8$,同时计算各分量峭度值,见表 1。基于峭度准则,当峭度最大时,确定 VMD 最佳分解层数 $k=4$ 。

表 1 各 IMF 分量峭度值

Tab. 1 Kurtosis value of each IMF component

IMF	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8
峭度值	4.040	4.502	4.225	7.877	3.846	3.776	7.072	2.994

由图 2、图 3 可知,VMD 分解原始声音信号获得 4 个模态分量。当分解层数 $k=4$ 时,原始信号重构效果较好。

2.2 特征分析

2.2.1 MFCC 及 LPCC 特征参数

图 4 中,3 种声音在第 4、6 维 MFCC 特征参数的取值不同,表明其标准 MFCC 静态特征和一阶 MFCC 动态特征具有显著差异性。图 5 中,3 种声音在第 9~12 维 LPCC 特征参数的取值不同,表明 3 种声音的波形及共振峰特征差异明显。因此,可以利用不同类型声音在某些特征系数上的差异性进行识别。



图 2 VMD 处理的汽车鸣笛声时域及频域图

Fig. 2 An example of time-frequency domain diagram of car whistle processed by VMD

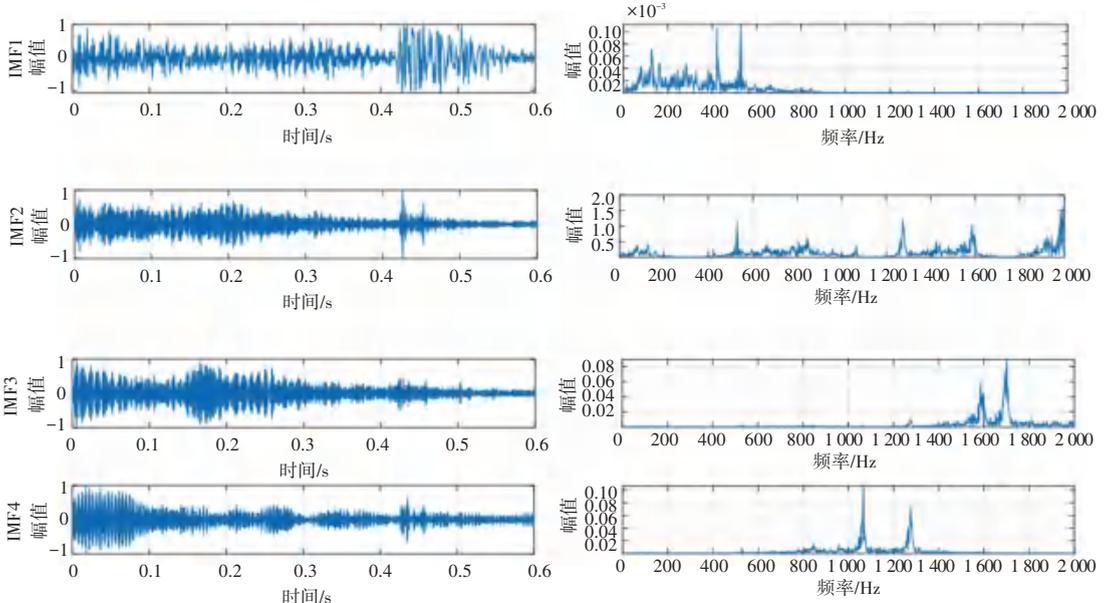


图 3 各 IMF 分量时域及频域图

Fig. 3 Time-frequency domain diagram of each IMF component

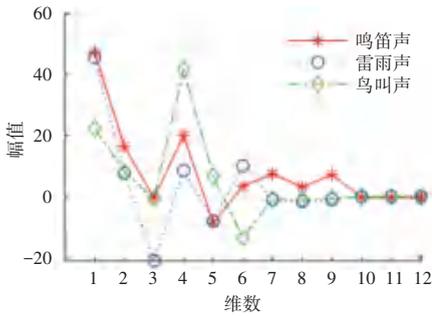


图 4 3 种声音信号中提取的某帧 12 维 MFCC 特征系数

Fig. 4 The 12 dimensional MFCC feature coefficient extracted from three kinds of sound signals in some frame

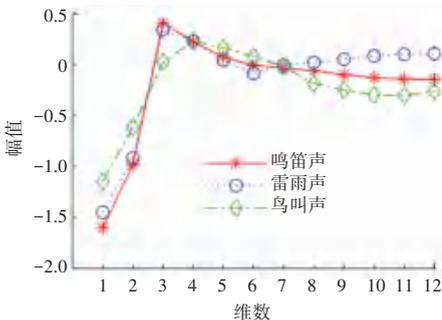


图 5 3 种声音信号中提取的某帧 12 维 LPCC 特征系数

Fig. 5 The 12 dimension LPCC feature coefficient extracted from three kinds of sound signals in some frame

2.2.2 融合特征

ReliefF 算法中, 权重值越大, 表明该特征参数对区分不同类样本的能力越强; 权重值越小, 表明该特征参数对区分不同类样本的能力越弱。

基于 Fisher 准则和 ReliefF 算法, 图 6、图 7 反映了不同维数特征在 MFCC 和 LPCC 特征中的权重值不同。

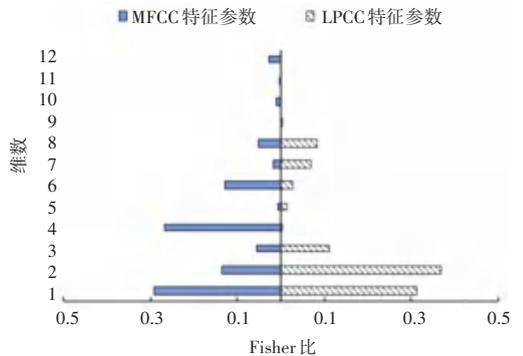


图 6 MFCC、LPCC 特征的 Fisher 比

Fig. 6 Fisher ratio of MFCC and LPCC features

由图 6 可见, 基于 Fisher 准则的第 3 维 MFCC 特征权重值为 0.293 6, 表明在 Fisher 准则下, 标准 MFCC 静态特征比一阶、二阶 MFCC 动态特征更能反映 3 种声音之间的差异性。基于 Fisher 准则的第 2 维 LPCC 特征权重值为 0.368 2, 表明 3 种声音的波形及共振峰

特征差异在第 2 维 LPCC 特征参数差异显著。

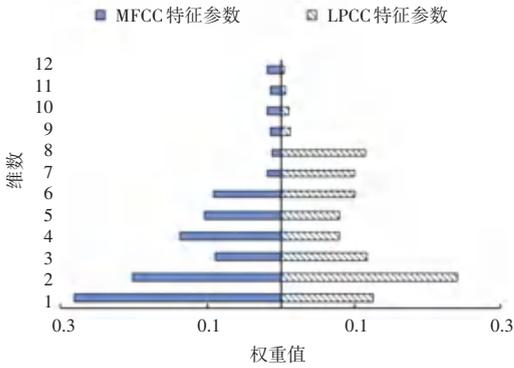


图 7 MFCC、LPCCC 特征的 ReliefF 权重

Fig. 7 ReliefF weight of MFCC and LPCCC features

由图 7 可见,基于 ReliefF 算法的第 1 维 MFCC 特征权重值为 0.280 5,并且反映动态特征的一阶、二阶参数的权重值显著高于 Fisher 准则下对应参数的 Fisher 比值。由此表明 ReliefF 算法能更加充分利用声音信号的动态特征,从而提高对声音信号帧与帧之间相关度的利用率。基于 ReliefF 算法的第 2 维 LPCCC 特征权重值为 0.240 4,高维 LPCCC 特征参数的权重值显著高于 Fisher 准则下同类特征,表明 ReliefF 算法对反映通道特性的特征利用更充分、更全面地反映 3 种声音波形及共振峰特征的差异性。

3 特征识别

3.1 分类器训练

文中采用 AudioSet 数据库的声音样本,样本总计 300 例声音信号。其中,100 例汽车鸣笛声、100 例鸟叫声、100 例雷雨声。声音样本预处理包括:预加重、分帧、加窗等。采样频率为 44.1 KHz,数字量化为 16 bit,帧长为 25ms,帧移为 10 ms,文件采用 Wav 格式保存。BP 神经网络模型网络设置:中间层 10 层,70%为训练样本、15%为验证样本、15%为测试样本。经过 VMD 处理的声音信号识别结果见表 2。

表 2 经过 VMD 的声音信号识别结果

Tab. 2 Recognition results of sound signals processed by VMD

特征	类型	总数	BP 识别率/%
MFCC	鸣笛声	100	94.7
	鸟叫声	100	
	雷雨声	100	
LPCCC	鸣笛声	100	72.5
	鸟叫声	100	
	雷雨声	100	

由表 2 可知,经过 VMD 处理的声音信号,MFCC 和 LPCCC 特征参数在 BP 中的识别率分别为 94.7% 和 72.5%。MFCC 特征的识别率显著高于 LPCCC 特

征的识别率,说明 MFCC 对汽车鸣笛声的表征能力更强。不同特征提取方法在 PB 神经网络中识别结果见表 3。

表 3 不同特征提取方法在 BP 神经网络中的识别结果

Tab. 3 Recognition results of different feature extraction methods in BP

特征	声音	总数	BP 识别率/%
MFCC	鸣笛声	100	94.7
	鸟叫声	100	
	雷雨声	100	
LPCCC	鸣笛声	100	72.5
	鸟叫声	100	
	雷雨声	100	
Fisher	鸣笛声	100	93.6
	鸟叫声	100	
	雷雨声	100	
ReliefF	鸣笛声	100	95.9
	鸟叫声	100	
	雷雨声	100	

由表 3 可知,通过对单一特征和融合特征在 BP 神经网络中识别准确率进行对比,本文提出的基于 ReliefF 融合特征算法识别率最高,达到 95.9%,优于其余 3 种特征的识别率。

3.2 实验及结果分析

实验采用真实车辆鸣笛声,音响播放鸟叫声和雷雨声。实验在半消声室(9.8 m×8.6 m×3.5 m,长×宽×高)内进行,采用 PCB 麦克风及 LMS SCADAS Mobile 数据采集仪测量声压信号。

纵向分析表 4 可知,在 BP 模型中,LPCCC 特征参数的识别率最低,仅为 80.3%,而其余 3 种特征提取方法的识别率均在 95% 以上。基于 ReliefF 融合特征方法识别率优于单一特征 MFCC 和基于 Fisher 融合特征方法,达到 98.9%,比 LPCCC 提高 23.2% 以上,表明本文所提融合特征方法优于单一特征方法。

表 4 4 种特征提取方法在 BP 神经网络中的识别率

Tab. 4 The recognition rate of four feature extraction methods in BP

特征	声音	总数	BP 识别率/%
MFCC	鸣笛声	28	97
	鸟叫声	28	
	雷雨声	28	
LPCCC	鸣笛声	28	80.3
	鸟叫声	28	
	雷雨声	28	
Fisher	鸣笛声	28	96.8
	鸟叫声	28	
	雷雨声	28	
ReliefF	鸣笛声	28	98.9
	鸟叫声	28	
	雷雨声	28	

(下转封三)