

doi:10.11835/j.issn.1000-582X.2021.11.003

基于基频的梅尔倒谱系数在车辆识别中的应用

李成娟^{1,2}, 易强^{1,2}, 李宝清¹, 王国辉¹

(1.中国科学院上海微系统与信息技术研究所微系统技术重点实验室,上海201800;
2.中国科学院大学,北京100049)

摘要:用传统的梅尔倒谱系数作为特征进行车辆识别时,识别效果易受噪声干扰。为增强特征鲁棒性,提出一种加权的基频自适应梅尔倒谱系数特征提取算法。首先用能熵比法对车辆声音信号进行端点检测;然后提取车辆信号的基频,自适应构建三角滤波器组,提高滤波器对基频的感知敏感度;最后对基频自适应梅尔倒谱系数进行F比加权。实验结果表明,与传统梅尔倒谱系数相比,在识别车辆时,加权的基频自适应梅尔倒谱系数识别准确率提高7.10%,虚警率降低3.93%,漏警率降低7.10%。

关键词:梅尔倒谱系数;特征提取;车辆识别;基频

中图分类号:TN912.16

文献标志码:A

文章编号:1000-582X(2021)11-017-07

The application of Mel-Frequency Cepstral Coefficients technology based on fundamental frequency in vehicle recognition

LI Chengjuan^{1,2}, YI Qiang^{1,2}, LI Baoqing¹, WANG Guohui¹

(1. Science and Technology on Microsystem Laboratory, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 201800, P. R. China;
2. University of Chinese Academy of Sciences, Beijing 100049, P. R. China)

Abstract: The Mel-Frequency Cepstral Coefficients (MFCC) are susceptible to noise in field vehicle recognition. To enhance the robustness of features, this paper proposed a weighted and adaptive feature extraction algorithm based on the MFCC method. Firstly, the energy to entropy ratio method was used to detect the endpoint of field vehicle's acoustic signal. Then, the fundamental frequency of vehicle's acoustic signal was extracted. The triangular filter bank was adaptively constructed according to the fundamental frequency so as to improve the filter's sensitivity to the fundamental frequency. Finally, the obtained frequency was weighted with fisher's ratio. Compared with the traditional MFCCs, the experimental results show that the improved MFCCs improve the recognition accuracy by 7.10%, reduce the false alarm rate by 3.93% and reduce the leakage alarm rate by 7.10% in field vehicle recognition.

Keywords: Mel-Frequency Cepstral Coefficient; feature extraction; vehicle recognition; fundamental frequency

收稿日期:2020-10-21

基金项目:微系统技术重点实验室基金项目(614280401020617)。

Supported by the Foundation of Science and Technology on Microsystem Laboratory (614280401020617).

作者简介:李成娟(1994—),女,硕士,主要从事声音信号识别方向研究,(E-mail)leecj1425@163.com。

通讯作者:王国辉(1966—),男,高级工程师,硕士生导师,主要从事信号识别方向研究,(E-mail)Wgh0415@163.com。

无线传感器网络凭借其通信能力强、记录信息全面、无需人工值守等优点,广泛应用于军事、农业、商业等各个领域^[1-2]。无线传感器网络通常由图像、声音、红外、震动传感器等几种类型的传感器构成^[3]。声音传感器有不易被察觉、便于布设、不易受地形影响的优点,被广泛使用。

根据车辆行进过程中采集的声音信号,对野外车辆目标进行识别,一般分为2个步骤:1)提取声音传感器采集的声信号特征;2)设计分类器得到判别结果。目前常用的特征有基于小波包变换提取的特征,线性预测倒谱系数^[4],梅尔倒谱系数(MFCC, mel-frequency cepstral coefficient)^[5],基于经验模态分解提取的特征^[6-7]等。其中,MFCC较常见,被广泛用于语种识别、说话人识别、声纹识别等领域,近年来被成功应用于车辆目标分类识别^[8-10]。但由于MFCC对噪声敏感^[11-13],在车辆识别领域的应用场景受限。

为提高特征抗噪能力,增强特征稳健性,能够体现野外车辆信号自身特征,提出一种改进的MFCC特征。根据车辆目标信号的频谱特征,自适应增加车辆基频信息在特征中的比重。然后进行 F 比加权,以降低同种车辆类别特征间的区分度,增大不同车辆类别特征之间的差别,从而提高特征在野外环境下的鲁棒性。使用高斯混合模型(GMM, gaussian mixture model)作为分类器,检验识别效果。提取传统MFCC特征,基频自适应MFCC特征,加权基频自适应MFCC特征作对比试验,期望获得更高的识别准确率,为野外环境的车辆识别,提供一种鲁棒特征提取方法。

1 传统的MFCC特征

人耳对声音信号的感知,与频率呈非线性关系。随着声音信号的频率降低,人耳接收信号的能力愈发敏感^[14]。为使人耳对信号敏感程度与信号频率呈现线性关系,MFCC参数将信号频谱作映射处理,映射后的刻度称作梅尔刻度。映射函数的表达式如下

$$M(f) = 2595 \times \lg\left(1 + \frac{f}{700}\right), \quad (1)$$

式中: $M(f)$ 为梅尔频率; f 为频率。

第一步,提取车辆目标声信号的MFCC,首先需要在时域尺度对其进行分帧、加窗、预加重等预处理。为减小频谱泄漏,窗函数选取旁瓣衰减较大的汉明窗。预加重,实际是让信号通过一个高通滤波器,目的是滤除车辆声音信号中的低频干扰成分,增加高频成分的比重^[15]。

第二步,为了将信号从时域尺度转换为频域尺度,作快速傅里叶变换,且计算每帧信号的能量谱。

第三步,对能量谱进行三角滤波。梅尔三角滤波器由公式(2)计算得出。

$$H_i(k) = \begin{cases} 0, & k < f_{i-1} \text{ or } k > f_{i+1}, \\ \frac{2(k - f_{i-1})}{(f_{i+1} - f_{i-1})(f_i - f_{i-1})}, & f_{i-1} \leq k \leq f_i, \\ \frac{2(f_{i+1} - k)}{(f_{i+1} - f_{i-1})(f_i - f_{i-1})}, & f_i \leq k \leq f_{i+1}, \end{cases} \quad (2)$$

式中: $H_i(k)$ 表示滤波器参数; f_i 表示三角滤波器的中心频率。滤波后取对数,得到对数能量。

最后,进行离散余弦变换,得到MFCC。为了获取动态信息,提取语音信号的MFCC特征时,通常包含动态差分步骤。与语音信号相比,车辆声音信号频带窄,周期性强,没有必要获取动态特性,因此特征提取算法省略动态差分步骤。

2 改进的MFCC特征

传统的MFCC在识别纯净声音信号时,可得到很好识别结果,但对噪声的敏感度很高,当采集的实际声音信号包含噪声信息时,识别效果易受干扰,难以预测识别效果是否符合实际需求。而且,当风力等级变化时,很容易造成训练样本集与测试样本集不匹配。为降低噪声对识别效果的干扰,减少应用场景的限制,对传统MFCC特征提取算法进行改进。

2.1 基频自适应梅尔倒谱系数

由于不同车型的传动系统、排气系统、发动机及其气缸数不同,不同车辆声信号的频域特性存在差异。

不同车辆声信号频谱图如图 1 所示。从图中可看出不同车型车辆声音信号的基频是不同的,因此车辆声音信号的基频可以作为车辆目标识别特征。将车辆声音信号的 MFCC 参数与基频特征结合,得到基频自适应梅尔倒谱系数。

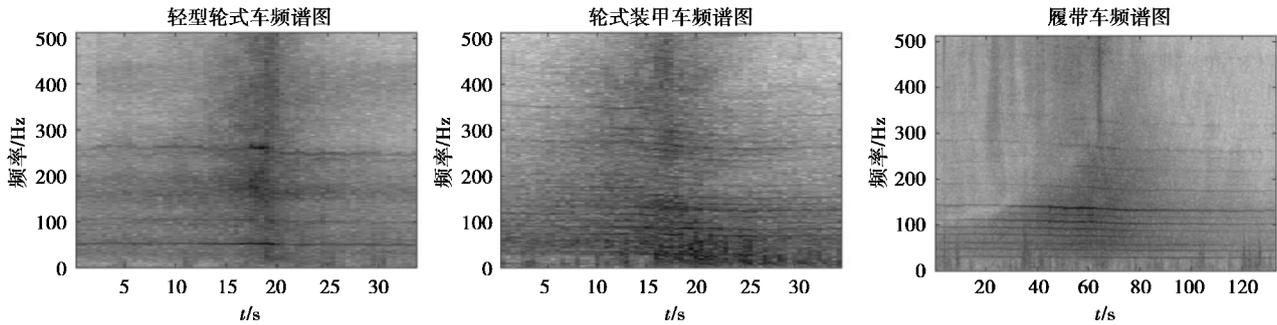


图 1 不同车辆声信号频谱图
Fig. 1 Different vehicles' spectrograms

提取传统的 MFCC 时,梅尔三角滤波器组中心频率根据梅尔刻度获得,随着频率提升,中心频率的密度逐渐降低。在野外情况风噪较大时,信号频谱中含有较多干扰频率,传统梅尔三角滤波器组有可能削弱基频在频谱中的比重,加大无关频率的干扰。为了更好体现车辆自身信号的频谱特征,在传统的滤波器组中,增添一个三角滤波器。该滤波器的中心频率为车辆信号的基频,通过此方法提取出来的特征称为基频自适应梅尔倒谱系数。改进的 MFCC 提取过程如图 2 所示。

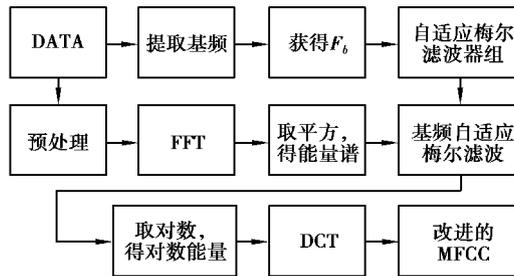


图 2 改进 MFCC 特征提取算法框图
Fig. 2 Flowchart of improved MFCC

自适应的梅尔三角滤波器组具体设计方法如下。假设传统梅尔三角滤波器组包含 N 个滤波器,其中心频率组成向量 $F=(f_1, f_2, \dots, f_N)$ 。提取车辆信号的基频 f_b ,将 f_b 与三角滤波器组的中心频率按升序排列,组合构成新的向量 $F_b=(f_1, f_2, \dots, f_b, \dots, f_{N+1})$ 。将 F_b 向量的各维分量作为滤波器的中心频率,代入式(2),即可得到自适应的三角滤波器组。传统梅尔三角滤波器组与基频自适应的三角滤波器组如图 3 所示。

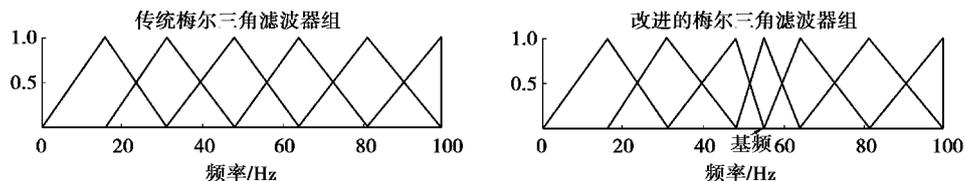


图 3 三角滤波器组
Fig. 3 Triangular filter-bank

为了得到基频自适应梅尔滤波器组,进一步获得基频自适应梅尔倒谱系数,需要选择合适的基频提取方法,准确提取车辆声音信号的基频。

2.2 基频提取方法

常用的单基频提取算法包括自相关函数法、倒谱法、谱减-自相关函数法、平均幅度差函数法、简化逆滤波法等^[16-17]。首先采用自相关函数法提取基频,考虑到野外环境噪声大的实际情况,同时用谱减-自相关函数法作对比实验。

自相关函数法提取单基频时,为了降低共振峰干扰,需要先对信号进行带通滤波。然后对信号进行加窗、分帧,计算短时自相关函数。设车辆声音信号第 i 帧的时间序列为 $data_i(m)$,其短时自相关函数 $R_i(k)$ 由式(3)得到

$$R_i(k) = \sum_{m=0}^{N-1-k} data_i(m) data_i(m+k), \quad (0 \leq k \leq K), \quad (3)$$

由于信号短时自相关函数周期与信号周期相同,因此基音周期可以取短时自相关函数的最大延迟量。对基音周期曲线进行平滑处理,使之通过中值和线性组合滤波,得到车辆的单基频信息。

谱减-自相关函数法,前提假设为车辆声音信号中噪声为加性噪声且噪声较平稳。将带噪车辆声音信号频谱与噪声频谱相减后,通过自相关函数法提取单基频,谱减法流程如图 4 所示。

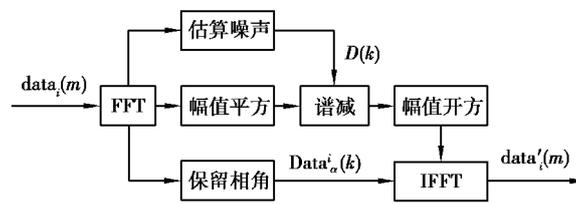


图 4 谱减法流程图

Fig. 4 Flowchart of spectral subtraction

为了进一步满足野外环境的实际应用需求,将单基频提取算法改进为多基频提取算法。运动目标在行驶过程中生成的声信号,可由其发声主要部件产生的声信号表示,其余部件产生的声信号比较微弱,可忽略不计。履带车产生声信号的主要部件为履带和引擎,轻型轮式车和轮式装甲车产生声信号的主要部件是排气系统和引擎,因此车辆运动过程中的声信号可由 2 个主要成分表示,每个主要成分产生一条基频。下面简要介绍多基频提取算法。

首先提取车辆声信号的 2 条基频,可参照单基频提取方法。然后,将车辆的 2 个基频进行融合,得到一个新的基频 $f_{\text{new}}(t)$ 。设某条车辆声信号第 t 帧的 2 个基频,其中较大的数值为 $f_{b1}(t)$ 、较小的数值为 $f_{b2}(t)$,则 $f_{\text{new}}(t)$ 可由式(4)得到^[18]

$$f_{\text{new}}(t) = \frac{1}{2} (f_{b1}(t) + \frac{\overline{f_{b1}(t)}}{\overline{f_{b2}(t)}} f_{b2}(t)), \quad (4)$$

其中, $\overline{f_{b1}(t)}$ 和 $\overline{f_{b2}(t)}$ 分别表示该条信号所有帧中, 2 条基频的平均值。多基频提取算法得到融合基频比单基频更符合车辆本身信号特征,因此用融合后的基频计算得到基频自适应梅尔倒谱系数,用于车辆识别,获得更高的识别精度。

2.3 加权基频自适应梅尔倒谱系数

MFCC 的每一维分量对表征特征的贡献率不同,为增加不同车型特征之间差距,使同一车型的特征更为紧凑,在基频自适应梅尔倒谱特征基础上,进行 F 比加权。 F 比参数,是在 Fisher 准则基础上提出的^[19],由类间离散度与类内离散度之比得到,可定量分析特征向量每一维分量有效性。设数据集包含 M 种车辆类别的车辆声音信号,其中第 i 个类型包含 N_i 条样本信号,则特征参数第 k 维特征分量的 F 比具体计算公式如下

$$F(k) = \frac{\sum_{i=1}^M (m_i(k) - m(k))^2}{\sum_{i=1}^M (\frac{1}{N_i} \sum_{j=1}^{N_i} (m_i^j(k) - m_i(k))^2)}, \quad (5)$$

式中: $m_i^j(k)$ 为第 i 类车辆类别第 j 条样本的第 k 维特征分量所有帧之和; $m_i(k)$ 为第 k 维特征分量在第 i 类车辆类别所有 N_i 条样本的所有帧之和; $m(k)$ 为第 k 维特征分量在整个数据集中所有帧之和。

特征参数中某维分量的 F 比参数数值越小,表明该特征分量在同类目标之间方差较大,在不同类目标之间方差较小,对表征信号特征的贡献率也越低。以 F 比为权重,对特征向量的每一维分量加权,能够削弱贡献率较低的特征分量干扰,提高贡献率较高分量的作用,从而提取更能体现不同车辆类别间差异的特征。

3 仿真实验

实验采用 MATLAB R2015b 平台仿真,样本信号采用 ICS40720 传声器设备进行采集,采样率为 8 192 Hz。外场实验场景如图 5 所示,目标车辆从起始位置行驶到终止位置,麦克风便采集到一条完整的样本。

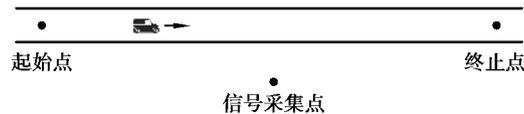


图 5 实验场景图

Fig. 5 Experimental scenario

仿真实验目的是根据野外无线传感器网络中声音传感器采集的信号,检验不同特征提取方法对车辆识别效果的影响。为了更贴近无线传感器网络的实际应用场景,声音传感器布设位置,距离道路中心不超过 50 m,起始位置到终止位置的车辆行驶总长度限制在 800 m 以内。当野外风噪声非常大时,无线传感器网络中识别车型主要通过图像和震动传感器,声音传感器难以采集到有效车辆声音信号。因此去掉数据集中风力级在 6 级以上采集的声音信号,保留数据对应的风力级主要集中在 1~5 级。

3.1 实验过程

为验证改进的特征提取算法的实际识别效果,共采集硬土路、砂石路、水泥路及草地 4 种不同路面的车辆声音信号作为样本。车辆类型包含一种小型车,两种大型车,不同车型的行驶速度范围不同。小型车运动速度为 25~60 km/h,代表车型为轻型轮式车,如卡车。大型车运动速度为 10~20 km/h,代表车型为重型轮式车、履带车,如坦克。按照实际需求以及声音信号的特征,在实验中将车辆分为履带车、轮式装甲车和轻型轮式车 3 种类型。样本集中每种车辆类别样本数和总帧数如表 1 所示,将样本数按照 70% 和 30% 的比例随机分为训练集和测试集,用于仿真实验。

表 1 野外车辆样本数

Table 1 Different field vehicles' specifications

参数	轻型轮式车	轮式装甲车	履带车
条数	72	66	30
帧数	10 999	7 311	12 887

采集数据后进行手工裁剪或拼接,由于存在人工误差,样本残留了部分噪声段。为获得更准确实验数据,对信号进行端点检测,得到有效运动目标数据后,再进行特征提取和分类识别。谱熵是声音信号的一种特征参数,对噪声不敏感,具有较好稳健性,因此,采用能熵比法对车辆信号进行端点检测。

检测到有效数据段后,首先提取维度为 23 的 MFCC 特征,然后提取维度为 24 的基频自适应梅尔倒谱特征,最后对其进行 F 比加权。

3.2 结果分析

识别效果采用识别准确率、漏警率和虚警率 3 个标准进行评判。结果如表 2 所示,表格中方法 1 表示用传统的 MFCC 作特征;方法 2 表示先对信号进行端点检测,然后用传统的 MFCC 作特征;方法 3 表示端点检测后用自相关法提取单基频,然后用基频自适应 MFCC 作特征;方法 4 表示端点检测后用谱减-自相关法提

取单基频,然后用基频自适应 MFCC 作特征;方法 5 在方法 4 的基础上进行 F 比加权;将方法 5 中单基频提取算法改为多基频提取算法,即方法 6。提取特征后,选择在声音识别领域效果好,且计算量较小的 GMM^[20] 分类器进行识别,结果如表 2 所示。

表 2 野外车辆识别结果
Table 2 Field vehicles' classification accuracies %

方法	准确率	漏警率	虚警率
方法 1	87.46	12.54	6.37
方法 2	90.85	9.15	4.25
方法 3	92.05	7.95	3.71
方法 4	93.64	6.36	3.03
方法 5	94.12	5.88	2.53
方法 6	94.56	5.44	2.44

分析表中数据可得,在野外车辆识别时,方法 2 与方法 1 相比识别准确率提高 3.39%,漏警率降低 3.39%,虚警率降低 2.12%,表明在提取特征之前,加入端点检测可以提高识别效果。对比方法 3 和方法 4 数据可以看出,在野外车辆识别中,谱减-自相关法提取比自相关法提取的基频更加适用。对比方法 4 和方法 5 数据可以得出结论,对特征参数进行 F 比加权后,识别效果提升。对比方法 5 和方法 6 数据可以看出,与单基频提取算法相比,多基频提取算法识别效果有所提升,更能满足实际应用需求。

F 比加权的基频自适应梅尔倒谱参数与传统的 MFCC 相比,识别准确率提高 7.10%,漏警率降低 7.10%,相对值降低 56.62%,虚警率降低 3.93%,相对值降低 61.70%,有更高的鲁棒性,更适用于野外环境中车辆识别。

4 结 论

在野外无线传感器网络中,使用声音传感器对车辆进行识别时,基于车辆声音信号的基频,针对 MFCC 容易受噪声影响的情况,对特征提取算法进行改进,提出 F 比加权的基频自适应梅尔倒谱特征。使用能熵比法对车辆信号进行端点检测后,用谱减-自相关算法提取基频。在提取特征过程中,根据车辆基频重新构建滤波器组。为进一步加大三种车辆类别间的特征差距,对改进的特征向量进行 F 比加权优化。最后使用 GMM 模型建模,得出分类识别结果。实验结果表明, F 比加权的基频自适应梅尔倒谱特征识别野外车辆时,准确率提升,虚警率和漏警率降低,因此更加适用于野外环境中车辆识别。

虽然 F 比加权的梅尔倒谱特征具有较好鲁棒性,但在进行特征提取之前,对野外采集的信号进行增强并降噪也非常重要。如何有效去除运动目标声信号中包含的噪声,将是下一步研究的重点内容。

参考文献:

- [1] Shaikh F K, Zeadally S. Energy harvesting in wireless sensor networks: a comprehensive review[J]. Renewable and Sustainable Energy Reviews, 2016, 55: 1041-1054.
- [2] Kurt S, Yildiz H U, Yigit M, et al. Packet size optimization in wireless sensor networks for smart grid applications[J]. IEEE Transactions on Industrial Electronics, 2017, 64(3): 2392-2401.
- [3] Pandey M, Mishra G. Types of sensor and their applications, advantages, and disadvantages[C]//Emerging Technologies in Data Mining and Information Security, India: Springer, 2019: 791-804. DOI:10.1007/978-981-13-1501-5_69.
- [4] Witkowski M, Kacprzak S, Zelasko P, et al. Audio replay attack detection using high-frequency features[C]//Interspeech 2017. ISCA: Swedon, 2017: 27-31.
- [5] Juvela L, Bollepalli B, Wang X, et al. Speech waveform synthesis from MFCC sequences with generative adversarial

- networks[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). April 15-20, 2018, Calgary, AB, Canada: IEEE, 2018: 5679-5683.
- [6] Sharma R, Vignolo L, Schlotthauer G, et al. Empirical mode decomposition for adaptive AM-FM analysis of speech: a review[J]. *Speech Communication*, 2017, 88: 39-64.
- [7] Sharma R, Mahadeva Prasanna S R. A better decomposition of speech obtained using modified empirical mode decomposition[J]. *Digital Signal Processing*, 2016, 58: 26-39.
- [8] Huang J C, Zhang X, Guo F, et al. Design of an acoustic target classification system based on small-aperture microphone array[J]. *IEEE Transactions on Instrumentation and Measurement*, 2015, 64(7): 2035-2043.
- [9] Guo F, Huang J C, Zhang X, et al. A classification method for moving targets in the wild based on microphone array and linear sparse auto-encoder[J]. *Neurocomputing*, 2017, 241: 28-37.
- [10] Zhao Q, Guo F, Zu X S, et al. An acoustic-based feature extraction method for the classification of moving vehicles in the wild[J]. *IEEE Access*, 2019, 7: 73666-73674.
- [11] Al-Ali A K H, Dean D, Senadji B, et al. Enhanced forensic speaker verification using a combination of DWT and MFCC feature warping in the presence of noise and reverberation conditions[J]. *IEEE Access*, 2017, 5: 15400-15413.
- [12] Shahnawazuddin S, Deepak K T, Pradhan G, et al. Enhancing noise and pitch robustness of children's ASR[C]//2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). March 5-9, 2017, New Orleans, LA, USA: IEEE, 2017: 5225-5229.
- [13] Bhattacharjee U, Gogoi S, Sharma R. A statistical analysis on the impact of noise on MFCC features for speech recognition[C]//2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE). December 23-25, 2016, Jaipur, India: IEEE, 2016: 1-5.
- [14] Palo H K, Chandra M, Mohanty M N. Recognition of human speech emotion using variants of mel-frequency cepstral coefficients advances in systems[J]. *Control and Automation*, 2018: 491-498. DOI:10.1007/978-981-10-4762-6_47.
- [15] Ludeña-Choez J, Gallardo-Antolín A. Feature extraction based on the high-pass filtering of audio signals for acoustic event classification[J]. *Computer Speech & Language*, 2015, 30(1): 32-42.
- [16] Kim J W, Salamon J, Li P, et al. Crepe: a convolutional representation for pitch estimation[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). April 15-20, 2018, Calgary, AB, Canada: IEEE, 2018: 161-165.
- [17] Jouvét D, Laprie Y. Performance analysis of several pitch detection algorithms on simulated and real noisy speech data[C]//2017 25th European Signal Processing Conference (EUSIPCO). August 28-September 2, 2017, Kos, Greece: IEEE, 2017: 1614-1618.
- [18] Huang J C, Guo F, Zu X S, et al. A novel multipitch measurement algorithm for acoustic signals of moving targets[J]. *Mechanical Systems and Signal Processing*, 2016, 81: 419-432.
- [19] Hegde S, K A K, Shetty S. Feature selection using fisher's ratio technique for automatic speech recognition[J]. *International Journal on Cybernetics & Informatics*, 2015, 4(2): 45-52.
- [20] Chettri B, Sturm B L. A deeper look at Gaussian mixture model based anti-spoofing systems[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). April 15-20, 2018, Calgary, AB, Canada: IEEE, 2018: 5159-5163.