

基于矢量量化的猪咳嗽声识别

龚永杰¹ 黎 煊^{1,2} 高 云^{1,2}
雷明刚^{2,3} 刘望宏^{2,3} 杨 专¹

1. 华中农业大学工学院, 武汉 430070; 2. 生猪健康养殖协同创新中心, 武汉 430070;
3. 华中农业大学动物科技学院动物医学院, 武汉 430070

摘要 针对猪的规模化养殖中频发的呼吸道疾病问题, 提出通过监测咳嗽状况对猪的健康状况进行预警, 以谱减法去噪和端点检测为猪咳嗽信号主要预处理方法, 以矢量量化(vector quantization, VQ)匹配算法为核心算法, 分别构建基于标准梅尔频率倒谱系数(Mel-frequency cepstral coefficient, MFCC)和改进的 MFCC 2 种猪咳嗽声识别模型。测试结果显示, 以标准 MFCC 为特征矢量构建的识别系统的识别率、误判率和综合识别率分别达到 88%、14% 和 87.3%, 基于改进的 MFCC 特征矢量构建的识别系统与之相比有很大提高, 其识别率、误判率和综合识别率分别达到 91%、12% 和 90.0%。试验表明, 采用改进的 MFCC 与矢量量化相结合构建猪咳嗽识别系统是可行的, 能够应用于猪的呼吸道疾病预警。

关键词 矢量量化; 猪; 咳嗽; 识别; 梅尔频率倒谱系数

中图分类号 TP 391.42 **文献标识码** A **文章编号** 1000-2421(2017)03-0119-06

我国生猪养殖正逐步由传统的农户散养向规模化集约化转变, 如何对养殖过程中出现的呼吸道疾病进行快速监测, 为现代化养猪事业的可持续发展提供服务和支撑, 是畜牧科学领域研究的热点和难点。猪咳嗽信号中包含被测对象的喉部或气管受刺激时的振动信息, 是猪呼吸道疾病的主要表现, 通过监测咳嗽可以进行相关呼吸道疾病的识别和预警^[1]。传统方法多采取人为蹲点记录猪的咳嗽状况, 这种方法虽然简单易行, 但它不仅消耗大量的人力成本, 而且数据有很大的主观性, 监测准确度难以得到保证。

随着计算机技术的发展, 借助声音特征分析技术能够从咳嗽声中提取足够的信息用于检测和分析, 进而开发猪的咳嗽自动识别系统, 这种监测方法具有高效率、低成本、非接触和无侵入的特点^[2-5]。国内外最先将咳嗽识别应用于患者呼吸道疾病的临床诊断, 以咳嗽声的梅尔倒谱系数(Mel-frequency cepstral coefficient, MFCC)、时频能量分布等为特征参数, 采用神经网络(artificial neural networks, ANN)、隐马尔科夫(hidden markov model, HMM)、时间规整(dynamic time warping, DTW)模型对咳嗽声进行自动检测^[6-10], 这些对人的咳嗽声的研究, 为猪的咳嗽识别提供了理论依据。Silvaa 等^[11]、Farrari 等^[12]、Bahr 等^[13]研究了猪的尖叫和咳嗽所传达的健康信息, 并对不同原因引起的咳嗽在动态能量、均方根(root mean square, RMS)、峰值频率(Hz)等特征参数上的差异性进行了分析。Hirtum 等^[14]分别从建立环境噪声模型和比较不同去噪方法两方面出发, 研究了如何降低环境低频噪声对猪咳嗽声时频特性的影响。Exadaktylos 等^[15]、Hirtum 等^[16]、Ferrari 等^[17]研究了动物咳嗽的声学特征, 并提取猪咳嗽信号的时域和频域特征为参数, 构建了基于模糊 c 均值聚类算法的咳嗽识别模型, 综合识别率可达 85%。Guarinoa 等^[18]使用滤波器组与振幅解调相结合的方法提取猪咳嗽声的特征向量, 运用动态的时间规整算法建立模型进行匹配, 咳嗽声正确识别率达到了 85.5%。这些研究从不同角度对猪咳嗽声识别展开研究, 取得了较好

收稿日期: 2016-09-18
基金项目: 湖北省自然科学基金项目(2014CFB317); “十三五”国家重点研发计划项目(2016YFD0500506); 现代农业产业技术体系项目(CARS-36)
龚永杰, 硕士研究生, 研究方向: 农业智能化检测与控制. E-mail: gongliangong@webmail.hzau.edu.cn
通信作者: 黎 煊, 博士, 副教授, 研究方向: 农业智能化检测与控制. E-mail: x@mail.hzau.edu.cn

的识别效果,为猪的健康实时监测和预警系统的开发打下了基础。

目前,国内对猪咳嗽识别的研究仍然鲜见报道。矢量量化针对孤立词识别具有计算速度快、识别率高的优点,在语音识别领域得到了广泛的应用。本研究应用谱减法、端点检测和分帧加窗对猪的咳嗽信号进行预处理,提取猪咳嗽音的标准 MFCC 特征参数和改进的 MFCC 特征参数以构建基于矢量量化的猪咳嗽音识别模型,实现对猪咳嗽音的快速识别,旨在为猪的呼吸道疾病在线监测预警提供新的途径和方法。

1 咳嗽声采集和预处理

1.1 咳嗽声采集

猪咳嗽数据的采集选择在华中农业大学生猪养殖基地进行,采集设备选用先科 A20 录音笔,采样频率 48 kHz,采样精度 16 位,该设备能够连续录音 24 h,且可在较远距离获得无损声音,保证了录音质量。录音在 2~3 月龄猪呼吸道疾病多发期进行,共采集 6 头体质量 75 kg 左右的长白猪的咳嗽和非咳嗽样本,每次采集时间为 18 h。通过人工标记对采集的声音进行分类获得用于预处理的有效的咳嗽声和其他尖叫、打斗、喷嚏、甩耳朵等干扰声音,本次试验共选取 600 个咳嗽声和 100 个非咳嗽声,保存格

式均为 wav,其中 400 个咳嗽声作为训练样本集,200 个咳嗽声和 100 个非咳嗽分别作为测试样本集。

1.2 声音预处理

猪场采集的声音信号包含很多噪声和无用的信号,需要进行预处理。预处理主要为了达到 2 个目的:一是从信号中提取目标片段,主要手段是端点检测,分帧加窗;二是减小环境噪声等异常声音对特征提取的影响,主要方法包括谱减法去噪。

1)谱减法去噪。噪声信号主要来源于麦克风、声卡等硬件设备产生的噪声,主要包括直流电中包含的交流声、三极管和集成电路中的无规则电子运动产生的噪声、滤波不良产生的噪声等。实际环境中的信号是原信号和噪声信号的叠加,谱减法利用这一叠加原理从输入信号的能量值中减去估计得来的噪声平均能量值,去除信号中的稳定噪声部分,最终得到纯净的语音。噪声段平均能量值 $D(k)$ 为

$$D(k)=\frac{1}{N_{\text{IS}}}\sum_{i=1}^{N_{\text{IS}}}|x_i(k)|^2,k=0,1,\cdots,N$$
 (1)

式中 N_{IS} 为对应帧数, N 为第 i 帧信号的帧长, $x_i(k)$ 为一帧数据第 k 个采样点的数值。图 1 为咳嗽声经过去噪后的对比图,谱减后的噪声信号明显受到抑制,信噪比得到增强。

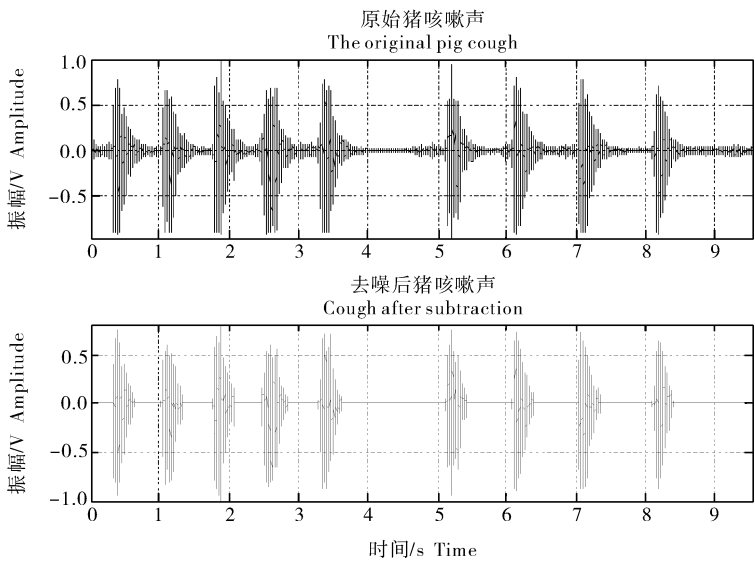


图 1 原始和谱减法去噪后的猪咳嗽声对比图

Fig.1 The comparison between the original pig cough and the cough after subtraction

2)端点检测。语音信号一般分为无声段、清音段和浊音段,清音段可视为背景噪声段,浊音段即为

语音段,浊音段平均能量最高、平均过零率最小,噪声段则平均能量居中、平均过零率最大。基于语音

信号的能量和过零率的特点,本研究采用基于平均过零率和平均能量的双重判决的双门限法检测语音信号的起止点。第 i 帧语音信号的平均能量 $E(i)$ 和平均过零率 $Z(i)$ 计算公式分别为

$$E(i)=\sum_{n=0}^{L-1}x_i^2(n) \quad (2)$$

$$Z(i)=\frac{1}{2}\sum_{n=0}^{L-1}|\operatorname{sgn}[x_i(n)]-\operatorname{sgn}[x_i(n-1)]| \quad (3)$$

式中, L 为帧长, $x_i(n)$ 为第 i 帧数据第 n 个采

样点的数值, $\operatorname{sgn}[\cdot]$ 为符号函数, 即

$$\operatorname{sgn}[x]=\begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases} \quad (4)$$

双门限检测法能够降低环境噪声对端点检测的影响, 更加准确地截取语音段信号, 图 2 为应用双门限法从一段咳嗽信号中截取单独的咳嗽声的效果图, 每一竖直的实线和虚线的间隔为一个单独声音。

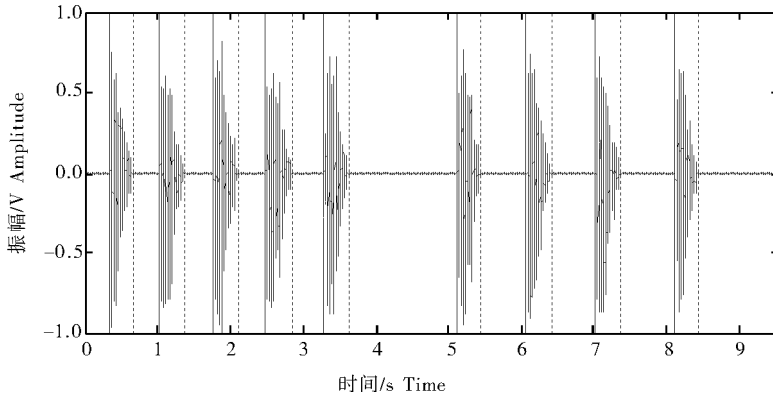


图 2 双门限法猪咳嗽声端点检测

Fig.2 The cough with double threshold endpoint detection method

3)分帧和加窗。猪的咳嗽声是一种瞬变、非周期、离散性大的信号,无法直接用于语音分析,但短时可以认为是平稳的、时不变的、连续的信号,因此,在信号处理前必须先做分帧处理,再对每一帧信号提取特征参数。加窗是分帧的主要手段,相当于用信号与一个有限长的窗函数相乘,这个过程中不同的窗函数会产生不同程度的带宽和频谱泄漏。矩形窗、海宁窗和汉明窗是语音分析中最常用的 3 种窗函数,本文选择汉明窗,相对于矩形窗和海宁窗,汉明窗主瓣宽,旁瓣明显减小,能有效减少泄漏,其定义为:

$$\omega(n)=\begin{cases} 0.54-0.46\cos(2\pi n/(L-1)) & 0 \leq n \leq L-1 \\ 0 & \text{其他} \end{cases} \quad (5)$$

式中 L 为窗长,声音采样频率为 48 kHz,当帧长取 10~30 ms 时可认为是平稳的,本研究选择帧长 512,帧移 200。

2 咳嗽声特征参数提取

咳嗽信号的特征参数主要体现在时域特征和频域特征 2 个方面,相对于时域特征,频域特征能够反映声音信号的最独特的感知特性,可以更好地描述不同声音的差异性。频域特征主要包括 Mel 频率、MFCC(Mel 频率倒谱系数)、LPC(线性预测系数)、

LPCC(线性预测倒谱系数),其中比较常用的特征参数为 LPCC 和 MFCC。

在 1 kHz 以下时,人耳对声音的感知能力与频率成线性关系,大于 1 kHz 时,感知能力则与频率成对数关系,且频率越高感知能力越差,而在 Mel 频域内,人对声音的感知能力与 Mel 频率呈现线性关系。基于人对声音的感知规律,先将线性频谱映射到基于听觉感知的 Mel 非线性频谱中,然后转换到倒谱上,最终得到的 MFCC 为原始声音信号对应的 Mel 频率上的倒谱系数。其中 Mel 频率与赫兹频率的对应关系为:

$$\operatorname{Mel}(f)=2595 \lg (1+f / 700) \quad (6)$$

式中, f 为对应频率,研究表明, MFCC 特征参数相对于 LPCC 特征参数能够更好地对语音信号进行描述,具有更好的鲁棒性^[19-20]。本研究选择分别提取标准 MFCC 和改进的 MFCC 作为输入矢量, MFCC 特征提取的过程主要包括先对声音进行预处理,通过快速傅立叶变换(fast fourier transform, FFT)得到对应的频谱,再通过滤波器组获得 Mel 频谱,最后通过离散余弦变换(discrete cosine transform, DCT)得到倒谱系数。三角带通滤波器能够减少谐波使语音的共振峰更加突出,频谱更加

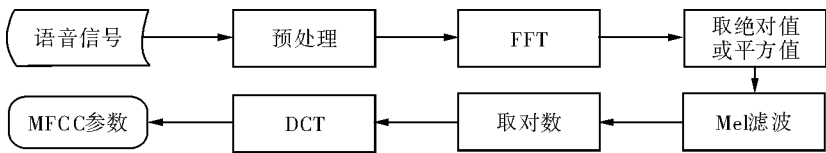


图 3 梅尔频率倒谱系数的提取

Fig.3 The extraction of Mel-frequency cepstral coefficients

平滑,本研究选择包含 24 个三角带通滤波器的滤波器组,最终获得 12 维标准 MFCC 系数,MFCC 特征向量的具体提取步骤如图 3 所示。

语音特性分为静态特性和动态特性两方面,标准 MFCC 只能反映语音的静态特性,其动态特性可以用静态参数的差分来描述,计算过程如下:

$$D_t = \begin{cases} C_{t+1} - C_t & t < K \\ C_t - C_{t-1} & t \geq Q - K \\ \frac{\sum_{K=1}^K K(C_{t+K} - C_{t-K})}{\Delta x} & t \text{ 为其他} \end{cases} \quad (7)$$

其中, D_t 为第 t 个一阶差分, C_t 为第 t 个标准 MFCC 系数, Q 为标准 MFCC 的阶数, K 为一阶导数的时间差,取 2。标准 MFCC 与差分 MFCC 相结合组成 24 维改进的 MFCC,图 4 为提取的猪咳嗽声的标准 MFCC 三维视图,静态特性前端变化较为尖锐、后端变化较为平滑,声音特异性主要集中在后端部分。

图 5 为改进的 MFCC 特征参数的三维视图,与图 4 相比,动态特性部分进一步加强了不同声音之间的差异性,表明改进的 MFCC 将静态特性和动态特性相结合对声音进行更加详细的描述,能够更好地反映咳嗽声的特异性,提高整体识别率。

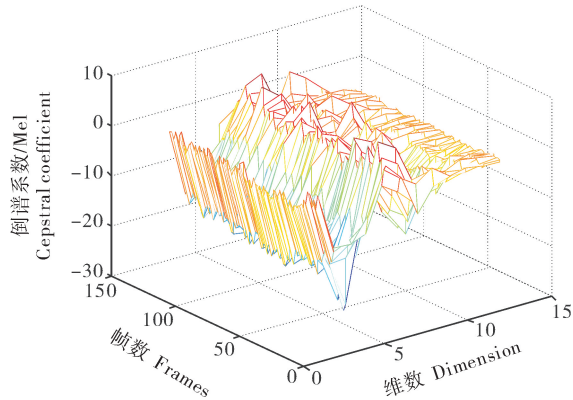


图 4 标准梅尔倒谱系数

Fig.4 Standard Mel-frequency cepstral coefficients

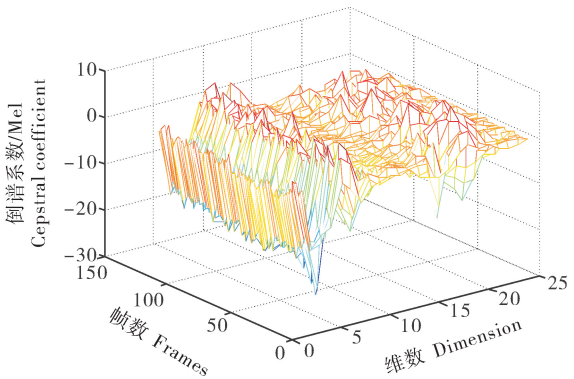


图 5 改进的梅尔倒谱系数

Fig.5 Improved Mel-frequency cepstral coefficients

3 猪咳嗽声的识别

3.1 VQ 模型参数库的建立

首先解决的问题是如何获取初始码书,其主要途径是随机选取法和分裂法,分裂法克服了随机选取法难以得到理想的初始码书的缺点,具有计算效率高、初始码书性能好的优点,以此码书设计的矢量量化器有更好的识别性能。本研究选用分裂法,先设定初始码书 $J=1$,即只包含 1 个码字,计算训练序列的形心,以此形心作为初始码字,按照 $Y_{J(2)} = Y_{J(1)} \pm \epsilon$ 分裂成 2 个元素,训练得到了第 2 个码字,继续分别将 2 个码字进行分裂,获得包含 4 个码字的码书。这个过程不断地重复,直到形成所需的含 J 个码字的码书。

VQ 模型是依据给定的训练集 S 和初始码书,找到平均失真度最小的码书和空间划分,即获得失真率最小的最佳码本。本试验选择 LBG 算法获取最佳码本,LBG 算法以一定的迭代算法无限逼近最终得到最优码本。具体过程如下:

(1)初始化:给出全部参考矢量集 S 和初始码本,设定最小迭代误差 δ ,算法最大迭代次数 L ,设置总失真 $D^{(0)} = \infty$,初始迭代次数 $m=1$,最大迭代次数为 L 。

(2)迭代:根据最邻近准则将矢量集 S 分成 J 个子集,并计算总失真 D^m ,此时形成新的以子集为质心的 J 个码字。如果相对失真改进量 $\delta^m = |D^{m-1} - D^m| / D^m$ 小于最小迭代误差 δ ,则停止迭代,否则继续迭代直到达到最大迭代次数为止。

将 400 个猪咳嗽样本训练获得最佳码本,在识别阶段,采用欧式距离算法计算测试样本与码本中每个码字的距离 d ,公式如下:

$$d(X,Y)=\frac{1}{K}\sum_{i=1}^K(x_i-y_i)^2$$

(5)

式中 X 为测试样本码字的 K 维矢量 $X=(x_1,x_2,\cdots,x_i)$, Y 为码本中码字的 K 维矢量 $Y=(y_1,y_2,\cdots,y_i)$ 。计算结果与设定的最小距离比较,小于最小设定距离则计数器加 1,否则计数器不变,最终计数结果如大于设定的阈值则评估为咳嗽声,否则为非咳嗽声。

3.2 猪咳嗽识别结果与分析

本研究选择 4 个指标描述 VQ 模型识别性能,分别为识别率(正确识别咳嗽音数目占总咳嗽音的百分比)、误判率(被误识别为咳嗽音的非咳嗽音数占总非咳嗽音的百分比)和综合识别率(正确识别的咳嗽音和非咳嗽音总数占测试样本总数的百分比),校验正确率(正确识别的训练样本中的咳嗽音占训练样本总数的百分比)。在样本的容量和划分上,参考文献[3]和[5],共选择 900 个样本,其中 400 个咳嗽样本作为训练样本,200 个咳嗽样本和 100 个非咳嗽声样本作为测试样本,分别提取标准 MFCC 参数和改进的 MFCC 参数作为 VQ 模型的输入,试验识别结果如表 1 所示。

表 1 表明,改进的 MFCC 参数作为识别系统的输入时,识别性能得到了较好的改善,识别率、误判率和总识别率分别由 88%、14%和 87.3%提高到

表 1 猪咳嗽声识别结果
Table 1 The results of pig cough recognition

特征参数 Characteristic parameter	识别率 Recognition rate	误判率 Misjudgment rate	校验正确率 Check correct rate	综合识别率 Comprehensive recognition rate
标准 MFCC Standard MFCC	88	14	91	87.3
差分 MFCC Difference MFCC	91	12	92	90.0

91%、12%和 90%。这是因为改进的 MFCC 参数不仅包含声音的静态特性,而且包含声音的动态特性,对咳嗽声的特征描述更加全面,能够提高系统的识别性能。

4 讨 论

本研究基于矢量量化(VQ)技术对猪咳嗽音进行识别,应用谱减法、端点检测和分帧加窗等技术将采集的咳嗽信号进行预处理,分别提取标准 MFCC 和改进的 MFCC 特征参数构建猪咳嗽声识别模型,拓展猪咳嗽声检测的思路,为猪的呼吸道疾病的实时监测和预警提供理论依据。矢量量化模型对咳嗽声识别具有计算速度快、同时可以保证相对较高的识别率,在应用到实时监测时有更好的效果。谱减法是降噪的常规方法,能够有效减少咳嗽信号中的噪声干扰,提高系统识别率。端点检测能够去掉咳嗽数据中的静音部分减少识别系统工作量,提高工作效率。而分帧加窗则是处理猪咳嗽声必不可少的部分,这是由于猪的咳嗽声作为整体是离散性信号,无法进行分析,但其短时是平稳的,一般选择 10~30 ms 为一帧进行分析。直接使用猪场采集的猪咳嗽声信号对矢量量化模型进行训练和测试,提高了

模型对实际复杂环境的适应能力,为模型的移植奠定了基础。测试结果表明,改进的 MFCC 参数比标准 MFCC 参数能更好地反映声音的特征,具有更好的识别能力,综合识别率达到了 90%,可用于构建咳嗽声识别系统。本次试验针对 75 kg 左右长白猪取得了较好的识别效果,而不同品种和日龄的猪的咳嗽音特征具有很大的差异性,会导致构建模型参数的改变,影响识别系统的稳定性和识别率,如何扩展模型的适用范围是进一步研究的重点内容。

参 考 文 献

[1] 赵丽杰,孙加道.以相伴症状对引发猪咳嗽的几种疾病的鉴别与诊断[J].养殖技术顾问,2014(10):197.

[2] 朱伟兴,赵凡.声音匹配识别算法的研究与实践[J].中国传媒大学学报(自然科学版),2007,14(1):20-26.

[3] 宣传忠,武佩,张丽娜,等.羊咳嗽声的特征参数提取与识别方法[J].农业工程学报,2016,47(3):342-348.

[4] MEEN G H,SHELLEKENS M A,SLEGGERS M H M,et al. Sound analysis in dairy cattle vocalisation as a potential welfare monitor[J]. Computers and electronics in agriculture,2015, 118:111-115.

[5] 徐亚妮,沈明霞,闫丽,等.待产梅山母猪咳嗽声识别算法的研究[J].南京农业大学学报,2016,39(4):681-687.

- [6] MATO S, BIRRING S, PAVORD D, et al. Detection of cough signals in continuous audio recordings using hidden Markov models[J]. IEEE transactions on biomedical engineering, 2006, 53(6): 1078-1083.
- [7] AMRULLOH Y, ABEYRATNE U, SWARNKAR V, et al. Automatic cough segmentation from non-contact sound recordings in pediatric wards[J]. Biomedical signal processing and control, 2015, 21(8): 126-136.
- [8] DRUGAN T. Using mutual information in supervised temporal event detection: application to cough detection[J]. Biomedical signal processing and control, 2014, 10(3): 50-57.
- [9] TUZUNER A, DEMIRCI S, BILGIN G, et al. Voice assessment after treatment of subacute and chronic cough with inhaled steroids[J]. Journal of voice, 2015, 29(4): 484-489.
- [10] HIRTUM A, BERCJMANS D. Automated recognition of spontaneous versus voluntary cough [J]. Medical engineering & physics, 2002, 24: 541-545.
- [11] SILVAA M, EXADAKTYLOS V, FARRARI S, et al. The influence of respiratory disease on the energy envelope dynamics of pig cough sounds[J]. Computers and electronics in agriculture, 2009, 6(8): 80-85.
- [12] FARRARI S, SILVAA M, GUARINO M, et al. Cough sound analysis to identify respiratory infection in pigs[J]. Computers and electronics in agriculture, 2008, 64(2): 318-325.
- [13] BAHR C, FONTANA L, KASHIHA M, et al. Discerning pig screams in production environments[J]. PLoS ONE, 2015, 10(4): 123-137.
- [14] HIRTUM A, BERCJMANS D. Considering the influence of artificial environmental noise to study cough time-frequency features[J]. Journal of sound and vibration, 2003, 266(3): 667-675.
- [15] EXADAKTYLOS V, SILVA M, AERTS J, et al. Real-time recognition of sick pig cough sounds[J]. Computers and electronics in agriculture, 2008, 63(2): 207-214.
- [16] HIRTUM A, BERCJMANS D. Fuzzy approach for improved recognition of citric acid induced piglet coughing from continuous registration[J]. Journal of sound and vibration, 2003, 266(3): 677-686.
- [17] FERRARI S, PICCININI R, SILVA M, et al. Cough sound description in relation to respiratory diseases in dairy calves[J]. Preventive veterinary medicine, 2010, 96(3): 276-280.
- [18] GUARINO M, JANS P, COSTA A, et al. Field test of algorithm for automatic cough detection in pig houses[J]. Computers and electronics in agriculture, 2008, 62: 22-28.
- [19] 朱春媚, 黎萍. 基于子带能量变换改进 MFCC 的咳嗽识别[J]. 计算机工程与应用, 2016, 52(11): 148-151.
- [20] CHIA O, HARIHARAN M, YAACOB S, et al. Classification of speech dysfluencies with MFCC and LPCC features[J]. Expert systems with application, 2012, 39(2): 50-57.

Recognition of pig cough sound based on vector quantization

GONG Yongjie¹ LI Xuan^{1,2} GAO Yun^{1,2} LEI Minggang^{2,3}
LIU Wanghong^{2,3} YANG Zhuan¹

1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China;

2. The Cooperative Innovation Center for Sustainable Pig Production, Wuhan 430070, China;

3. College of Animal Science and Technology, College of Animal Medicine,
Huazhong Agricultural University, Wuhan 430070, China

Abstract Respiratory tract diseases, frequently outbreaking in pig barns, bring economy lost in modern pig production. In this paper, a method was proposed to predict and monitor respiratory tract diseases by recognition of the cough sound of pigs. The sounds of pig were processed by the spectral subtraction denoising and the endpoint detection before being made into the improved Mel-frequency cepstral coefficients (MFCC), which are feature vectors input into the vector quantization matching algorithm (VQ algorithm). Thus a pig cough recognition model based on the VQ algorithm was developed to recognize the sound of cough. The results showed that the recognition rate, the false positive rate and the comprehensive recognition rate of the model based on the improved MFCC were 91%, 12% and 90.0%, respectively, which were better than 88%, 14% and 87.3% of the standard MFCC, suggesting the feasibility of the early warning of the respiratory tract disease of pigs in the future.

Keywords vector quantization (VQ); pig; cough; identification; Mel-frequency cepstral coefficients (MFCC)

(责任编辑:边书京)