

应用科技  
*Applied Science and Technology*  
ISSN 1009-671X, CN 23-1191/U

## 《应用科技》网络首发论文

题目: 育肥猪舍中的猪叫声端点检测算法  
作者: 熊梓奥, 苍岩  
收稿日期: 2020-04-23  
网络首发日期: 2020-10-20  
引用格式: 熊梓奥, 苍岩. 育肥猪舍中的猪叫声端点检测算法. 应用科技.  
<https://kns.cnki.net/kcms/detail/23.1191.U.20201020.1436.004.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 育肥猪舍中的猪叫声端点检测算法

熊梓奥, 苍岩

哈尔滨工程大学 信息与通信工程学院, 黑龙江 哈尔滨 150001

**摘要:** 针对在猪舍声监测系统中采集的声音中包含许多冗余信息, 整个系统的待检测数据量大, 导致检测效率低下这一问题, 提出了一种基于长短时记忆 (LSTM) 网络单元的端点检测算法, 结合猪叫声信号的梅尔频率倒谱系数 (MFCC) 与对数能量特征进行训练和检测。仿真实验采用育肥舍猪场的猪叫声和风扇噪声作为数据集, 使用双阈值检测和支持向量机 (SVM) 检测算法进行对比试验, 同时探究了提出算法在不同信噪比环境下的性能。实验结果表明, 在育肥猪舍的风扇噪声环境下, LSTM 检测算法具有良好的检测性能和鲁棒性。

**关键词:** 端点检测; LSTM; MFCC; 短时能量; 短时过零率; SVM; 猪叫声; 风扇噪声

**中图分类号:** TP391.4

**文献标志码:** A

## The voice activity detection for oink in fattening pig houses

XIONG Ziao, CANG Yan

College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

**Abstract:** In a pig cough detection system, the sound in the pig house needs to be collected for a long time. The collected sounds contain a lot of redundant information and the data to be detected is large, resulting in inefficiency. To solve this problem, a voice activity detection algorithm is proposed based on long short term memory (LSTM) network units. The Mel-frequency cepstrum coefficient (MFCC) and logarithmic energy of oink signals were extracted for training and detecting. In the simulation experiments, the oinks and fan noise in the fattening pig house are used as data sets, the double thresholds detection and support vector machines (SVM) detection algorithms are used for comparison test. Meanwhile, the performance of the proposed algorithm in different signal-to-noise ratio environment is explored. The results showed that LSTM detection algorithm has good performance and robustness in the fan noise environment of a fattening pig house.

**Keywords:** voice activity detection; LSTM; MFCC; short-term energy; short-term zero crossing rate; SVM; oink; fan noise

2019 年, 由于非洲猪瘟的影响, 我国生猪市场遭受冲击, 各地养猪场大量生猪病死。在养殖户经济损失巨大的同时, 也造成了猪肉价格飙升的情况<sup>[1]</sup>。猪在患上非洲猪瘟后, 最显著的病状就是咳嗽<sup>[2]</sup>。为预防相关呼吸道传染疾病, 提高人工诊断效率, 可通过监测猪舍中猪的咳嗽情况, 报告饲养员进行预警。猪咳嗽声监测系统是一个端对端系统, 需要对猪舍的声音长时间采集, 再来识别其中的猪咳嗽声。由于长时间采集的声音数据量大, 包含没有猪叫声的部分, 所以在对猪咳嗽声识别之前, 需要对采集得到的音频数据进行端点检测。在冗长的音频中确定猪叫声的起

点与终点, 提取出猪叫声, 从而提高系统的检测效率。同时, 端点检测算法的检测效果也会直接影响监测系统后续咳嗽声识别的准确率。因此, 一个有效且鲁棒的端点检测算法是整个猪咳嗽声监测系统的关键之一。

现阶段端点检测算法大致可分为两类: 第一类是基于阈值比较的方法。这类方法提取样本每帧的特征参数, 将其与人为设定的阈值进行比较, 然后判定得到语音帧, 其最为经典的是基于短时能量和短时过零率的双阈值端点检测算法<sup>[3]</sup>。它使用短时能量与短时过零率作为特征参数, 具有计算量小, 检测速度快的特点, 但在低信噪比条件下算法的性能急剧下降。由于语音的特征参数易受背景噪声影响, 有些研究者便尝试找到更抗噪的特征, 文献[4]将子带能量与子带谱

收稿日期: 2020-04-23.

作者简介: 熊梓奥, 男, 硕士研究生;

苍岩, 女, 讲师, 博士。

通信作者: 苍岩, E-mail: cangyan@hrbeu.edu.cn.

熵相结合, 提出一种新型语音特征参数——子带能量熵比。除此之外, 文献[5]引入了特征组合的方法, 以改善单一特征易受噪声影响的问题。第二类是基于模型匹配的方法。这类方法通过样本构建语音模型, 再与测试样本进行匹配。在低信噪比情况下, 第二类方法比第一类方法的检测结果更准确, 但是它需要大量数据进行训练, 而且算法相对复杂, 计算量大。早期有研究者使用隐马尔可夫模型 (hidden Markov model, HMM) 算法<sup>[6]</sup>、支持向量机算法 (SVM) <sup>[7]</sup>以及单一的神经网络作为端点检测的模型。文献[8]对单一神经网络的检测效果做了比较。随着深度学习研究的快速发展, 对于端点检测算法的探究从单一神经网络, 转向多算法融合。文献[9]将深度神经网络 (deep neural networks, DNN) 与维特比算法结合, 进一步提高检测的准确率。文献[10]提出了一种增强统计噪声抑制算法, 并将其作为一个模块, 应用于在卷积神经网络前。除此之外, 文献[11]搭建了复杂的神经网络模型, 将卷积神经网络 (convolutional neural networks, CNN) 与循环神经网络 (recurrent neural network, RNN) 结合, 进行特征学习。模型输入语音样本原始波形, 实现了端到端的端点检测系统。

综上, 本文以生猪养殖场育肥舍的应用场景为背景, 提出一种基于长短时记忆 (LSTM) 网络单元的端点检测算法。算法首先通过分析猪舍中的猪叫声信号, 提取其梅尔频率倒谱系数 (MFCC) 与对数能量特征作为模型的输入, 再搭建以 LSTM 为主体的神经网络对猪叫声信号进行端点检测, 并探究了算法在猪舍风扇噪声下的鲁棒性。

## 1 算法总述

本文端点检测方案本质上是对猪叫声音频样本进行逐帧判断, 确定每一帧是猪叫声还是非猪叫声, 从而得到猪叫声的起点和终点。相比于人类, 猪叫声种类较少, 主要有正常的哼叫声、打架的嚎叫声、患病的咳嗽声, 而育肥猪舍的背景噪声也相对单一, 大部分情况为风扇声、水声以及猪撞击猪栏的金属声。因此, 可以针对猪舍的特点设计整个检测方案。由图 1 所示, 猪叫声端点检测方案由预处理、特征提取、模型训练、模型检测 4 部分组成。

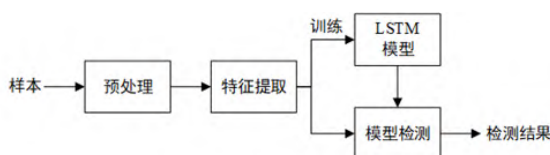


图 1 猪叫声端点检测方案

预处理阶段: 将样本按比例分为训练集、验证集、

测试集, 并对每个样本进行预加重、分帧、加窗处理。

特征提取阶段: 对预处理后的训练集、验证集中样本提取可以表征样本的特征参数 MFCC 和对数能量的组合, 减少原始样本的冗余信息, 提高训练效率。

模型训练阶段: 根据 LSTM 可以对样本时间信息学习的特点, 搭建以 LSTM 为主的神经网络, 将提取的特征参数输入网络中训练, 并保存训练完成的网络。

模型检测阶段: 对测试集样本进行特征提取, 输入训练好的神经网络中, 输出结果, 最后判定得到猪叫声的起点和终点。

## 2 基于 LSTM 的猪叫声端点检测算法

### 2.1 预加重

语音信号有着低频信噪比大, 高频信噪比小的特点。为了改善高频信噪比, 需要对猪叫声样本进行预加重处理, 提升其高频部分, 从而增加高频分辨率。将样本信号通过一个高通滤波器即可实现预加重, 高通滤波器的传递函数如下,

$$H(z) = 1 - \alpha z^{-1}$$

假设第  $n$  时刻的语音采样值为  $x(n)$ , 经过预加重处理后的信号如下,

$$x_p(n) = x(n) - \alpha x(n-1)$$

式中  $\alpha$  为预加重系数, 通常取 0.97。

### 2.2 加窗分帧

语音信号是一种时变信号, 但它具有短时平稳特性。在一个较短的时间内, 语音信号特性基本保持不变即相对稳定, 可以将其视为一个准稳态过程。MFCC 特征参数提取采用短时谱分析, 因此对猪叫声信号进行分帧。分帧时相邻帧之间应有部分交叠, 以保证帧与帧之间的连贯性。由于分帧后语音帧的两端急剧变化, 会导致频域中帧与帧之间过渡时信号发生丢失。将每个语音帧乘以 Hamming 窗, 使两端平滑过渡到零, 从而减小语音帧的截断效应。Hamming 窗函数如下,

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N$$

式中  $N$  为窗的长度。

### 2.3 特征提取

MFCC 是由 Davis 和 Mermelstein<sup>[12]</sup>在 1980 年提出。研究表明, 人耳对低频信号更加敏感。当频率小于 1 kHz 时, 频率与人耳感知能力呈线性关系; 当频率大于 1 kHz 时, 其呈对数关系。梅尔 (Mel) 频率就是将实际频率由线性转换为非线性的方法, 转换公

式如下,

$$M(f) = 1125 \times \ln(1 + \frac{f}{700})$$

式中  $f$  为实际频率。

作为语音信号常见的特征之一, MFCC 模拟了人的听觉特性, 适合于语音信号处理的相关工作中。此外, 能量也是有效衡量语音和非语音的有效特征。因此, 本文选择 MFCC 加对数能量作为表征猪叫声样本的特征参数, 特征提取步骤如图 2 所示。

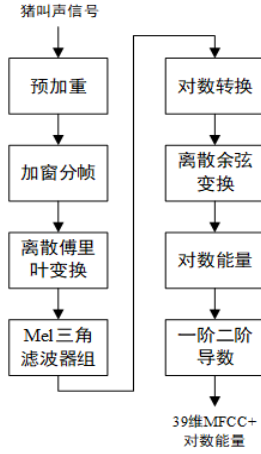


图2 MFCC加对数能量特征提取步骤

1) 将之前预加重、分帧加窗后的猪叫声信号, 按帧进行离散傅里叶变换 (discrete Fourier transform, DFT) 得到样本的频谱,

$$Y_i(k) = \sum_{n=1}^N y_i(n)w(n)e^{-\frac{j2\pi kn}{N}}, 1 \leq k \leq K$$

计算功率谱为

$$P_i(k) = \frac{1}{N} |Y_i(k)|^2$$

式中:  $y_i(n)$  为预加重、分帧后的信号;  $k$  为傅里叶变换的点数;  $n$  表示第  $i$  帧中的第  $n$  个采样点。

2) 将上面的功率谱通过 Mel 滤波器组, 得到 Mel 频谱, 公式如下,

$$S_i(m) = \sum_{k=1}^N P_i(k) H_m(k), 1 \leq m \leq M$$

式中:  $M$  为滤波器组中三角滤波器的个数;  $H_m(k)$  为 Mel 滤波器的频率响应。

3) 计算每个滤波器组输出的对数能量, 再通过离散余弦变换 (discrete Cosine transform, DCT) 得到 MFCC 系数  $C$ , 公式如下,

$$C_{MFCC-i} = \sum_{m=1}^M \ln(S_i(m)) \cos(\frac{\pi i(m-0.5)}{M}), 1 \leq i \leq I$$

式中:  $M$  为滤波器组中三角滤波器的个数;  $I$  为 MFCC 维数。

4) 取 DCT 后的第 2 个到第 13 个系数, 组成 12 维 MFCC, 再与这一帧的对数能量组合, 最终得到这帧语音的 13 维特征

5) 通常情况, 由于 MFCC 与对数能量的组合只能反映猪叫声的静态特征, 为了提高模型的检测效果, 在 13 维特征的基础上计算得到其一阶和二阶导数, 组成 39 维特征阵。

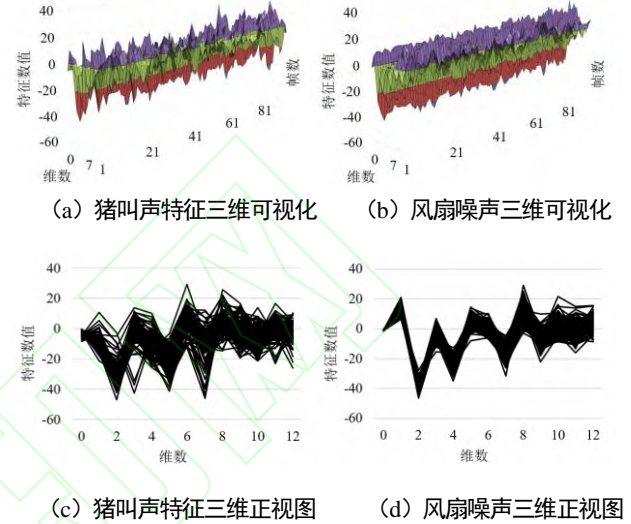


图3 猪叫声与风扇噪声特征对比

图 3 (a)、(b) 分别为猪叫声与风扇噪声 MFCC 加对数能量前 13 维特征矩阵的三维特征图, 图 3 (c)、(d) 分别对应的三维特征图的正视图。由于风扇噪声波形幅值随时间变化较小, 相对猪叫声更平稳。从图 3 (c)、(d) 中可以看出, 每帧的风扇噪声相比于猪叫声, 特征值方差更小, 特征曲线重合度更高。

## 2.4 检测模型

长短时记忆 (LSTM) 网络单元在 1997 年由 Hochreiter 与 Schmidhuber<sup>[13]</sup>提出, 它是一种特殊的 RNN 结构。它能够对输入的长时依赖关系进行建模, 同时一定程度上解决了 RNN 在较长的时间序列上反向传播时带来的梯度消失问题。

在 LSTM 的模块中, 包含一个输入门, 一个输出门和一个遗忘门, 通过 3 个门的协作来学习权重, 达到能够存储长期信息的效果。LSTM 网络单元的结构图如图 4 所示。

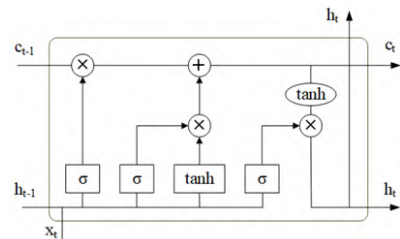


图4 MFCC+对数能量特征提取步骤



LSTM 内部相关计算公式为

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$

$$c_t = f_t \times c_{t-1} + i_t \times \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$

$$h_t = o_t \times \sigma_h(c_t)$$

式中:  $h_t$  是隐藏状态;  $c_t$  是单元状态;  $W$ 、 $U$ 、 $b$  为训练中需要学习的参数;  $f_t$  为遗忘门、 $i_t$  为输入门、 $o_t$  为输出门, 三者的输入均为  $h_{t-1}$  和  $x_t$ , 经过激活函数, 将值缩放至 0 和 1 之间。当遗忘门的值取 0 时, 单元状态  $c_t$  的值也变为 0, 相当于遗忘掉上一时刻的状态, 只关注此时此刻的输入。输入门决定是否接收此时此刻的输入, 输出门决定是否输出单元状态。

图 5 为 LSTM 检测模型的网络结构。其中, Linear 表示线性层, ReLU 和 Sigmoid 为激活函数。

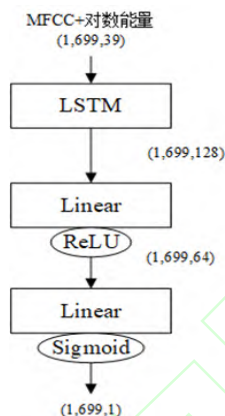


图 5 LSTM 检测模型结构

## 2.5 算法实现步骤

1) 对训练集样本进行预加重、分帧和加窗处理。其中, 预加重系数  $\alpha$  为 0.97, 帧长为 25 ms, 帧移为 10 ms, 窗函数为 Hamming 窗。将时长为 7 s 的音频样本转换成 699 个语音帧, 得到 (1, 699) 的帧

2) 将人工标记的猪叫声起止时间点, 转换为 (1, 699) 的帧标签, 语音帧的值为 1, 非语音帧的值为 0。

3) 计算训练集样本的 39 维 MFCC 加对数能量特征。得到尺寸为 (39, 699) 的特征矩阵。

4) 将每个特征矩阵转换为 (1, 699, 39) 的张量, 送入 LSTM 模型中训练, 模型的输出为 (1, 699, 1) 的张量。

5) 对模型的输出进行判定, 大于阈值判定为语音帧, 小于阈值判定为非语音帧, 得到检测结果。

## 3 实验结果及分析

### 3.1 实验过程

#### 3.1.1 实验数据

本文实验采用的猪叫声数据来自吉林省某生猪养殖场育肥舍, 由工作人员使用手机录制。猪舍内有用于控制温度的风扇, 当室内温度达到一定程度便自动开启。样本的采集时间是冬季, 风扇转动的频率较低。相比于夏季, 采集数据没有被风扇噪声干扰的情况更多。数据经过截取处理, 去除被风扇声干扰的数据, 组成由 200 个时长为 7 s 的样本构成的数据集。样本的采样频率为 16 kHz, 采样精度为 16 bit。使用 Audacity 软件标记样本中猪叫声的起止点, 并保存为 json 文件, 作为语音标签。最后把实验数据按 8:1:1 的比例分为训练集、验证集、测试集。此外, 还额外录制了不含猪叫声的风扇噪声, 用于算法的鲁棒性验证。

#### 3.1.2 实验内容

**实验一** 对于端点检测算法, 根本目标是要完整检测出猪叫声。模型在输出时由 Sigmoid 函数将输出值映射到 (0, 1) 之间, 因此需要进行判别, 使其值转换为代表语音帧的 1, 非语音帧的 0。如大于阈值 0.5 的概率值, 则判别为语音帧, 反之则判别为非语音帧。这里阈值可以视为一种容忍度, 即在保证准确率的前提下, 接受一定的误检, 而避免漏检, 保证检测猪叫声的完整性。为探究判别阈值大小对模型准确率的影响, 本实验多次改变阈值的大小, 将阈值设置在 0.3~0.7, 步长为 0.1。先使用训练集和验证集训练模型, 再使用测试集对训练完成的模型进行评估。

**实验二** 使用基于短时能量和短时过零率的双阈值端点检测算法、基于 SVM 的端点检测算法作为对照实验。双阈值算法先提取样本的短时能量和短时过零率, 分别设置 2 个特征阈值, 再沿着时间方向将 2 种特征和阈值做比较, 得到检测结果。阈值的计算采用自适应方法, 对不同的样本选取不同的阈值, 以确保算法的检测效果。SVM 算法通过提取样本的多种时频特征构成组合特征, 来训练 SVM 模型, 再使用训练好的模型对测试样本的每一帧进行语音、非语音的二分类, 最终得到整个样本的检测结果。SVM 的核函数采用线性核, 经过多次实验测试, 惩罚因子  $C$  选取最优值 1, 实验结果见图 6。

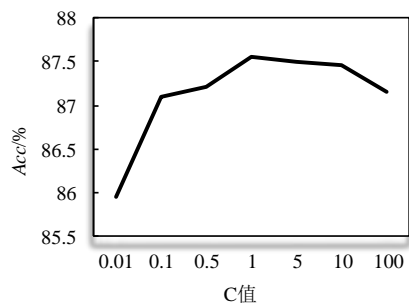
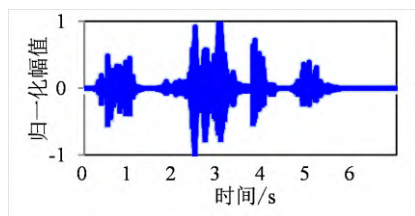
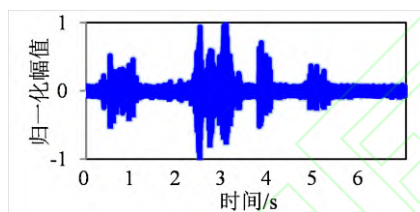


图6 SVM在不同C情况下的检测结果

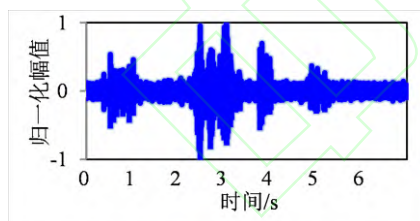
为验证本文算法的鲁棒性，本实验对比3种算法在加入不同信噪比噪声的数据集下的检测效果。将风扇噪声分别以10、5、0、-5、-10 dB的信噪比，加入到数据集中，再使用加噪的数据集训练模型。图7显示了一个样本的原始波形图及其加噪后的波形图。



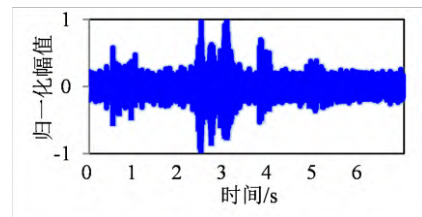
(a) 未加噪



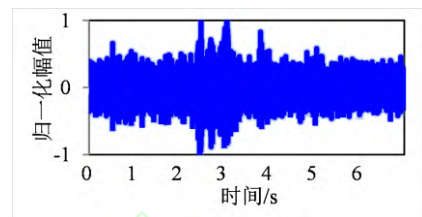
(b) SNR=10 dB



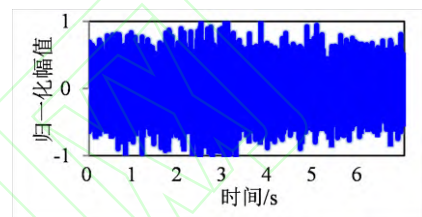
(c) SNR=5 dB



(d) SNR=0 dB



(e) SNR=-5 dB



(f) SNR=-10 dB

图7 样本的原始波形图及其加噪后的波形

### 3.2 实验评价指标

为客观评价算法的性能，实验需要比较算法的检测结果与人工标记的起止点，并从以下3个方面对算法进行评估，

准确率 (Accuracy):

$$A = N_T / N_s \times 100\%$$

漏检率 (False Rejection Rate):

$$R_{FR} = N_{FR} / N_s \times 100\%$$

误检率 (False Acceptance Rate):

$$R_{FA} = N_{FA} / N_s \times 100\%$$

式中： $N_s$ 为总帧数； $N_T$ 为检测正确的帧数； $N_{FR}$ 为检测为非语音帧的语音帧数量； $N_{FA}$ 为检测为语音帧的非语音帧的数量。

### 3.3 实验结果分析

#### 3.3.1 实验一

表1显示了LSTM模型在不同阈值下，在各个数据集上的表现。

表1 LSTM模型在不同判定阈值情况下的准确率

阈值	未加噪	SNR=10	SN=5	SN=0	SN=-5	SNR=-10
0.3	87.68	88.81	87.40	82.22	79.89	73.62
0.4	87.78	89.15	87.67	85.04	80.19	75.14
0.5	87.86	90.82	88.06	84.46	78.48	71.90
0.6	87.79	91.22	88.49	84.09	77.89	72.22
0.7	87.60	91.11	88.18	83.76	78.35	70.82

可以看出，未加噪时，阈值对准确率的影响相对较小；在信噪比较大（10、5 dB）的情况下，最佳阈

值等于 0.6; 信噪比较小 (0, -5, -10 dB) 时, 最佳阈值等于 0.4。在测试样本未加噪的情况下, 样本中的猪叫声容易被模型检测出来, 因此语音帧对应的输出更接近标签值 1, 准确率没有因为阈值的变化而大幅改变。当测试样本加噪后, 在信噪比较大的情况下, 猪叫声的幅值远大于风扇噪声的幅值, 噪声对检测结果有一定干扰, 但干扰相对较小, 漏检情况较少。此时, 非语音帧对应的输出值会大于 0, 但又不会大很多, 因此较大的阈值会将这类输出值判定为非语音帧, 降低了误检率, 使得准确率提高。在信噪比较小的情

况下, 风扇噪声会淹没部分猪叫声, 检测结果会受到噪声的严重干扰, 语音帧的输出值会更接近非语音帧输出值。此时, 较小的阈值会降低检测标准, 增加误检率, 但保证了更多的猪叫声不被噪声影响而漏检, 反而增加了整体的准确率。

### 3.3.2 实验二

表 2 显示了 3 种算法在添加了不同信噪比风扇噪声的数据集上的表现, 图 8 显示了 3 种算法在不同信噪比风扇噪声下的准确率。由图 8 可以看出, 3 种算法的准确率会随着信噪比的减小而降低。

表 2 3 种算法在不同信噪比的风扇噪声下的检测结果

%

SNR/dB	双阈值检测算法			SVM 检测算法			LSTM 检测算法		
	A	$R_{FR}$	$R_{FA}$	A	$R_{FR}$	$R_{FA}$	A	$R_{FR}$	$R_{FA}$
未加噪	82.29	17.46	0.25	87.55	0.67	11.78	87.86	3.33	8.81
10	82.74	16.94	0.32	85.20	3.10	11.70	91.22	3.72	5.06
5	84.15	14.60	1.25	74.43	6.72	18.85	88.49	3.23	8.28
0	78.03	19.76	2.21	67.64	7.54	24.82	85.04	6.37	8.59
-5	60.20	4.83	34.97	59.56	5.99	34.45	80.19	9.96	9.85
-10	54.20	2.44	43.36	50.02	10.84	39.14	75.14	14.34	10.52

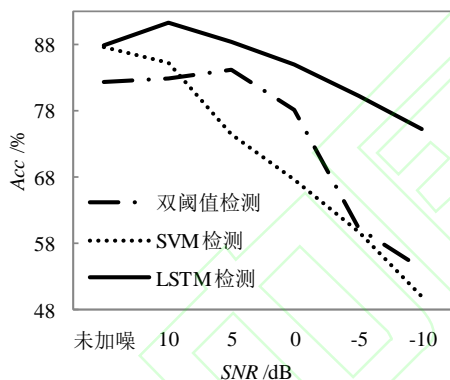


图 8 3 种算法在不同信噪比的风扇噪声下的准确率

双阈值检测算法在高信噪比 (10、5 B) 的情况下检测效果较理想。但当风扇噪声强度增加, 信噪比小于 0 时, 部分猪叫声被风扇噪声所淹没, 猪叫声的短时能量与短时过零率便没有了区分度。这使得算法几乎将整个样本都判定为猪叫声, 更少的猪叫声被漏检, 更多的风扇噪声被判定为猪叫声, 造成了漏检率  $R_{FR}$  大幅降低与误检率  $R_{FA}$  的激增。图 9 为图 7 (d) 样本的双阈值检测结果, 从图中可以看出, 由于噪声影响, 双阈值检测算法将部分噪声误判为猪叫声。

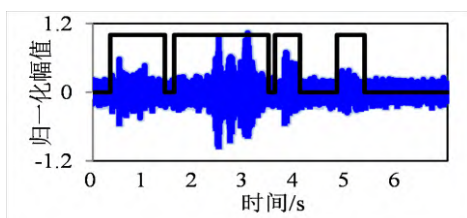


图 9 SNR=0 dB 时双阈值检测结果

SVM 算法在信噪比降低的过程中, 整体的检测效果变差, A 的变化几乎是呈线性降低,  $R_{FR}$  和  $R_{FA}$  一直增加。这说明了随着噪声强度的增加, 猪叫声的时频特征与风扇噪声的时频特征区分度越来越低, SVM 模型很难对猪叫声和风扇噪声进行正确分类。图 10 为图 7 (d) 样本的 SVM 检测结果, 在信噪比为 0 dB 的情况下, SVM 模型已无法对样本正确检测。

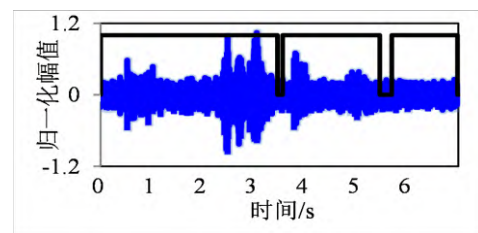


图 10 SNR=0 dB 时 SVM 检测结果

LSTM 算法相比于其他 2 种算法, A、 $R_{FR}$ 、 $R_{FA}$  三者的变化幅度更小, 检测结果更加稳定。在信噪比高的情况下, 算法检测结果优于其他, 即便在较低的信噪比 (-5、-10 dB) 下也仍有一个理想的检测效果。这说明 LSTM 模型通过训练样本学习到了猪叫声的特征, 即使在低信噪比时, 也能从含噪样本中判定得到猪叫声。因此也证明了 LSTM 算法在猪舍的风扇噪声下有着更好的抗噪鲁棒性。图 11 为图 7 (d) 样本的 LSTM 检测结果, 由图可以看出 LSTM 模型在较低信噪比时仍然能检测出猪叫声。

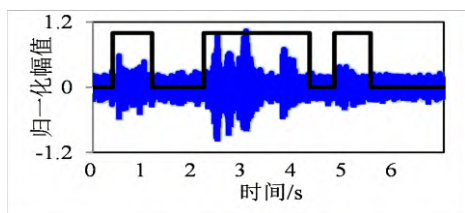


图 11 SNR=0 dB 时 LSTM 检测结果

## 4 结论

1) 传统端点检测算法在高信噪比情况下, 检测效果良好, 但随着噪声强度的增加, 算法的准确率也随之大幅降低。本文提出了一种基于 LSTM 网络单元的端点检测算法, 针对风扇噪声的环境, 通过大量猪叫声数据训练得到更精准的端点检测模型。

2) 本文进行了 2 组实验, 实验一为测试得到 LSTM 端点检测模型的最佳判决阈值, 实验二在这个阈值的基础上与双阈值端点检测和 SVM 端点检测进行比较。仿真实验结果表明, 在猪舍中不同信噪比的风扇噪声下, 相比于传统端点检测算法, 提出算法的检测效果稳定, 鲁棒性好。

本文所提出的 LSTM 模型较小, 可以通过搭建较复杂的模型, 进一步提高检测的效果。未来也可以在数据集中增加其他猪场的猪叫声, 以提高模型的泛化能力。

## 参考文献:

- [1] 王宏栋, 王玉赛, 李冬刚, 等. 非洲猪瘟的诊断及防控措施[J]. 中国猪业, 2019, 14(8): 92-94, 98.
- [2] 李影. 几种引发猪咳嗽疾病的临床诊断[J]. 吉林畜牧兽医, 2018, 39(7): 14, 17.
- [3] 王满洪, 张二华, 王明合. 基于双门限算法的端点检测改进研究[J]. 计算机与数字工程, 2017, 45(11): 2223-2228.
- [4] ZHANG Yi, WANG Kejia, YAN Bo. Speech endpoint detection algorithm with low signal-to-noise based on improved conventional spectral entropy[C]//2016 12th World Congress on Intelligent Control and Automation. Guilin, China, 2016, DOI: 10.1109/WCICA.2016.7578597.
- [5] ZAW T H, WAR N. The combination of spectral entropy, zero crossing rate, short time energy and linear prediction error for voice activity detection[C]//2017 20th International Conference of Computer and Information Technology. Dhaka, Bangladesh, 2017, DOI: 10.1109/ICCITECHN.2017.8281794.
- [6] 朱杰, 韦晓东. 噪声环境中基于 HMM 模型的语音信号端点检测方法[J]. 上海交通大学学报, 1998, 32(10): 14-16.
- [7] 董恩清, 赵鹤鸣, 周亚同, 等. 支持向量机在语音激活检测中的应用研究[J]. 通信学报, 2003, 24(3): 70-77.
- [8] TONG Sibo, GU Hao, YU Kai. A comparative study of robustness of deep learning approaches for VAD[C]//2016 IEEE International Conference on Acoustics, Speech and Signal Processing. Shanghai, China, 2016, DOI: 10.1109/ICASSP.2016.7472768.
- [9] BAI Liang, ZHANG Zhen, HU Jun. Voice activity detection based on deep neural networks and Viterbi[C]//The 2017 2nd International Seminar on Advances in Materials Science and Engineering. Singapore, 2017, DOI: 10.1088/1757-899X/231/1/012042.
- [10] OBUCHI Y. Framewise speech-nonspeech classification by neural networks for voice activity detection with statistical noise suppression[C]//2016 IEEE International Conference on Acoustics, Speech and Signal Processing. Shanghai, China, 2016, DOI: 10.1109/ICASSP.2016.7472772.
- [11] ZAZO R, SAINATH T N, SIMKO G, et al. Feature learning with raw-waveform CLDNNs for voice activity detection[C]//Interspeech 2016. San Francisco, American, 2016, DOI: 10.21437/Interspeech.2016-268.
- [12] DAVIS S, MERMELSTEIN P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1980, 28(4): 357-366.
- [13] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.