

一种强噪声干扰下的炮控系统声音识别算法

张雷,袁博,查晨东

(陆军装甲兵学院 兵器与控制系,北京 100072)

摘要:声音作为一种重要的信息媒介,能够为维修人员提供大量的装备信息;但实际维修环境受到车辆启动噪声的干扰,难以准确直观地对声音进行判断;为实现对炮控系统各主要声音部件启动过程的识别,提出了一种基于改进谱减法降噪和多类型识别策略的声音识别算法;通过对炮控系统各部件与发动机声音信号的分析,利用改进谱减法对声音样本进行了降噪处理,并通过实验优化了谱减参数,进一步提升了降噪性能,解决了强噪声干扰的问题;利用滑窗校正和短时能量同步检测的方法制定了具体的识别策略,解决了实际应用中识别结果不稳定以及多类型过程识别的问题;通过实验验证,该声音识别算法对炮控系统各部件启动状态识别准确率达92.4%,具有较好的识别性能。

关键词:声音识别;环境感知;可穿戴维修

A Sound Recognition Algorithm for Gun Control System under Interference of Strong Noise

Zhang Lei, Yuan Bo, Zha Chendong

(Arms and Control Department, Army Academy of Armored Forces, Beijing 100072, China)

Abstract: However, the actual maintenance environment is disturbed by vehicle starting noise, so it is difficult to accurately and intuitively judge the sound. In order to recognize the start-up process of the main sound components of gun control system, a sound recognition algorithm based on improved spectral subtraction noise reduction and multi-type recognition strategy was proposed. Through the analysis of the sound signals of gun control system components and engine, the improved spectral subtraction method is used to reduce the noise of sound samples, and the spectral subtraction parameters are optimized through experiments, which further improves the performance of noise reduction and solves the problem of strong noise interference. A specific recognition strategy is developed by using sliding window correction and short-time energy synchronization detection, which solves the problems of unstable recognition results and multi-type process identification in practical application. Experiments show that the recognition accuracy of this algorithm is 92.4% for the start-up status of each component of gun control system, and it has good recognition performance.

Keywords: sound recognition; environmental awareness; wearable maintenance

0 引言

坦克炮控系统启动时各部件会按照一定的顺序上电工作,具有较为明显的音频特征。维修人员能够根据炮控系统的启动声音来判断各部件是否正常上电工作,从而缩小可能的故障范围,快速进入更为具体的测试诊断工作^[1]。将声音识别技术引入可穿戴维修检测设备中符合可穿戴设备的发展方向,即辅助增强用户的感知、判断和任务处理能力^[2]。将声音识别技术应用与可穿戴维修检测设备能够增强维修人员的听觉能力,在分析音频信号的同时给出部件的状态信息,为维修操作提供更加快速准确的判断^[3-4]。

基于可穿戴设备的环境声音识别技术是声音事件识别(Acoustic Event Classification)和环境感知(Context A-

ware)技术在特定环境中的一种应用^[5-6]。目前,环境声音识别的应用大多以语音识别技术为基础,根据不同的应用特点与实际需求再建立合适的识别模型。文献[7]提出了一种基于Mel子带参数化特征的声音识别算法,采用高斯混合模型拟合的方法进行声音事件的自动分段,并利用SVM分类器实现了11种鸟鸣声音的识别。文献[8]结合工厂的实际噪声环境实现了电机异常噪声的检测识别,在特征提取过程中利用汉宁卷积窗代替汉宁窗,有效抑制了频谱泄露,提高了电机异常噪声特征提取的准确性。文献[9]提出了一种基于小波包分析的非接触式玻璃破碎声音识别系统,通过提取小波包系数与短时特征参数,从时频域两个角度对玻璃破碎声进行了描述,提高了事件识别的准确率。文献[10]提出了一种基于特征分析的环境声音识别算法,通过提取声音事件的4种典型特征进行简单分类,该算法比模型训练匹配具有更好的实时性和环境适应性,但在多类型复杂事件的分类中还存在特征区分度不佳、识别准确率不足的问题。

针对某型坦克基层级维修中伴随的强噪声干扰问题,本文提出了一种以改进谱减法降噪为基础、以梅尔频率倒

收稿日期:2018-11-20; 修回日期:2018-12-10。

作者简介:张雷(1974-),男,吉林九台人,博士,副教授,主要从事武器系统诊断与评估方向的研究。

通讯作者:袁博(1993-),男,硕士研究生,主要从事穿戴维修检测技术方向的研究。

谱系数 (MFCC) 为特征向量的声音识别算法。该算法能够在车辆启动的强噪声干扰下识别炮控系统中各主要部组件的启动状态。针对降噪性能不好以及识别准确率不高的问题, 进一步对谱减参数与识别策略进行了优化。

1 声音样本分析

根据炮控系统的实际启动状况, 变流机、油泵电机、电磁离合器和电机扩大机四个部组件的启动过程的声音特征较为明显, 是维修人员的主要判断目标, 具体安装位置如图 1 所示。

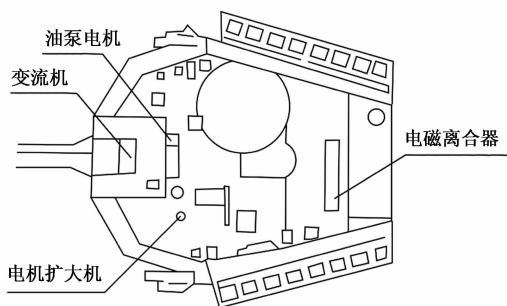


图 1 各部组件安装位置示意图

在基层级维修中, 炮控系统一般有两种启动方式。一种是外接电源供电, 这种方式不需要启动车辆, 环境噪声较小, 坦克乘员与维修人员可直接根据启动声音对各部件状态进行判断; 另一种是车辆发动供电, 这种方式通常会带来严重的噪声干扰, 难以对启动部件声音进行准确的判定。大多数基层级维修与现场抢修环境无法供给外接电源设备, 炮控系统启动声音难以避免的会受到发动机噪声的覆盖。图 2 和图 3 分别为车辆启动供电和外接电源供电下的声音样本。能够看出在车辆启动噪声下, 炮控系统启动声音的时频特征受到覆盖, 而外接电源供电下的辨识度较高。为较好的展示炮控系统的启动过程, 在图 3 中对炮控系统启动的各个阶段进行了标识, 定义 a 为静音段, b 为变流机启动段, c 为油泵电机启动段, d 为电磁离合器闭合声音, e 为电机扩大机启动段。

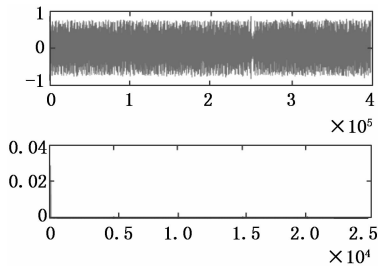


图 2 车辆启动供电声音样本

声音事件识别在很多领域中都已经有了较为成熟的应用方案。然而炮控系统启动过程的声音识别存在一定的特殊性, 难以直接采用现有的声音识别方案, 主要原因有以下几点:

1) 炮控系统启动声音受到持续的强噪声覆盖, 这种背

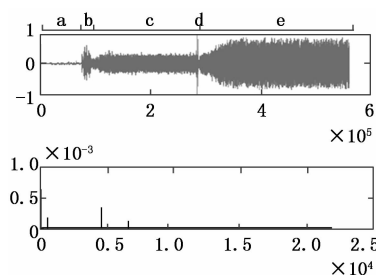


图 3 外接电源供电声音样本

景噪声远大于常见的环境噪声, 有效信号的信噪比极低, 信号特征难以提取。

2) 即使通过预加重与滤波降噪等方式减少噪声干扰, 各电机启动段也存在短时能量与过零率区分度不足的情况, 阈值也难以把握, 无法实现准确的端点检测。

3) 部组件启动状态识别是一种事件级识别, 即识别出各部件是否按照一定顺序上电工作, 这与常见的声音检测目的存在一定的区别。

4) 各部组件声音具有多样性, 一方面电磁离合器闭合声音表现为极短时间内的一个高频信号, 而其他部件为持续性信号, 若采用单一分类标准容易存在虚警和漏检现象; 另一方面变流机、油泵电机与电机扩大机存在一个 1~2 s 的趋稳过程, 随后进入稳定运行状态, 单一标签内的特征状态往往呈现出一定的变化性, 难以准确划分各阶段的分界点。

因此, 声音信号的抑噪性能和合理的声音识别策略是准确识别各部组件启动过程的关键。

2 改进谱减法降噪

2.1 基本谱减法

谱减法降噪最早由 boll 提出, 具有算法简单、计算量小、实时性好的特点, 因此成为声音识别中最为常用的去噪方法之一。该方法的基本假设是将噪声视为不相关的加性噪声。假设车辆启动噪声信号 $d(t)$ 线性叠加到炮控系统声音样本 $s(t)$ 中, 见式 (1):

$$x(t) = s(t) + d(t) \tag{1}$$

通过计算含噪声音信号的短时功率谱, 可以按照式对含噪声音样本中的炮控系统启动部分进行估计。

$$|\hat{S}(e^{j\omega})|^2 \approx |X(e^{j\omega})|^2 - |\bar{D}(e^{j\omega})|^2 \tag{2}$$

由于噪声信号的变化性, 估计噪声与有效噪声会存在一定差异。利用基本谱减法估计车辆启动噪声, 能够在一定程度上保留炮控系统启动的时频特征, 但当实际噪声信号大于平均值的时候, 容易在处理样本中出现残差噪声。这种残差噪声在声音波形上表现为随机性的尖峰凸起, 听觉上类似于音乐的起伏, 因而被称为音乐噪声 (Music Noise)。

2.2 改进谱减法

为解决音乐噪声带来的负面影响, 提高去噪样本的信噪比, Berouti 提出了一种带参数的改进谱减法。这种算法

能够将噪声功率谱中的过估计部分减去,并将得到的频谱限制在一个预定的水平值以下,具体计算方法见式。

$$\begin{cases} |\hat{S}(e^{j\omega})|^2 = \\ \left\{ \begin{array}{l} |X(e^{j\omega})|^2 - \alpha |D(e^{j\omega})|^2, \quad |\hat{S}(e^{j\omega})|^2 > \beta |\bar{D}(e^{j\omega})|^2 \\ \beta |\bar{D}(e^{j\omega})|^2, \quad \text{其它} \end{array} \right. \end{cases} \quad (3)$$

其中: α 为相减因子 ($\alpha \geq 1$), β 为频谱下限阈值参数 ($0 < \beta \leq 1$)。相减因子的加入有利于增强去噪效果,减少残留噪声,但过大的相减因子也会增加声音样本的失真度,需要在提高信噪比与降低失真度中找到一个较为合理的设置参数。通过调整下限阈值参数实际上是留给了残留噪声一个幅值区域,将小于这个下限值的部分统一设置为固定值,以此减小残留噪声峰值与 0 之间的差值幅度,缓解音乐噪声的影响。

2.3 谱减参数优化

声音去噪效果的评价方法包括主观评价方法和客观评价方法,主观评价方法常用于语音增强应用中,利用用户对降噪后语音的实际感受进行评分。声音事件识别是为了将有效信号的时频特征突出出来,不需要将去噪声音返还给用户,因此采用客观评价方法来判断实际去噪效果。

信噪比是衡量声音降噪性能的常规方法,但是炮控系统启动过程变化幅度较大,经典信噪比的整体评价方法并不准确,因此采用分段式信噪比 (Segmental SNR, Seg-SNR) 计算方法,定义见式 (4):

$$\text{SegSNR} = 10 \log_{10} \frac{\sum_{K=0}^{N-1} |X(e^{j\omega})|^2}{\sum_{K=0}^{N-1} |N(e^{j\omega})|^2} \quad (4)$$

其中: N 为帧长。利用对每一帧信噪比进行计算再求取平均的方法,能够更好地评价声音样本的去噪质量,一般 SegSNR 值越大,说明含噪信息越小,声音波形也越接近外接电源下的炮控系统启动声音。

失真度 LSD 用来描述去噪前后声音样本的保留程度, LSD 值越小表示降噪后的声音样本与纯净声音样本越接近,降噪质量越好。LSD 值越大表示对有效声音的破坏越大,降噪质量越差,定义见式。

$$\text{LSD} = \frac{1}{L} \sum_{l=0}^{L-1} \left\{ \frac{1}{N/2+1} \sum_{k=0}^{N/2} [10 \lg X(k,l) - 10 \lg \bar{X}(k,l)]^2 \right\} + \quad (5)$$

其中: $X(k,l)$ 和 $\bar{X}(k,l)$ 分别为降噪后和纯净信号的傅里叶变换分量, LSD 越小表明信号质量越好,越大表明降噪操作对有效信号的损失越大。

以 44.1 kHz 采样率分别对车辆启动和炮控系统启动声音进行采样,通过时域叠加的方式构造车辆启动状态下的炮控系统启动声音样本,计算方法见式。

$$y(n) = s(n) + ad(n) \quad (6)$$

$y(n)$ 为构造出来的含噪声音样本, $s(n)$ 为外接电源下炮控系统启动声音样本, $d(n)$ 为车辆启动噪声, a 为噪

声混合系数。样本由相同环境条件下的实际采样获得,对谱减参数 α 和频谱下限阈值参数 β 取值,观察 SegSNR 与 LSD 变化,从而确定改进谱减法参数的最佳取值。经过试验求得 $\alpha = 10, \beta = 0.05$ 时能够较好的平衡信噪比与失真度。将原始信号分帧加窗预加重后的信号经过改进谱减法降噪后的波形对比如图 4 所示。虽然与纯净信号相比存在了较大的失真度,但基本能够体现出各阶段特征的差异性,为特征提取与分类识别部分打下了基础。

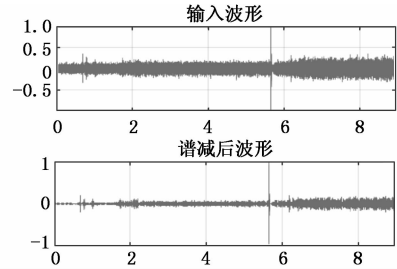


图 4 改进谱减法降噪前后波形对比

3 MFCC 特征提取

声音信号的常见特征参数有短时能量、短时过零率、线性预测倒谱系数等,其中 MFCC 模拟了人耳的感知效果,在有低频噪声干扰时具有较好的辨识度,具体计算步骤为:

- 1) 信号预处理: 预加重, 分帧, 加窗;
- 2) 短时傅里叶变换, 得到频谱与幅度谱;
- 3) 让幅度谱通过 Mel 刻度滤波器;
- 4) 对滤波输出做对数运算, 并通过离散化求得 MFCC 系数

其中 Mel 刻度滤波器由一组 N 个三角形带通滤波器组成 (N 一般在 20~30 范围内取值), 每个三角窗滤波器所覆盖的范围都近似于人耳的一个临界带宽, 从而模拟出人耳的掩蔽效应。Mel 频率与实际频率的具体关系见式, 实际频率在 1 kHz 以下时呈近似的线性分布, 1 kHz 以上时呈对数型增长。

$$f_{mel} = 2595 \cdot \log\left(1 + \frac{f}{700 \text{ Hz}}\right) \quad (7)$$

图 5 展示了提取谱减降噪后的音频信号的低 8 维 MFCC, 各种颜色代表了不同维数的特征参数, 其中横坐标代表采样帧序列, 纵坐标代表振幅的对数运算结果, 单位为分贝。

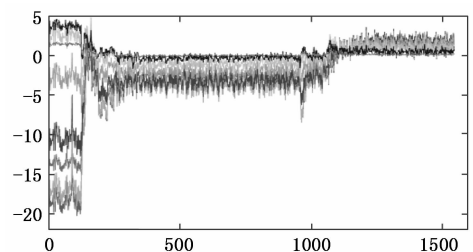


图 5 低 8 维 MFCC 特征参数

4 实验过程与结果分析

4.1 数据库与实验环境

通过实地采集声音样本的方式建立坦克炮控系统启动的声音数据库, 其中单一样本以车辆启动前 1~2 s 的静默段为起始, 以电机扩大机稳定运行 3~5 s 后结束。音频文件以 wav 格式保存, 采样频率为 44100 Hz, 32 bit 量化, 单通道。根据炮控系统启动时间的不同, 各音频文件长度在 8~25 s 范围之内。

本文针对炮控系统启动中变流机、油泵电机、电磁离合器、电机扩大机的 4 个启动过程进行识别, 具体实验条件: 1) 软件平台: Matlab2016a; 2) 录音设备: 飞利浦 vtr8800 录音笔。

4.2 滑窗法校正

基于 MFCC 特征分析结果可以看出各组件启动过程在一定范围内具有较为明显的区分, 但单帧特征变化幅度较大, 因此加入滑动窗口解决识别结果跳变的现象, 经过多次尝试, 窗口深度在 200 时识别效果较好, 滑窗相关代码如下:

```
win = 200; %窗口大小
step = 1; %滑动步长
p = predictLabels';
res = '';
for i = 1:step:length(p) - win
    w = p(1,i:i+win-1);
    res1 = res;
    res = judge( w );
    if ~strcmp(res1,res)
        disp(res);
    end
end
```

滑动校正前后对比结果如图 6 所示, 左侧为单帧识别结果, 右侧为窗口深度 200 时的识别结果, 跳变现象得到了明显的改善。

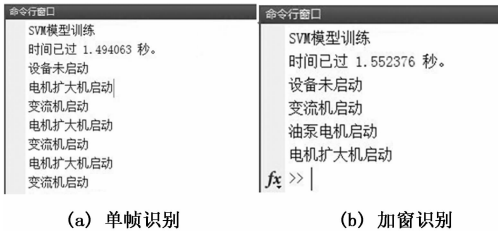


图 6 加窗校正前后对比

4.3 多类型过程识别策略

由于电磁离合器为短时间内的低频声音信号, 因此该状态的识别不适用于滑窗法。根据上述识别结果可以看出, 滑窗校正后“电磁离合器闭合”完全消失于识别结果中, 因此将电磁离合器闭合声音的识别作为独立部分进行处理。

经过实验分析, 对于预加重、谱减降噪后的声音, 电磁离合器闭合声音的短时能量特征具有极为明显的区分度。因此设计采取多类型过程的识别策略, 将 SVM 分类与短时能量检测同步进行, 在短时能量到达阈值时输出“电磁离合器闭合”的状态信息。由于电机扩大机状态识别在滑动窗口下具备一定的时间延迟, 因此不会影响到与电磁离合器闭合状态识别的先后关系。具体识别流程如图 7 所示。

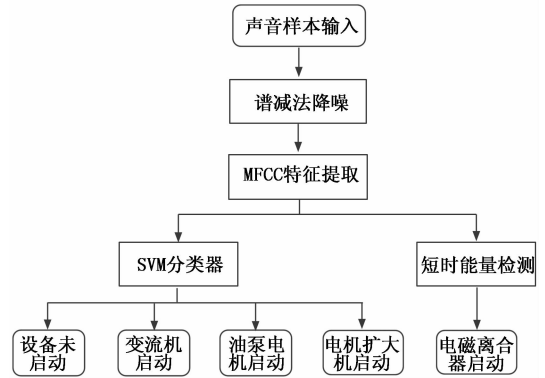


图 7 识别流程

4.4 实验结果分析

下一步利用实验验证该声音识别算法性能, 对降噪后的声音样本进行 MFCC 特征提取, 生成一个 $n \times m$ 矩阵, n 为帧数, m 为特征维度。将 MFCC 作为输入值进行训练, 其中每个帧都提前设定好类别标签 0, 1, 2, 3, 4, 分别对应着设备未启动、变流机启动、油泵电机启动、电磁离合器闭合、电机扩大机启动五个识别状态。在训练过程中, 先将帧序列打乱, 然后抽取 80% 的数据作为训练样本, 剩下 20% 数据作为测试样本。训练完毕后, 将测试数据的分类结果与真实标签作对比得到识别率, 如图 8 所示, SVM 分类器的单帧识别准确率达 86.44%。

```
命令窗口
nSV = 1701, nBSV = 0
.....*
optimization finished, #iter = 13111
C = 30.485418
obj = 650.406392, rho = 0.022460
nSV = 2889, nBSV = 0
.....*
optimization finished, #iter = 9557
C = 24.847949
obj = 346.256167, rho = 0.515775
nSV = 1870, nBSV = 0
Total nSV = 4165
Accuracy = 86.4407% (1224/1416) (classification)
fx >>
```

图 8 SVM 分类准确率

该应用中声音识别的最终目的在于准确识别出各组件的启动状态, 因此除了验证分类算法本身性能, 还需要对实际应用情况进行一个定量评估。具体方法为从声音库中提取 50 个声音样本, 逐个从单一样本中抽取固定标签数量的随机帧信号特征进行训练, 并用训练模型对包括该样本在内所有 50 个声音样本进行识别, 从总计 2500 次识别结

(下转第 111 页)