

ABSTRACT

The security of public places for social stability and harmony, people's lives and property is of great significance. Currently, video surveillance in public places with the core technology of video monitoring platform has played a positive role on reducing the crime rate and improving the efficiency of investigation and scientific evidence collection. Gunshot signal often contains very important information including the type of weapon, the size of the fire and so on. This information provides first-hand information including criminal fire intelligence and clues to solve crimes. Feature extractions of gunshot mostly follow the traditional methods of speech signal processing. Gunshot signal in public places is the typical non-speech signal with non-stationary and nonlinear. So, the study of gunshot signal provides a new train of thought for processing other types of non-speech signal. Therefore, Gunshot feature extraction and classification in public places is of significant practical value and academic significance. In this paper, First analysis gunshot signal of public and environmental background noise characteristics, then proposed an Ensemble Empirical Mode Decomposition (EEMD) model of the gunshot signal feature extraction and recognition method and final results are verified through the experiment.

This paper main work is as follows:

① Feature analyze of the typical gunshot and background noise in public places. Several kinds of typical gunshot signal and he background noise characteristics analyzed such as time domain, frequency domain, frequency domain, cepstrum domain have been done. Though statistical analysis of relevant features, this paper argues that gunshot has common property of nonlinear, non stability and the different types of shots such as pistols, rifles, machine guns have more similarities than differences. Environment in public places such as square, station, wharf and other background noise is very loud, often swamped the original gunshot signal characteristics, makes sound signal in public places is very irregular, even destroy the original gunshot signal of time-frequency and other related features. In one of the few studies in the literature, will be a public place noise model is assumed as Gaussian distribution, this paper analysis assumes that the background noise in public places for symmetric α stability ($S\alpha S$) than the Gaussian distribution is more accord with the actual situation of public places.

② Gunshot character description modeling in public places. From ① to know, gunshot signal have characteristics of nonlinear, non-stability. Empirical Mode Decomposition (EMD) method is an effective method for this kind of signal. Therefore, in this paper, on the basis of the EMD, gunshot in public signal feature extraction and recognition method is researched. Ensemble Empirical Mode Decomposition (EEMD) though adding the random Gaussian distribution sequence to analyzed signal to solve the problem of EMD, like mode mixing. Inspired by this, the article give full consideration to public environment background noise distribution of the actual situation, improve models of the EEMD put forward in the reconstructed signals use $S\alpha S$ distribution instead of Gaussian distribution, EMD decomposition to the reconstructed signal, get Intrinsic Mode Function (IMF), take the overall average as the final IMF of original signal.

③ Gunshot feature extraction methods in public places. Gunshot signal after EMD decomposition will generate a finite number of IMF components, and frequency of each IMF component is not fixed. To this end, this article will after ② in the gunshot character description method proposed in this paper decomposed IMF component of the generated Fast Fourier Transform (FFT), according to Parseval theorem to find the energy of the original signal and each IMF component of energy, eventually take the ratio of each IMF energy to the firsthand signal energy as feature vector

Experiment results show that the proposed improved EEMD model compared with traditional speech processing method of Mel Frequency Cepstrum Coefficient (MFCC) and Linear Prediction Cepstrum Coefficient (LPCC) characterization parameters, such as more accurate description of gunshot. At the same time, compared with the original EMD and EEMD our method has also a certain degree of increase, and has better robustness.

Keywords: Ensemble Empirical Mode Decomposition, Public Places, Feature Extraction of Gunshot, Ratio of Energy

目 录

中文摘要.....	I
英文摘要.....	III
1 绪 论.....	1
1.1 课题的研究背景及意义.....	1
1.2 课题的研究现状.....	2
1.3 课题的研究难点.....	6
1.4 论文的主要研究内容.....	7
1.5 论文的组织安排.....	7
2 公共场所典型枪声信号及背景噪声特性分析.....	9
2.1 引言.....	9
2.2 公共场所典型枪声信号的特性分析.....	9
2.2.1 典型枪声信号的时域分析.....	9
2.2.2 典型枪声信号的频域分析.....	13
2.2.3 典型枪声信号的时频域分析.....	15
2.2.4 典型枪声信号的倒谱分析.....	18
2.3 公共场所背景噪声的特性分析.....	20
2.3.1 公共场所背景噪声的特性分析.....	20
2.3.2 公共场所背景噪声分布模型.....	20
2.4 本章小结.....	22
3 公共场所枪声特征描述建模与特征提取方法研究.....	23
3.1 引言.....	23
3.2 经验模态分解及其改进.....	23
3.2.1 EMD.....	23
3.2.2 EEMD.....	25
3.3 公共场所枪声特征描述建模及特征提取方法.....	26
3.3.1 公共场所环境下枪声信号重构.....	26
3.3.2 改进 EEMD 的枪声特征描述建模.....	27
3.3.3 公共场所枪声特征提取方法.....	31
3.4 分类器.....	32
3.4.1 BP 神经网络.....	32
3.4.2 支持向量机.....	33

3.5 本章小结	34
4 实验结果与分析	35
4.1 引言	35
4.2 预处理	35
4.2.1 端点检测	35
4.2.2 自适应滤波	38
4.3 实验结果与分析	42
4.3.1 不同特征描述方法对比实验	42
4.3.2 不同分类器对比实验	45
4.4 本章小结	46
5 总结展望	47
5.1 工作总结	47
5.2 工作展望	47
致 谢	49
参考文献	50
附 录	53
A. 作者在攻读硕士学位期间发表的论文及取得科研成果	56

1 绪 论

公共场所的安全防范对于维护社会和谐稳定,保障公民生命财产安全意义重大。目前,以视频监控为核心技术的公共场所视频监控平台,为减少犯罪率,提高破案效率和取证的科学性发挥了积极的作用。然而社会对公共场所安全监控、安全防护提出了更快、更全、更好的要求。枪声信号通常包含武器的类型、火力的大小等,这些信息既为警方提供了破案的线索又为警方提供罪犯的火力情报。目前,公共场所枪声信号特征提取方法大多沿用传统的语音信号处理方法。对于枪声这种典型的非语音信号,传统的语音信号处理方法存在明显的不足。因此,对公共场所枪声特征提取及识别研究具有重要的社会意义和学术价值。本章首先介绍课题研究的背景和意义,然后着重对公共场所枪声特征提取及识别的研究现状进行分析,并针对研究难点提出本文的主要研究内容。

1.1 课题的研究背景及意义

公共场所是指公众从事各种社会活动的地方。在公共场所环境下人们日常生产活动中所发生的各种危险事件、交通事件、犯罪事件等等都与人们的生命财产安全息息相关。目前,以视频监控为核心技术的公共场所视频监控平台,为减少犯罪率,提高破案效率和取证的科学性发挥了积极的作用。但像校园枪击案一样令人悲痛的事件仍时有发生,当地时间 12 月 14 日,美国康涅狄格胡克小学发生校园枪击惨案,造成 28 人死亡,其中包括 20 名儿童。这是 2012 年以来美国发生的最血腥、死亡人数最多的枪击事件,这也是继崔荫案、卢刚案、科伦拜恩校园枪击案、古腾堡中学校园枪击案、维吉尼亚理工大学枪击案、约凯拉校园枪击案、北伊利诺伊大学枪击案和温嫩登校园枪击案后的另一起最严重的校园枪击案。北京时间 3 月 19 日 19 时 42 分,某驻渝部队营房哨兵遭歹徒持枪袭击,经抢救无效身亡,歹徒抢走自动步枪 1 支,此事引发社会广泛关注。枪击事件往往由于事发突然,而致警方无法在第一时间得到准确信息做到实时反应,致使不能将损失降到最低。这主要是因为单纯的摄像机录像监控方式,夜晚和阴天情况下效果不佳,而且犯人往往有意避开摄像头存在的区域进行作案,更加加大了预警的难度。而声音信号则不受光线和天气的影响,可以全天候、大范围、隐形化的对公共场所进行监控。枪声信号通常包含了至关重要的信息,包括武器的类型、火力的大小等,这些信息既为警方提供了破案的线索又为警方提供罪犯的火力情报,为下一步的武力镇压提供第一手资料。社会公共安全监控对预防和防止公共场合枪击事件的发生,以及公共安全应急管理提出了更高的要求。因此,对公共场所枪声的

特征提取及识别方法进行深入研究，对维护国家安全、社会稳定和公共安全有很重要的社会意义及实用价值。

然而，目前业界对公共场所典型枪声还没有统一明确的定义，本文根据射程和破坏力将典型枪声分为三类（手枪、步枪、机枪）。同时在一些现有的文献中，对于枪声的特征提取与识别方法仍然沿用传统的针对语音信号处理的方法如 MFCC、LPCC 等^[1]。由于公共场所枪声为典型的非语音信号，传统的特征提取方法显然不能满足实际需要。另外，公共场所背景噪声对公共场所枪声特征提取和识别有较大的影响。另一方面，声音识别技术是一个非常活跃的研究领域，其研究涉及心理学、生理学、人工智能、模式识别、数字信号处理、发声机理和听觉机理等多个学科，更是模式识别的典型案例分析之一^[2]。对本课题的研究可以促进语音与非语音信号处理方法的发展。因此对本课题的研究具有重要的理论研究价值和学术意义。

1.2 课题的研究现状

公共场所枪声特征提取及识别主要涉及预处理、特征提取及分类器设计三个部分^[3]。公共场所枪声特征提取及识别流程如图 1.1 所示。① 预处理：主要包括抑制 50Hz 工频干扰预滤波，信号的分帧、加窗等。因此，对于公共场所的背景噪声的预处理有两个主要的步骤：一是端点检测，即从背景噪声提取感兴趣的信号片段；二是尽量减小工频噪声对于枪声的干扰。② 特征提取：公共场所枪声特征提取有两个主要步骤：一是对枪声在时域、频域、时频域、倒谱域上进行信号分析以寻找适当的特征提取描述方法；二是通过特征提取方法获取枪声信号的特征向量。③ 分类器选择：分类器的好坏对识别结果有直接影响，因此选择性能优良的分类器能获得更好的识别结果。

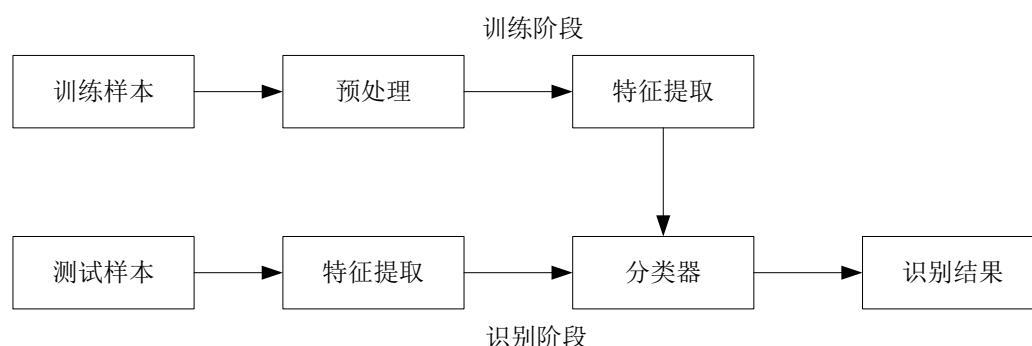


图 1.1 公共场所枪声特征提取及识别流程图

Fig.1.1 Flow chart of gunshot feature extraction and classification in public places

枪声的发生由两部分构成：一是在子弹发射时，弹体被燃烧的发射火药推出，灼热的弹体在钢壳体中高速前进的过程中与枪管剧烈摩擦，从而产生高温高压气体，瞬间与氧气混合释放出巨大的爆炸压力波；二是子弹速度高于声速突破音障与空气高速摩擦产生的呼啸声。通过调研得到，对语音信号的研究已经有很长的历史，而对非语音信号尤其是枪声信号的研究，是由于近几年对公共安全的迫切需要才得到研究人员的关注。目前对枪声信号的处理大多沿用传统的语音信号处理方法如 MFCC、LPCC 等。因此，研究适合于非语音信号的特征提取方法成为解决枪声信号特征提取及识别的关键。本节首先对传统的语音信号处理方法所涉及的主要理论及技术的发展做一个简要的介绍。然后，对目前典型的非语音信号特征提取方法进行分析。

① 语音信号特征提取方法

20 世纪 50 年代初期贝尔实验室研制的 Audry 系统^[4]标志着语音信号处理方法的出现。该系统通过提取语音信号的共振峰位置特征参数，采用模拟电路实现对待测样本的匹配识别，成为第一个可以实现英文数字的语音识别系统。从此人们对声音识别作了大量的研究工作，提出了很多行之有效的方法。特征提取的任务是要从语音信号中提取出表现语音个性并揭示语音本质的特征，然后对其进行编码得到特征序列。使用不同的技术方法分析相同的语音信号，可以得到反映语音信号不同性质的特征序列。较早出现的是线性预测编码 (Linear Prediction Coefficient, LPC)^[5]。LPC 利用声音信号过去采样点的线性组合来逼近现在或者未来的语音信号采样点，使得提取的特征能用极少的参数就可以有效地表征语音信号，同时计算简单，具有较好的时间复杂度。结合 LPC 技术和倒谱分析出现了线性预测倒谱系数 LPCC。LPCC 是一种基于语音信号数字模型的方法，利用信号过去的若干个样点值来估计未来的样点值，求解过程中少量的噪声就会对线性预测系数带来极大的干扰^[6]。语音信号处理领域使用最为广泛的特征提取方法是基于倒谱分析的梅尔频率倒谱系数 MFCC^[7,8]。该方法打破传统的线性正比关系，建立了梅尔频率尺度，该尺度更加符合人耳对外界声音信号的频率感知关系。MFCC 是基于人耳听觉的模型，其提取特征的关键在于梅尔滤波器组对信号的滤波。滤波器组在梅尔频率域是均匀分布的，但枪声的信息在其频谱范围不是均匀分布的，于是就产生了各个通带统计输出不能均等的问题。

② 非语音信号特征提取方法

目前对非语音信号的研究主要延续了语音识别的技术。典型的有：

1) 卡内基梅隆计算机科学学院的 Hoiem 提出的复杂声学环境下声音目标检索方法：该方法的核心是采用决策树选择声音信号时域中最具有区分性的声学特征，并用这些特征为每个研究对象建模，通过相似度计算识别出感兴趣的信号^[9]。该研

究包括汽车鸣笛、狗叫、猫叫、门铃声、关门声、尖叫、拔剑声、男人的笑声、激光枪等 13 种不同的从电影片段中提取的声音。Hoiem 将电影中剪辑出的音频数据分为成短时、等长且交叠的音频帧，计算每帧短时傅里叶变换，然后在 200Hz~4000Hz 范围将音频帧分成 16 个频段，使每个频段的平均能量近似相等。在此基础上计算每个频段的能量值、能量波动值、基音以及响度等特征构成特征集合。通过决策树选择最具有区分性的特征，来保证每个样本正确分开。该方法存在的问题在于特征提取参数选择不当：对于多种相似的枪声信号，其时域特征都具有脉冲性，因此仅仅时域上的声学特征如能量、响度等对于信号本质的揭示明显不足，区分性也较差。另外，研究对象为模拟信号，结果可能与真实信号有所区别。2) 日本东北大学工程研究院的 Akinori Ito 提出从闭路电视系统拾音器采集到的音频信息中检测出枪声的方法^[10]。该方法实验室的环境下，以空调和计算机风扇声音作为背景噪声，消防警铃、玻璃碎声、男女的尖叫声、枪声以及蜂鸣声作为研究对象，采用 16 维 MFCC、1 维短时能量以及它们的一阶差分共 34 维参数作为特征，利用多阶高斯混合模型对每类对象进行建模，最后通过设置阈值、计算相似度识别对象。该方法存在的问题是泛化性能差，不同的研究对象、研究背景都需要通过大量实验重新确定不同阶高斯混合模型的参数。3) 新加坡国立大学信息通信研究所的 K. Atrey 以办公室走廊为研究背景，谈话声、叫喊声、敲门声、脚步声作为研究对象^[11]。首先通过实验得到 50ms 为感兴趣事件的最小长度，以此为依据将声音传感器采集到的音频数据流分割为 50ms 一帧。该方法将 ZCR、LPC、LPCC 等作为特征集，高斯混合模型作为分类器。其核心表现在多层次分类过程，第一层区分背景音、前景音；第二层区分前景音中的语音与非语音；第三层识别非语音中的敲门声和脚步声，语音中的谈话声和喊叫声。该方法的优点是考虑了背景噪声对识别的影响，通过高斯混合模型对背景噪声建模以区分感兴趣的对象和背景噪音，但同时带来了实验设计复杂的缺点，对于不同的研究背景、研究对象，都需要大量的实验进行建模。4) 韩国首尔高丽大学的 Woohyun Choi 提出在不同背景环境下枪声目标检索方法：该方法的核心是采用最大似然线性回归选择声音信号时域中最具有区分性的声学特征，并用这些特征为每个研究对象建模，通过相似度计算识别出感兴趣的信号^[12]。该研究包括汽车鸣笛、狗叫、门铃声、玻璃破碎声、尖叫声、枪声等 9 种不同的声音。该方法采用 24 维 MFCC、24 维 Delta、12 维 LPCC 构成特征，GMM 作为分类器在无噪声、街道、地铁、音乐等背景下进行了对比实验，从实验结果来看在高信噪比时效果很好，但低信噪比时效果急剧下降，鲁棒性不好。

目前国内已有少数研究机构开始对枪声识别进行研究，其中主要包括重庆大学^[9]、中科院声学所、西安交通大学、西南交通大学等。研究人员在枪声识别研究

领域进行了许多有意义的尝试,积累了一些经验,并取得了一定的成果。

而针对枪声信号这一典型的非语音信号,国外学者主要开展的研究工作如下:

1) 巴西军事工程学院的 L. Freire 出于执法和军事的考虑,针对枪声信号提出了基于信号自相关的检测方法。该方法通过计算自相关函数的阈值来检测枪声^[13]。对于枪声这种冲激信号,研究者将拍掌声,气球爆炸声和说话时的爆破音作为环境中的干扰因素。最后通过传统语音识别的方法进行对比实验,采用 8 维 LPCC、13 维 MFCC 构成特征, HMM 作为分类器。该方法优点是计算简单,但缺点是背景简单,而在实际环境下背景复杂,该方法显然不再适用。2) 意大利科摩的计算声学 and 音响工程实验室的 G. Valenzise 提出一种枪声和尖叫声检测方法,该方法的特点在于采用了大量的时域及频域特征参数对信号进行表征^[14]。研究者以广场为背景,对枪声和尖叫声进行了研究,还开发出了相关 Audio-Surveillance 系统,该系统已在意大利米兰广场上使用。该方法对枪声信号提取特征集,包括 1 维的 ZCR、8 维的谱距相关特征、30 维的 MFCC 以及基于信号自相关函数的 10 维特征,然后进行特征子集的优化选择并降维。最后使用并行的高斯混合模型分别对枪声建模并进行分类。该方法存在的问题是:传统的特征参数对枪声信号没有足够的表征力,该方法通过罗列 49 维不同的传统特征参数对信号表征,以弥补表征力的不足;枪声种类单一,只考虑与尖叫声区分而不考虑不同种类枪声的识别;另外该方法中没有考虑广场噪声的影响。

相比国外的研究工作,国内研究者在枪声特征提取及识别领域开展的工作还在起步阶段。重庆大学的栾少文等人利用 MFCC 作为特征提取方法、HMM 作为分类器意图区分枪声和尖叫声、爆炸声,虽然通过改进参数的方法一定程度提高了识别率,但由于原理上使用语音信号特征提取方法提取枪声信号,因此并不十分适合,导致系统对噪声的鲁棒性较差^[15]。西南交大的吕霄云等人选取枪声的 MFCC 与短时能量构成组合参数,利用高斯混合模型做分类器区分枪声与爆炸声,并在电梯环境下做了相关实验进行验证,但由于选取的特征提取方法针对性不强,导致结果并不十分理想^[16]。

综上所述,目前对枪声信号特征提取及识别方法大多沿用传统的语音信号处理方法。特征描述方法主要有时域、频域、时频域和倒谱域特征参数。其中,时域特征有过零率、基音周期、短时能量、基音频率等;频域分析有频谱、线谱对分析等;时频域分析方法有短时傅立叶变换、Wigner-Ville 分布等、Gabor 变换、小波变换等^[17];倒谱域参数有感知线性预测倒谱系数、对数频率倒谱系数、梅尔频率倒谱系数等。经验模态分解方法是一种用于解决非线性、非平稳信号的方法。EMD 是由美国国家宇航局的 Norden E. Huang 于 1998 年提出的^[18],它能对信号进行时频局部化分析,摆脱了傅里叶变换理论的束缚,能够充分地解释瞬时频率

的定义。EMD 方法的本质就是通过平稳化处理信号,逐级分解得到在某些尺度上的波动或趋势,因此非常适合分析非线性、非平稳信号,具有很高的信噪比^[19]。一个复杂信号通过 EMD 分解会得到有限个本征模态函数 IMF,原始信号的局部特征信息就隐藏在各阶 IMF 分量之中并且 EMD 分解具有自适应性^[20]。目前,该方法已经广泛应用于海洋、大气、生物医学、天体观测资料与地震记录分析、合成孔径雷达图像滤波、机械故障诊断、密频动力系统的阻尼识别以及大型土木工程结构的模态参数识别方面等科学研究领域,并取得了较好的效果^[21-28]。但 EMD 存在边缘效应,不稳定与模态混叠的问题,尤其是在处理间歇信号(信号某一频段分量不连续)时偏差很大^[29,30]。本文通过对 EMD 方法的研究以及对枪声信号的特性分析,尝试将 EMD 方法引入枪声的特征提取。经典的总体经验模态分解 EEMD 方法中,由于随机白噪声信号具有在各个频段能量一致的特点,在原始信号中混入随机白噪声,再 EMD 分解就可以保证每个 IMF 在时域上的连续性,由于白噪声是均值为零的随机过程,求所有 IMF 的总体均值就可消除噪声的影响。EEMD 方法的思路很值得借鉴。

③ 分类器

人工神经网络(Artificial Neural Network, ANN)是一种模仿生物神经网络结构和功能的数学模型^[31]。大多数情况下人工神经网络能在外界信息的基础上改变内部结构,是一种自适应系统。BP (Back Propagation)神经网络是一种按误差逆传播算法训练的多层前馈网络,是目前应用最广泛的神经网络模型之一。支持向量机(Support Vector Machine, SVM)是一种监督式学习的方法,可广泛的应用与统计分类及回归分析。支持向量机能够同时最小化经验误差以及最大化几何边缘区,表现出了它本身独特的优良性^[32]。可以说,这两种分类器的性能相对较好,因此本文采用这两种分类器。

1.3 课题的研究难点

公共场所枪声特征提取与识别技术就是通过对枪声信号进行分析,找到合适的特征提取方法,从而使得枪声的类内距离较小,类间距离较大,进而能够从各种嘈杂的背景环境中识别出枪声的种类。现实环境中的许多情况影响枪声信号的准确识别,也正因为如此,枪声的特征提取与识别面临着各种各样的困难和挑战,本文试图就以下几个方面的难点进行攻坚:

① 公共场所枪声信号的特性决定了识别的难度。首先,业界没有对公共场所枪声的标准数据库,需要自行搜集数据建库;其次,原始数据不能直接进行特性分析,需要先进行预处理;最后,为选择最好的特征描述方法,还需要对公共场所典型枪声信号进行特性分析。公共场所典型枪声包括手枪声、步枪声、机枪声

三种声音，这种声音信号的多样性给提取兼顾简约性和区分性的特征描述带来了不可忽视的困难。

② 寻找合适的公共场所枪声特征描述方法。对于枪声的特征描述，既要有针对性，能够反映枪声信号本质信息，又要有包容性，能够兼顾到不同类别的枪声；既要具有同一性，使得同类枪声类内距离很小，又要具有鉴别性，使得不同类型的枪声类间距离很大。为此，寻找合适的公共场所枪声特征描述方法是解决公共场所枪声识别的重大难题。

③ 采用 EMD 方法将会遇到的难点：有效提取本征模态函数中的特征信息量。经验模态分解产生的本征模态函数不能直接作为特征向量输入到分类器得到，需要进一步处理、提取出能够作为特征向量的参数。

1.4 论文的主要研究内容

通过 1.3 节对公共场所典枪声特征提取及识别的难点分析，本文将从以下几个方面开展公共场所枪声特征提取识别的研究。本文的主要研究内容如下所述：

① 公共场所枪声信号及环境背景噪声特性分析。对典型的几种枪声信号公共及环境下背景噪声进行特性分析。并通过对时域、频域、时频域、倒谱域上枪声信号相关特性的统计分析，得到一些枪声信号具有的共性和不同类型的枪声如手枪、步枪、机枪声之间的差异性。公共场所环境如广场、车站、码头等背景噪声非常嘈杂，常会淹没原始枪声信号的特征，使得公共场所枪声信号变得很不规则，甚至破坏原始枪声信号的时频及其他相关特征。寻找一种更符合公共场所的实际情况的分布替代高斯分布。

② 公共场所枪声特征描述建模。EEMD 通过每次将随机的高斯分布序列人为添加到被分析的信号中，用以解决 EMD 存在的模态混叠问题。受此启发，本文充分考虑公共场所环境下背景噪声分布情况，对 EEMD 进行改进，提出在重构枪声信号时用 $S\alpha S$ 分布代替高斯分布，将重构信号进行 EMD 分解，对多次 EMD 分解得到的 IMF 求整体平均，将其作为原枪声信号的最终 IMF。

③ 公共场所枪声特征提取方法研究。枪声信号经过 EMD 分解后，会生成有限个 IMF 分量，而每个 IMF 分量的频率并不固定。为此，需要寻找新的特征。为此，本文将经过 EMD 分解生成的各阶 IMF 分量进行 FFT，再根据 Parseval 定理求出原信号的能量以及各个 IMF 分量的能量，最终以各个 IMF 分量的能量相对于原信号能量的能量比作为特征向量。

1.5 论文的组织安排

全文共分为 5 章，章节内容安排如下：

第 1 章为绪论，介绍课题的研究背景及意义，着重阐述语音信号与非语音信号特征提取方法的发展和国内外研究现状，特别是枪声特征提取方法的研究，在此基础上总结分析本课题的研究难点，提出本文的主要研究内容。

第 2 章对公共场所典型枪声信号和背景噪声进行特性分析。首先，由于业界没有公认的枪声数据库和公共场所噪声数据库，本文首先自建枪声数据库，并采集背景噪声建立公共场所噪声数据库。通过对时域、频域、时频域、倒谱域上枪声信号相关特性的统计分析，本文得到枪声信号具有非线性、非平稳性的共性。经过对公共场所噪声的特性分析，本文假设公共场所噪声更符合 $S\alpha S$ 分布，并以此替代高斯分布。

第 3 章为全文的重点。首先，介绍作为理论基础的 EMD 和 EEMD 方法。借鉴 EEMD 方法假设环境噪声为高斯分布重构原始信号从而改进 EMD 的思路。本文在此基础上做了如下工作：假设公共场所背景噪声为 $S\alpha S$ 分布，替代高斯分布对原始枪声信号进行重构。其次，推导得出改进 EEMD 的枪声特征描述模型。对多次 EMD 分解得到的 IMF 求整体平均，提出将其作为原枪声信号的最终 IMF，将各阶 IMF 分量进行 FFT，再根据 Parseval 定理求出原信号的能量以及各个 IMF 分量的能量，最终以各个 IMF 分量的能量相对于原信号能量的能量比作为特征向量。

第 4 章为实验结果与分析。首先，对枪声数据进行预处理，由于无法直接在公共场所录制枪声信号，因此采用模拟枪声信号与实际公共场所环境噪声混合的方法得到待分析信号，并进行端点检测和自适应滤波。其次，用第 3 章提出的改进 EEMD 的公共场所枪声特征提取方法得到枪声的特征向量。将该特征向量输入分类器进行分类识别。实验证明，本文提出的改进 EEMD 相比传统的语音处理方法中 MFCC、LPCC 等特征描述参数能够更准确的描述枪声特性。同时，相比原始的 EMD 与 EEMD 方法亦有一定程度的提高，而且鲁棒性更好。

第 5 章为全文总结和工作展望。回顾和总结本文的研究工作，从本课题的相关研究为出发点，为研究工作提出进一步的思路。

2 公共场所典型枪声信号及背景噪声特性分析

2.1 引言

信号的特性分析是信号处理过程中十分重要的环节。信号的特性分析是信号特征提取的前提，只有对信号特性进行深入分析，才有可能得到既可以表征信号本质又具有良好鉴别性的特征。传统信号处理理论研究的对象大多限于平稳信号。但是判断枪声信号这样典型的非语音信号是否是线性的、平稳的需要特性分析来总结其共性了。

本章 2.2 节对公共场所典型枪声信号进行特性分析。首先，由于没有权威的枪声数据库和公共场所噪声数据库，本文首先收集数据自建枪声数据库并独立采集背景噪声建立公共场所噪声数据库。通过对时域、频域、时频域、倒谱域上枪声信号相关特性的统计分析，本文得到一些枪声信号的共性。传统研究假设，在广场、车站、校园公共场所等公共场所环境下的噪声以高斯白噪声和粉红噪声为主，但实际上往往存在一些大气噪声、人为噪声如汽车鸣笛声、脚步声等脉冲噪声的干扰。这类噪声具有非常显著的脉冲特性，并且其统计密度函数具有长拖尾^[43]。本章 2.3 节对公共场所噪声进行特性分析，假设公共场所噪声符合一种新的模型。

2.2 公共场所典型枪声信号的特性分析

目前，业界并没有标准的公共场所枪声数据库，需要自己整理建库。而受环境所限无法在公共场所直接录制枪声，因此本文从网上收集整理 3 大类（手枪、步枪、机枪）共 82 种枪声的模拟数据建立了原始枪声库。该库中共有 3 类枪声，每类有 80 种样本，单个样本的采样率为 44.1KHz，位速为 705Kbps，位数为 16 位，样本长度为 60-120s。公共场所噪声库是通过在三峡广场、重庆市内道路两旁以及火车站录制而建立的。通过将原始枪声库与公共场所噪声库进行混合得到实验用的公共场所枪声信号。

公共场所枪声是一个时间序列，对其进行分析时，时域波形具有直观明了的特点。时域分析主要包括短时能量、短时过零率、短时自相关函数等，这是音频信号的一组最基本的短时参数。常用的频域分析方法有傅立叶变换法，离散余弦变换法等。

2.2.1 典型枪声信号的时域分析

① 短时平均能量及短时平均幅度

假设枪声波形信号为 $x(m)$ ，经过加窗、分帧处理后得到 n 帧信号，第 n 帧为 $x_n(m)$ ，则 $x_n(m)$ 满足下式：

$$x_n(m) = w(m)x(n+m), 0 \leq m \leq N-1 \quad (2.1)$$

$$w(m) = \begin{cases} 1, 0 \leq m \leq N-1 \\ 0, m < 0 \text{ 或 } m > N-1 \end{cases} \quad (2.2)$$

式中, $n=0, T, 2T, \dots, nT$ 并且 N 为帧长, T 为帧移长度。设 E_n 为第 n 帧枪声信号 $x_n(m)$ 的短时能量, 则其计算公式如下:

$$E_n = \sum_{m=0}^{N-1} x_n^2(m) \quad (2.3)$$

E_n 能够很好对枪声信号幅度变化进行度量, 但高电平会影响其结果。为此, 可采用短时平均幅度函数 M_n , 它的定义为:

$$M_n = \sum_{m=0}^{N-1} |x_n(m)| \quad (2.4)$$

M_n 能够表征一帧枪声信号的能量大小, 它与 E_n 的区别在于计算时不会因取平方而造成大小取样值的差异。

② 短时过零率分析

短时过零率是时域分析中最简单的一种特征, 它是指每帧内信号通过零值的次数。对于连续信号, 可以考察时域波形通过时间轴的情况。而对于离散信号, 实质上就是信号采样点符号变化的次数。短时过零率是一种简便而有效的方法, 具有计算迅速的优点。

定义枪声信号 $x_n(m)$ 的短时过零率 Z_n 为:

$$Z_n = \frac{1}{2} \sum_{n=0}^{N-1} |\text{sgn}[x_n(m)] - \text{sgn}[x_n(m-1)]| \quad (2.5)$$

式中, sgn 是符号函数, 即:

$$\text{sgn}[x] = \begin{cases} 1, x \geq 0 \\ -1, x < 0 \end{cases} \quad (2.6)$$

可以将短时平均过零率和短时能量结合起来判断枪声信号的起止点的位置, 即进行端点检测。在背景噪声较小的情况下, 短时能量比较准确, 但当背景噪声较大时, 短时过零率可以获得较好的检测效果。

③ 短时相关分析

一般情况下, 相关函数用于测定两个信号在时域内的相似程度, 可以分为互相关函数和自相关函数。互相关函数主要研究两个信号之间的相关性, 如果两个信号完全不同、相互独立, 那么互相关函数接近零; 如果两个信号的波形相同, 则互相关函数会在超前和滞后处出现峰值, 可据此求出两个信号之间的相似程度。自相关函数主要研究信号本身的同步性、周期性。这里主要讨论自相关函数的性质及应用。

对于枪声信号 $x(m)$, 它的自相关函数定义如下:

$$R_k = \sum_{m=-\infty}^{\infty} x(m)x(m+k) \quad (2.7)$$

如果信号是随机的或是周期的，这时的定义为：

$$R_k = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{m=-N}^N x(m)x(m+k) \quad (2.8)$$

自相关函数具有很多性质：1) 如果信号 $x(m)$ 具有周期性（设周期为 N_p ），那么它的自相关函数也具有周期性，并且周期与信号 $x(m)$ 相同，即 $R(k) = R(k + N_p)$ ；2) 自相关函数是一个偶函数，即 $R(k) = R(-k)$ ；3) 当 $k = 0$ 时，自相关函数取得极大值，即 $R(0) \geq |R(k)|$ 。利用这些性质重新定义短时自相关函数：

$$R_k = \sum_{m=-\infty}^{\infty} x(m)w(n-m)x(m+k)w(n-m-k) \quad (2.9)$$

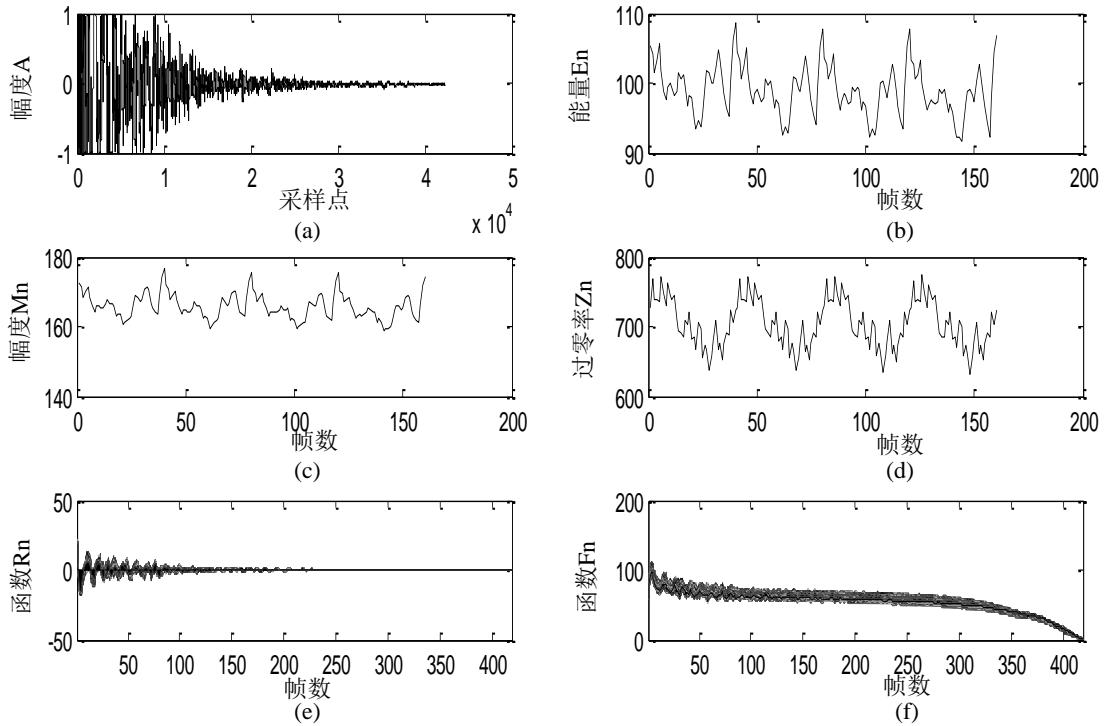


图 2.1 手枪信号时域分析：(a)原始信号波形；(b) 短时平均能量；(c) 短时平均幅度；(d) 短时过零率；(e) 短时相关函数；(f) 平均幅度函数

Fig. 2.1 Pistol signal time domain analysis: (a) Original signal waveform; (b) Short time average energy; (c) Short time average magnitude; (d) Short-time zero crossing rate; (e) A Short time correlation function; (f) average amplitude of function

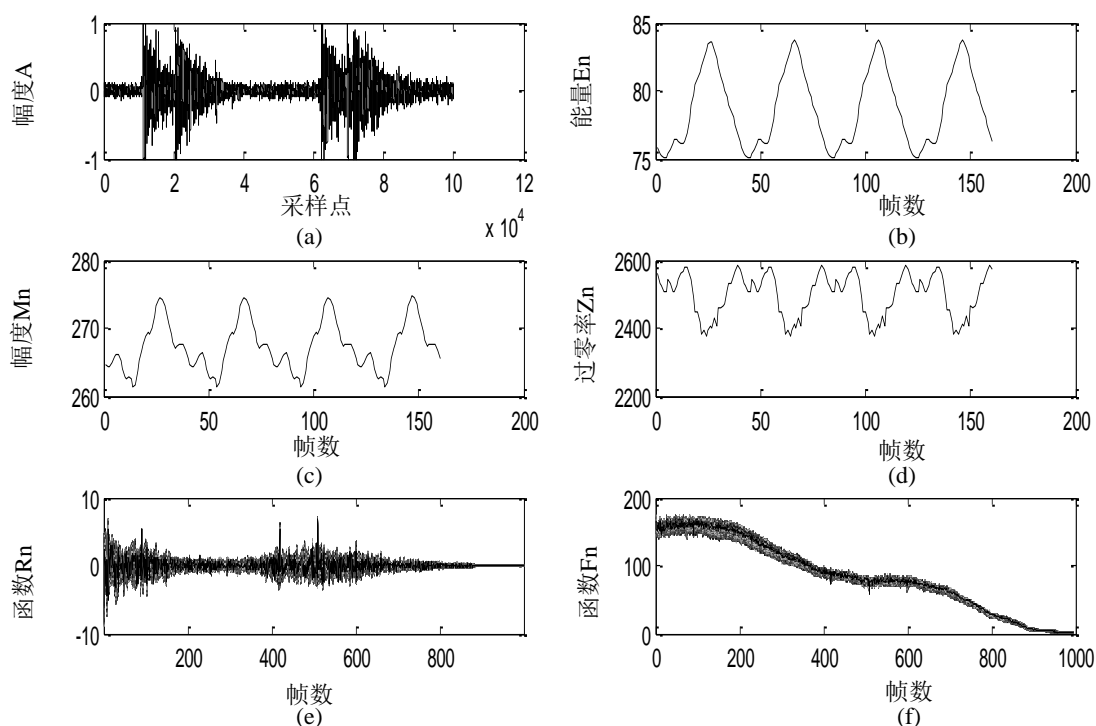


图 2.2 步枪信号时域分析: (a)原始信号波形; (b) 短时平均能量; (c) 短时平均幅度; (d) 短时过零率; (e) 短时相关函数; (f) 平均幅度函数

Fig. 2.2 Rifle signal time domain analysis: (a) Original signal waveform; (b) Short time average energy; (c) Short time average magnitude; (d) Short-time zero crossing rate; (e) A Short time correlation function; (f) average amplitude of function

通过统计三种典型枪声信号的时域特征, 本文发现三种典型枪声信号短时平均能量在 0-500 范围内波动, 短时平均幅度在 0-1000 范围内波动, 短时过零率在 0-5000 范围内波动, 短时平均幅度函数趋势都很相似, 虽然通过时域分析得到了枪声一些共性特征以便与背景噪声区分开来, 但依旧很难单凭时域特性就简单的区分出枪声的种类, 因此本文继续做频域分析以求找到更好的表征枪声信号特征的特征提取方法。

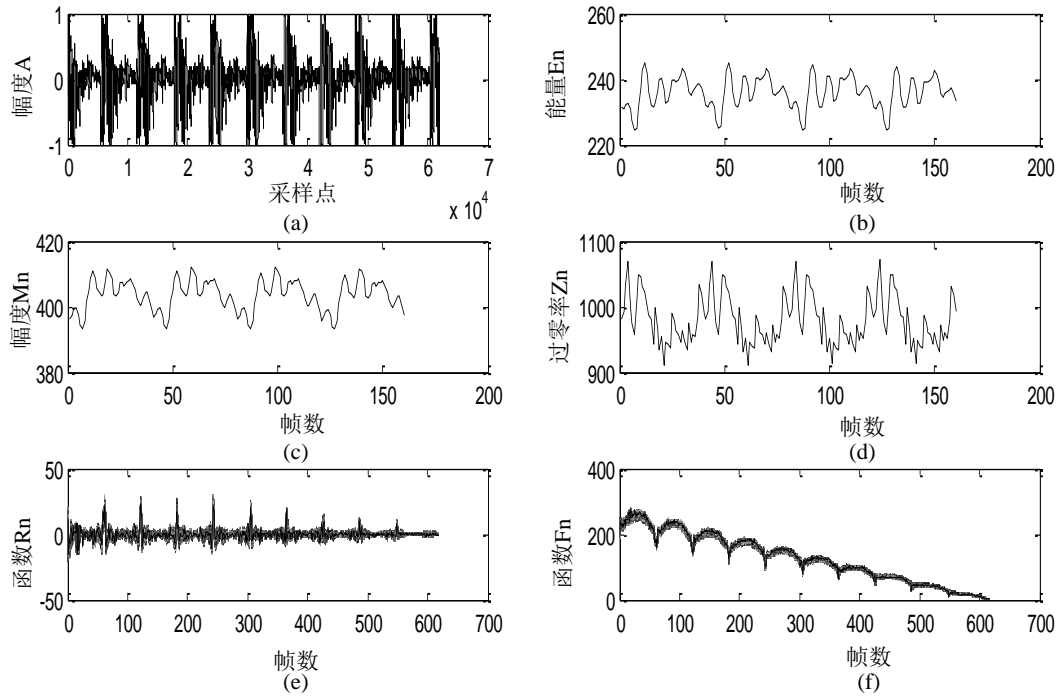


图 2.3 机枪信号时域分析：(a)原始信号波形；(b) 短时平均能量；(c) 短时平均幅度；(d) 短时过零率；(e) 短时相关函数；(f) 平均幅度函数

Fig. 2.3 Machine gun signal time domain analysis: (a) Original signal waveform; (b) Short time average energy; (c) Short time average magnitude; (d) Short-time zero crossing rate; (e) A Short time correlation function; (f) average amplitude of function

2.2.2 典型枪声信号的频域分析

① 傅里叶变换

傅里叶变换是傅里叶频谱分析的基础，傅里叶谱自相关函数、功率谱、倒谱等都可以用傅里叶变换及其反变换求得。与时域特征分析相似，本节也采用相同的短时分析技术来进行傅里叶频谱分析。

定义信号 $x(n)$ 的短时傅里叶变换为：

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-j\omega m} \quad (2.10)$$

式中， $w(n)$ 为窗口函数。在实际计算中，一般用离散傅里叶变换代替连续傅里叶变换，这就需要对信号进行周期性扩展，即把 $x(n)w(n)$ 看成某个周期信号的一个周期，然后对它做离散傅里叶变换，这是得到的是功率谱。短时功率谱实际上是短时傅里叶变换的平方，不难证明，它是信号 $x(n)$ 的短时自相关函数的傅里叶变换，即

$$P_n(e^{j\omega}) = |X_n(e^{j\omega})|^2 = \sum_{k=-\infty}^{\infty} R_n(k) e^{j\omega k} \quad (2.11)$$

式中, $R_n(k)$ 是前面讨论的自相关函数。

② 离散余弦变换

作为与傅里叶变换相关的一种变换, 离散余弦变换与离散傅里叶变换十分类似, 区别是离散余弦变换只使用了实数。离散余弦变换是根据下面的公式把 n 个实数 $x(0), x(1), \dots, x(n-1)$, 变换到另外 n 个实数 $D_{-1}, D_0, D_1, \dots, D_{n-1}$ 的操作。

$$D_m = \sum_{k=0}^{n-1} x(k) \cos \left[(k+0.5) \frac{\pi}{n} m \right], 0 \leq m \leq n-1 \quad (2.12)$$

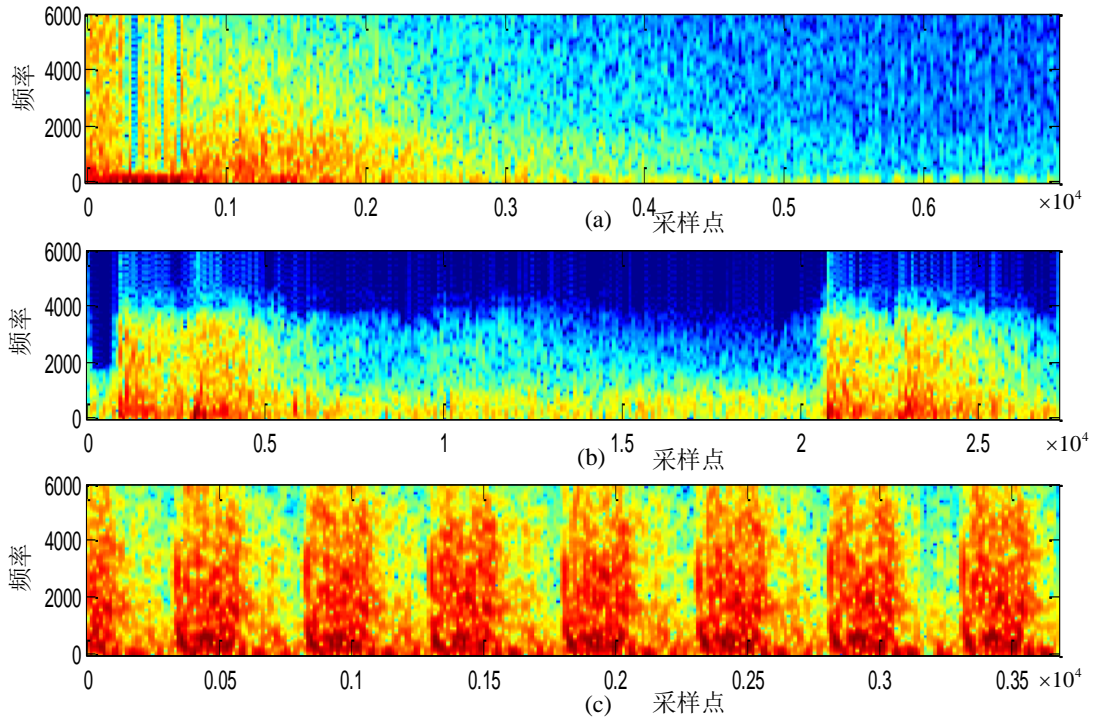


图 2.4 三种典型枪声信号声谱图: (a)手枪声信号声谱图; (b) 步枪声信号声谱图; (c) 机枪声信号声谱图

Fig. 2.4 Three kinds of typical sound signal spectrograms: (a) Pistol signal spectrograms; (b) Rifle signal spectrograms; (c) Machine gun signal spectrograms

图 2.4 是三种典型枪声信号的声谱图。从枪声声波形成原理来看, 弹丸被火药推出后, 在枪管中高速前进产生摩擦并且挤压空气, 产生弹道声波。谱图中具有浓郁颜色的能量分布主要就表现为弹道声波, 能量主要集中在 0-6000Hz 之间。从枪口射出的声波对空气分子作用形成非线性扰动, 形成一些能量值较小的杂波。

通过统计三种典型枪声信号的频域分析, 本文发现三种典型枪声信号频域特性都很相似而且由于枪声信号具有非线性、非平稳性的特征, 因此很难单凭频域特性就简单的区分出枪声的种类, 因此本文继续做时频域分析以求找到更好的表

征枪声信号特征的特征提取方法。

2.2.3 典型枪声信号的时频域分析

一般的信号都是随着时间的变化而发生变化,然而要深入理解信号的本质,需要研究信号的不同表现方式。时域和频域是观察一个信号的两种方式,时域分析和频域分析技术也是目前信号处理的主要方法。时域分析方法完全是在时间域中分析信号,时间分辨率理论上可以达到无穷大,但频率分辨率零,而频域分析方法则相反。一般在频域里分析信号可以得到更多的信息,因此以往人们更重视在频域内对信号加以分析。通过统计 2.2.1 和 2.2.2 节的分析,本文发现所枪声信号并不是平稳的,至少在观测的全部时间段内它不是平稳的,所以为了有效提取典型枪声的特征需要进行时频域分析。

① Gabor 变换

Gabor 变换的基本思想是:将信号用很多小的时间间隔分隔开,为了确定信号在每个时间间隔存在的频率对其用 Fourier 变换进行分析。具体方法是对信号 $f(t)$ 加一个滑动窗 $w(t-\tau)$ (τ 是移位因子,代表滑动窗的位置)后,再作 Fourier 变换,即

$$SF_f(\omega, \tau) = \int f(t)w(t-\tau)e^{-j\omega t} dt \quad (2.13)$$

Gabor 变换虽然在一定程度上克服了 Fourier 变换不具有局部分析能力的问题,但它自身存在着不可克服的缺陷。公式 2.13 也可以看成是 $f(t)$ 和 $g(t) = w(t-\tau)$ 的内积,即

$$SF_f(\omega, \tau) = \int f(t)w(t-\tau)e^{-j\omega t} dt = \langle f(t), w(t-\tau)e^{j\omega t} \rangle = \langle f(t), g(t) \rangle \quad (2.14)$$

假设 $w(t)$ 是高斯型函数,当 ω 从 ω_0 变成 $2\omega_0$ 时, $G(\omega)$ 的中心频率也随之变化,而且它的品质因数 Q 是变化的。Gabor 变换的 Q 值随 ω 变化而变化,所以 Gabor 变换所得结果不能准确地反映信号的特性。Gabor 变换的不足之处还在于, $G(\omega)$ 的带宽保持不变,一旦窗口函数选定后,时频窗口的形状便保持不变,割断了频率与窗口宽度的内在联系, Gabor 变换实质是具有单一分辨率的分析。而且可以证明,无论 ω 和 τ 如何离散化, $g'(\omega)$ 、 $\tau(t)$ 都不能形成 $L^2(R)$ 上的正交基。因此,为了不丢失信息,在信号分析或数值计算时必须采用非正交的冗余基,这就增加了不必要的计算量和存储量越小越严重。此外,对原信号 $f(t)$ 施加一个窗口函数 $w(t-\tau)$,必然会导致原信号 $f(t)$ 的 Fourier 频谱失真,这也就是 Gabor 变换的内在缺陷。

② 小波变换

小波变换对 Gabor 变换的局部化思想予以继承并进一步发展,改进了 Gabor 变换的缺点如不恒 Q 、缺乏离散正交基等。特别地,当分析对象是非平稳信号时,需要多分辨率分析:当信号波形变化剧烈时,要求窗口在时间轴上要窄一些,以得到较高的时间分辨率;当信号波形变化平缓时,要求窗口在频率轴上要窄一些,

以得到较高的频率分辨率，这样细致的操作是 Fourier 变换和 Gabor 变换都无法做到。

设 $x \in L^2(R)$ ，则称 $\psi(t)$ 为基本小波或母小波的函数，则

$$WT_x(a, \tau) = \frac{1}{\sqrt{a}} \int x(t) \psi\left(\frac{t-\tau}{a}\right) dt = \langle x(t), \psi_{a\tau}(t) \rangle \quad (2.15)$$

称为 $x(t)$ 的小波变换。式中 $a > 0$ 是尺度因子， τ 代表位移，正负皆可。尺度因子 a 能对基本小波 $\psi(t)$ 进行伸缩， a 越大 $\psi(t/a)$ 越宽，即小波的持续时间在不同尺度下随 a 增大而增宽。可以证明，小波变换的等效频域表示为

$$WT_x(a, \tau) = \frac{\sqrt{a}}{2\pi} \int X(\omega) \psi(a\omega) e^{j\omega\tau} d\omega \quad (2.16)$$

由此可见，如果 $\psi(\omega)$ 是带通函数且幅频特性比较集中，则小波变换便能够表征待分析信号 $X(\omega)$ 频域上的局部性质。例如 Morlet 小波的频谱是高斯函数且中心频率在 ω_0 ，如此则通过改变 ω_0 就可以对 $X(\omega)$ 在 ω_0 附近的局部性质进行表征。各个 $\psi(a\omega)$ 的中心频率和带宽会随 a 值的变化而变化，而品质因数却不会。因此，小波变换可以达到多分辨率分析的效果。虽然分析频率的高低不尽相同，但却可以保持品质因数 Q 不变，所以幅频特性就可以被准确的反映出来。这是小波变换比 Gabor 变换有优势的特点，并且这一优点也很符合实际工作需要，因为越要在时域上观察的更细致，就越要提高分析频率并缩小观察范围。

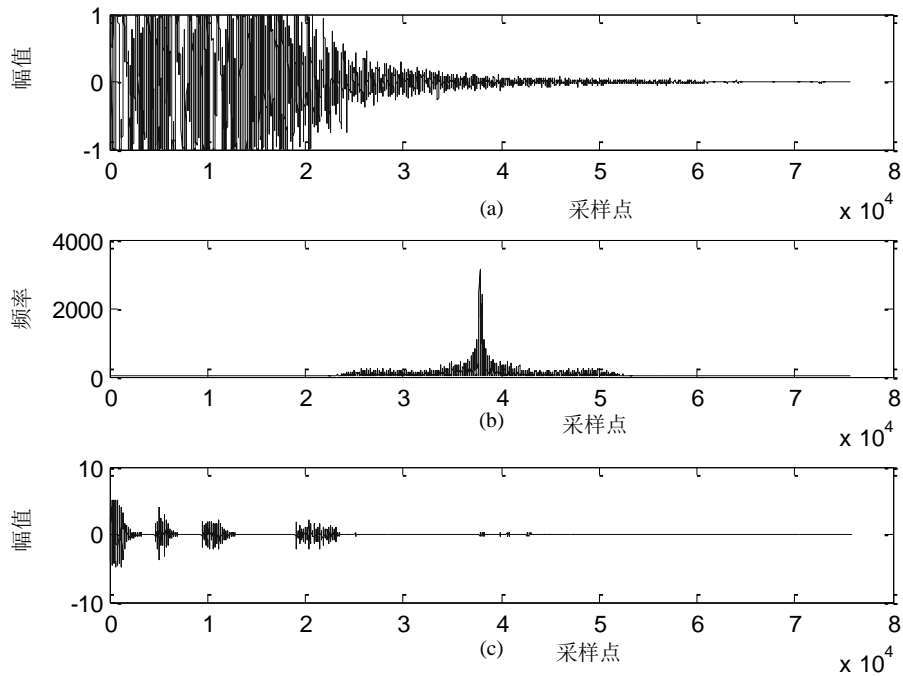


图 2.5 手枪信号时频域分析：(a)原始信号波形；(b) Gabor 变换；(c) 小波变换

Fig. 2.5 Pistol signal time-frequency domain analysis: (a) Original signal waveform; (b) Gabor transform; (c) Wavelet transform

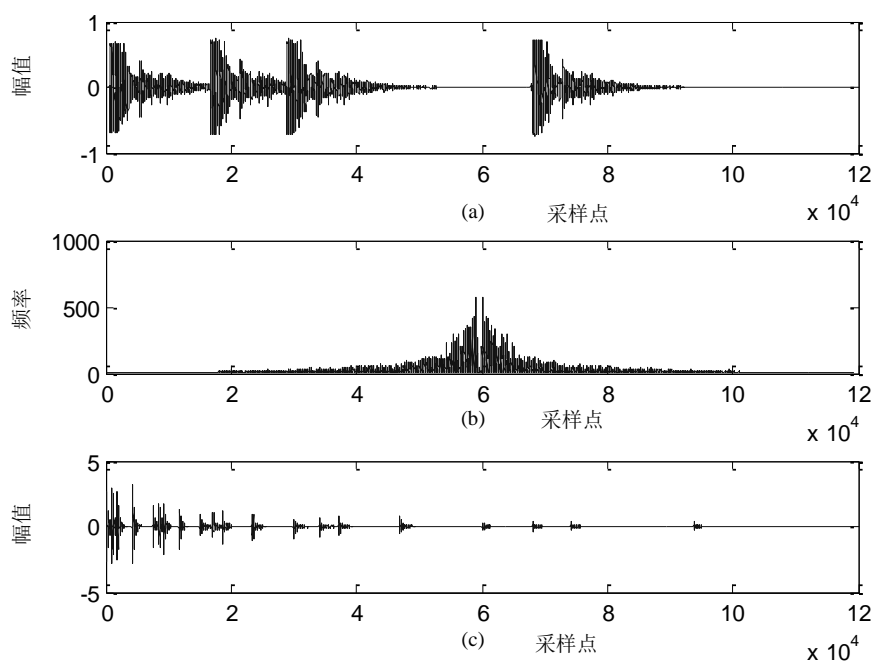


图 2.6 步枪信号时频域分析: (a)原始信号波形; (b) Gabor 变换; (c) 小波变换

Fig. 2.6 Rifle signal time-frequency domain analysis: (a) Original signal waveform; (b) Gabor transform; (c) Wavelet transform

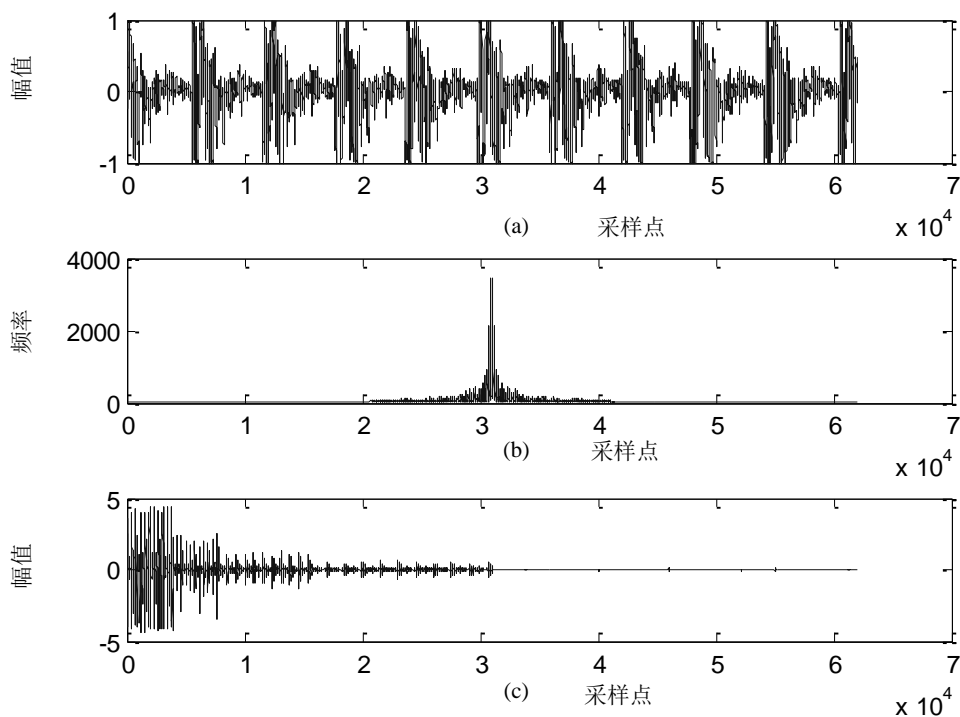


图 2.7 机枪信号时频域分析: (a)原始信号波形; (b) Gabor 变换; (c) 小波变换

Fig. 2.7 Machine gun signal time-frequency domain analysis: (a) Original signal waveform; (b) Gabor transform; (c) Wavelet transform

通过对三种典型枪声信号的时频域分析进行统计, 本文发现三种典型枪声信号 Gabor 变换非常相似, 频率在 0-4000Hz 之间而且有明显的脉冲特性, 而通过统计 4 阶小波变换发现枪声具有非线性的共性特征, 虽然时频域分析方法可以解决非平稳信号的问题但是由于枪声是非线性的, 因此很难单凭时频域特性就简单的区分出枪声的种类, 因此本文继续做倒谱域分析以求找到更好的表征枪声信号特征的特征提取方法。

2.2.4 典型枪声信号的倒谱分析

通常的加性信号可以用线性系统来处理, 这种系统是满足线性叠加原理的。显然通过 2.2.1-2.2.3 节的分析本文知道枪声信号中各组成分量的组合方式并不是加性的而是乘积性或卷积性的。显然, 线性系统对这种信号不在适用, 而需要非线性系统且满足该组合规则才行。但是非线性系统相比线性系统更为复杂, 需通过同态分析将非线性问题转化为线性问题来处理。对音频信号进行同态分析后, 将得到音频信号的倒谱参数, 因此同态分析也称为倒谱分析。倒谱分析中最常用的两种分析方法是线性预测倒谱分析和 Mel 频率倒谱分析。

① MFCC

由于 MFCC 参数将人耳的听觉感知特性很好的表达出来, 因此被广泛应用^[33]。提取及计算过程如下:

1) 原始枪声信号 $s(n)$ 经过预加重、分帧、加窗等处理, 得到每个信号帧的时域信号 $x(n)$ 。

2) 将时域信号 $x(n)$ 补零以形成长为 $N=512$ 的序列, 然后经过离散傅立叶变换 (Discrete Fourier Transform, DFT) 后得到线性频谱 $X(k)$, 转换公式:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi nk/N} \quad 0 \leq n, k \leq N-1 \quad (2.17)$$

在实际应用中, 常常用 FFT 替代 DFT 过程进行运算, 其中称 N 为窗宽。

3) 将上述线性频谱 $X(k)$ 转成 Mel 频谱, 并计算其对数, 得到对数频谱 $S(m)$ 。在枪声的频谱范围内设立一系列带通滤波器 $H_m(k)$, $0 \leq m < M$, M 为滤波器的个数。这些滤波器都是三角形滤波, 其中心频率为 $f(m)$, 相邻 $f(m)$ 之间的间隔随着 m 值变化而变化, 传递函数为:

$$H_m(k) = \begin{cases} 0, k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)}, f(m-1) \leq k \leq f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)}, f(m) < k \leq f(m+1) \\ 0, k > f(m+1) \end{cases}, 0 \leq m < M \quad (2.18)$$

$f(m)$ 可以用下面的方法加以定义:

$$f(m) = (N/F)B^{-1}(B(f_l) + m \frac{B(f_h) - B(f_l)}{M+1}) \quad (2.19)$$

式中 f_l 、 f_h 为滤波器的频率应用范围的最低频率和最高频率， F_s 代表采样频率， B^{-1} 是 B 的逆函数。为了提高结果对噪声和谱估计误差的鲁棒性，将上述经过 Mel 滤波器组得到的 Mel 频谱取对数能量。

则由线性频谱 $X(k)$ 到对数频谱 $S(m)$ 的总传递函数为

$$S(m) = \ln(\sum_{k=0}^{N-1} |X(k)|^2 H_m(k)), 0 \leq m < M \quad (2.20)$$

4) 将上述对数频谱 $S(m)$ 经过离散余弦变换到倒频谱域，即可得到 Mel 频率倒谱系数 $c(n)$ ：

$$c(n) = \sum_{m=1}^{M-1} S(m) \cos(\frac{\pi n(m+1/2)}{M}), 0 \leq m < M \quad (2.21)$$

② LPCC

为了求 LPCC，先要对枪声信号做线性预测分析，得到线形预测系数(Linear Prediction Coefficient, LPC)，再通过叠带算法求取枪声信号的 LPCC，作为区分不同种类枪声的信号特征参数^[34]。传输函数为：

$$H(Z) = \frac{G}{1 - \sum_{i=1}^P a_i Z^{-i}} \quad (2.22)$$

式中， P 代表预测器的阶数， G 代表滤波器的增益。因此，声音抽样 $S(n)$ 和激励信号 $e(n)$ 之间的关系可以用下列的差分方程来表示：

$$s(n) = Ge(n) + \sum_{i=1}^P a_i s(n-i) \quad (2.23)$$

上式表示枪声样点前后有相关性，可以用前 p 个样点值预测之后的样点值。

$$\tilde{s}(n) = \sum_{i=1}^P a_i s(n-i) \quad (2.24)$$

式中， a_i 即为线性预测系数。从而 P 阶线性预测器的系统函数可用下式表示：

$$P(z) = \sum_{i=1}^P z^{-i} \quad (2.25)$$

线性预测系数 a_i 的求解过程就是对枪声的 LPC 分析过程，也即给定枪声序列，使预测误差在最小均方误差准则下最小，从而得到最佳估值 \mathbf{n} 。某一帧内的预测误差为：

$$\varepsilon(n) = s(n) - \hat{s}(n) = s(n) - a_i s(n-i) \quad (2.26)$$

定义短时平均预测误差为：

$$E\{\varepsilon^2(n)\} = E\{[s(n) - a_i s(n-i)]^2\} \quad (2.27)$$

为使 $E\{\varepsilon^2(n)\}$ 最小，对 a_i 求偏导，并令其为零，有

$$E\{[s(n) - a_i s(n-i)] \cdot s(n-j)\} = 0, j = 1, \dots, p \quad (2.28)$$

倒谱本来应该是信号 Z 变换的对数模函数的反 Z 变换，一般通过信号的傅里叶变换取模的对数，再求反傅里叶变换得到。通过自相关法可以得到关于 a_i 的一组线性方程组，可以用下述递推公式对方程组进行求解得到相应的倒谱系数：

$$h(1) = a_1, h(n) = \begin{cases} a_n + \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) a_k h(n-k), & k < n \leq p \\ h(n) = \sum_{k=1}^p \left(1 - \frac{k}{n}\right) a_k h(n-k), & n \geq p \end{cases} \quad (2.29)$$

式中， $\{a_1, a_2, \dots, a_p\}$ 为 LPC 系数， $h(n)$ 为 LPCC 倒谱系数。

通过对三种典型枪声信号的倒谱域分析进行统计，本文发现三种典型枪声信号的 MFCC 系数能够一定程度的描述枪声的特征，而 LPCC 系数也对枪声特征进行有效的提取。

总而言之，通过统计三种典型枪声信号的时域、频域、时频域，本文发现三种典型枪声信号这些特征都很相似并且假设枪声信号的共性是具有非线性、非平稳性的特征，很难单凭时域特性就简单的区分出枪声的种类。

2.3 公共场所背景噪声的特性分析

2.3.1 公共场所背景噪声的特性分析

公共场所背景噪声是由各种类型声源叠加形成，并非完全服从高斯分布。比如步行街广场的背景声音就由小孩子的欢笑声、口哨声、汽车的鸣笛声、路人的谈话声等叠加形成。不同的声源有不同的频率范围，经过叠加之后，均匀地覆盖了 4000Hz 以下的低频段。

从理论上可知，分数低阶 α 稳定分布过程的显著特性就是存在较多脉冲状噪声与长拖尾特性，而这一点与公共场所背景噪声的特性十分相似。为此，本文将采用分数低阶 α 稳定分布噪声模型对公共场所存在的噪声进行理论分析^[43]。

2.3.2 公共场所背景噪声分布模型

α 稳定分布不同于其他统计模型，一般情况下 α 稳定分布的概率密度函数无法用统一的解析式表达。而往往是用特征函数的 4 个参数 $(\alpha, \beta, \gamma, \mu)$ 来描述：

$$\begin{cases} \varphi(t) = \exp\{j\mu t - \gamma |t|^\alpha [1 + j\beta \operatorname{sgn}(t)\omega(t, \alpha)]\} \\ \omega(t, \alpha) = \begin{cases} \tan \frac{\alpha\pi}{2}, & (\alpha \neq 1) \\ \frac{\pi}{2} \log |t|, & (\alpha = 1) \end{cases} \end{cases} \quad (2.30)$$

式中， α 称为特征指数 ($0 < \alpha \leq 2$)，概率密度函数的拖尾情况由 α 控制， α 值越小其分布的脉冲性越强，拖尾越长。 β 称为对称系数， $\beta = 0$ 时表示对称分布，

记为 $S\alpha S$ 分布, 本文主要研究这种分布。 μ 是位置参数($-\infty < \mu < +\infty$), 对应于中值或均值。 γ 为分散系数, 与高斯分布的方差类似。

当 $\alpha = 2$ 时, α 稳态分布就是高斯分布, 也就是说高斯分布是 α 稳态分布的一种特例。当 $0 < \alpha < 2$ 时, $S\alpha S$ 分布的特性相比高斯分布就明显不同。较明显特点是分布上有长拖尾即在远离均值或中值处的样本数较多, 体现在时间域波形上就是较多的尖峰脉冲。

中心极限定理阐明了随着有限方差的随机变量数量增长, 它们的和的分布趋向正态分布。假定服从 $S\alpha S$ 分布的公共场所噪声, α 值的大小随着拖尾厚度和环境噪声统计分布中尖峰脉冲的变化而变化。在图 2.11 中显示了公共场所背景噪声与 α 值不同时 $S\alpha S$ 分布过程的离散实现。从图 2.11 可以看出公共场所环境下的背景噪声与 $\alpha = 1.8$ 时的 $S\alpha S$ 分布非常接近, 都带有比较明显的脉冲值。

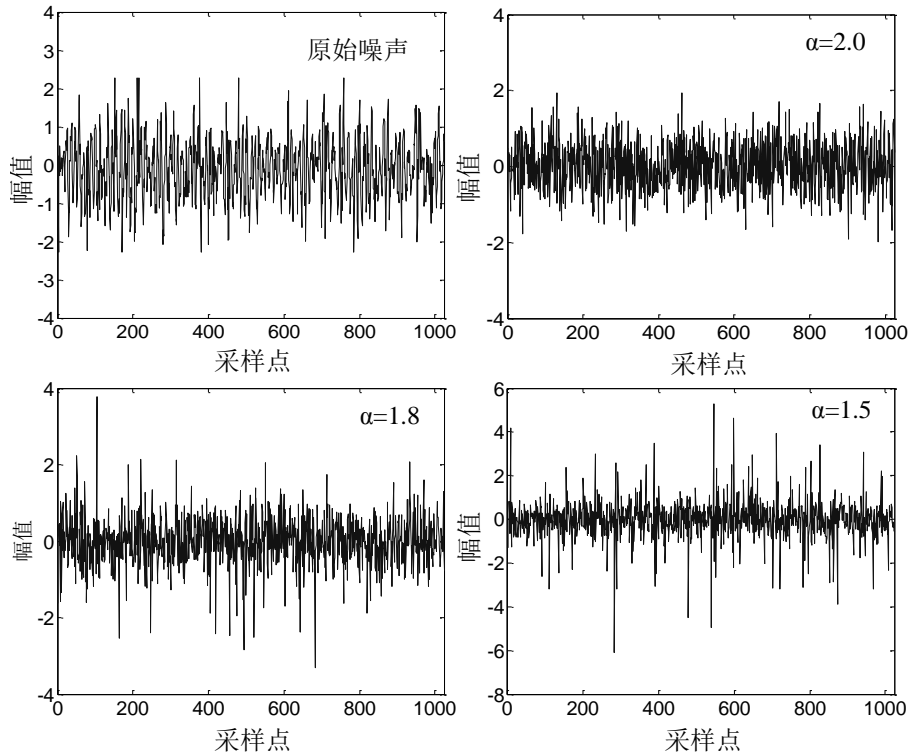


图 2.11 公共场所背景噪声与 $\alpha = 2.0, 1.8, 1.5$ 时的 $S\alpha S$ 分布过程的离散实现

Fig. 2.11 Waveform graph of noise under public place, discrete realization of $S\alpha S$ when $\alpha = 2.0, 1.8, 1.5$

通过公共场所背景噪声特性分析, 本文假设 $S\alpha S$ 分布与传统高斯分布相比能更好的表达公共场所背景噪声的分布, 据此重新构建公共场所背景噪声模型以适应公共场所枪声特征提取, 具体内容将在第 3 章展开。

2.4 本章小结

本章对公共场所典型枪声信号和背景噪声进行了特性分析。首先，由于没有权威的枪声数据库和公共场所噪声数据库，本文首先收集数据自建了枪声数据库并独立采集背景噪声建立公共场所噪声数据库。通过对时域、频域、时频域、倒谱域上枪声信号相关特性的统计分析，本文得到枪声信号具有非线性、非平稳性的共性。而公共场所噪声的特性分析也是不可或缺的，经过分析本文假设公共场所噪声为 $S\alpha S$ 分布，并替代高斯分布建立新的公共场所背景噪声模型。

3 公共场所枪声特征描述建模与特征提取方法研究

3.1 引言

特征是要揭示事物的本质，而且要对相近事物具有区分性。正如第 1 章所述，目前对非语音信号处理大多沿用了传统语音信号处理的方法，但由于针对性不强，效果并不尽如人意。而通过第 2 章对公共场所典型枪声的特性分析，本文归纳出枪声这种典型的非语音信号最重要的特性：非线性和非平稳的特性。因此选择一种针对性强的、适合枪声的特征提取方法是能否有效识别枪声的前提。本章 3.2 节介绍作为理论基础的 EMD 和 EEMD。本章 3.3 节是本章的重点，借鉴 EEMD 方法中假设环境噪声为高斯分布重构原始信号，从而改进 EMD 的思路。本文在此基础上做了如下工作：本文假设公共场所背景噪声为 $S\alpha S$ 分布，并以此替代高斯分布，重构原始枪声信号。其次，推导出改进 EEMD 的枪声特征描述模型。将所有 EMD 分解得到的 IMF 做整体平均，结果作为最终 IMF，将各阶 IMF 分量进行 FFT，再根据 Parseval 定理求出原信号的能量以及各个 IMF 分量的能量，最终以各个 IMF 分量的能量相对于原信号能量的能量比作为特征向量。本章 3.4 节简要介绍两种本文用到的分类器。

3.2 经验模态分解及其改进

3.2.1 EMD

EMD 是 N.E.Huang 于 1998 年提出的一种新的自适应信号处理方法，该方法基于信号局部极值特征，淡化先验信息，能够有效的处理非线性、非平稳信号。EMD 能够反映信号的本质特征并且具有自适应的信号分解和降噪能力。EMD 分解的实质就是一组滤波过程，它能逐级分解信号将其不同尺度的波动或趋势分开，EMD 分解产生的数据序列具有不同的频率尺度，这些序列被称为 IMF，每个 IMF 体现了包含在原信号中不同频率尺度的振荡特性^[35]。EMD 方法一经提出后，很快被用于雷达、医学、超声、机械故障诊断、地震信号处理分析、海洋等不同领域的信号处理中。

EMD 是一种循环迭代算法，它将任一时间信号 $x(t)$ 分解成有限个 IMF 和一个余项的和，三个基本假设是 EMD 分解的前提：

- ① 信号有一个以上的极大值和极小值点；
- ② 极值点间的时间推移定义特征的时间尺度；
- ③ 如果整个信号没有极值点只有曲折点，先对信号微分一次或多次找到极值点，再进行积分运算得到最后的结果。

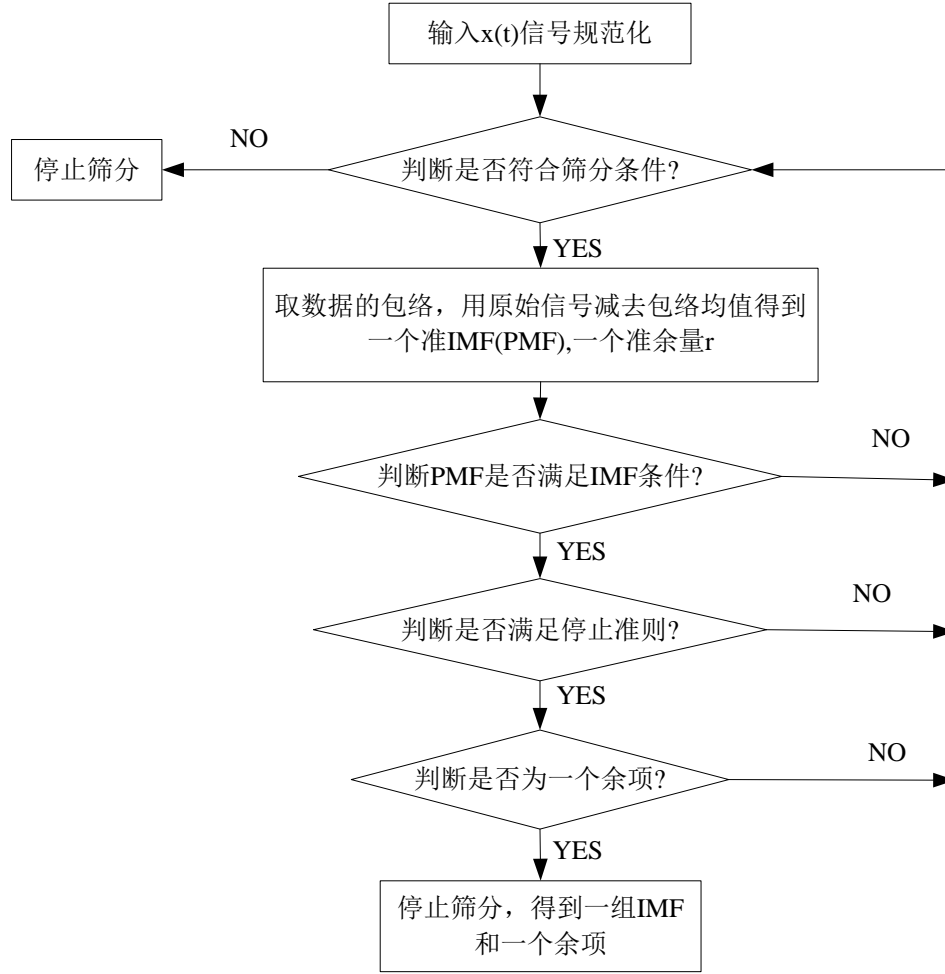


图 3.2 EMD 算法流程图

Fig.3.2 EMD Algorithm flow chart

分解成的有限个 IMF 反映了信号的内部特征, 一般低阶 IMF 频率较高, 随着 IMF 阶数的升高频率逐渐降低, 至余项几乎为 0。判断 IMF 需要满足以下两个条件^[36]:

- ① 信号极值点与零点的数目之差小于等于 1;
- ② 信号关于时间轴局部对称, 即无论在哪一点处, 两条包络线 (分别由局部极大值点和局部极小值点确定) 的平均值为零。

对信号 $x(t)$ 进行 EMD 分解的具体步骤如下:

- ① 首先确定信号 $x(t)$ 上所有的极大、极小值点, 然后用三次样条曲线将所有极大、极小值点分别连接起来, 信号所有点都在这两条曲线之间, 这两条曲线称为 $x(t)$ 的上下包络线 $x_{\max}(t)$ 和 $x_{\min}(t)$ 。称上下包络线的均值为平均包络线 $m_1(t)$ 。

- ② 将信号 $x(t)$ 与平均包络线 $m_1(t)$ 相减, 得到一个新序列 $h_1(t)$, 即:

$$h_1(t) = x(t) - m_1(t) \quad (3.1)$$

③ 判断 $h_1(t)$ 是否满足 IMF 条件, 若不满足, 则再将 $h_1(t)$ 替换原信号 $x(t)$, 重复步骤①、②, 得到 $h_{11}(t)$, 即:

$$h_{11}(t) = h_1(t) - m_{11}(t) \quad (3.2)$$

假设筛选 k 次后 $h_{1k}(t)$ 满足 IMF 条件, 即:

$$h_{1k}(t) = h_{1(k-1)}(t) - m_{1k}(t) \quad (3.3)$$

这样称 $h_{1k}(t)$ 为第一阶本征模态函数, 记为:

$$c_1(t) = h_{1k}(t) \quad (3.4)$$

④ 从信号中减去 $c_1(t)$, 得到第一阶剩余信号 $r_1(t)$, 即:

$$r_1(t) = x(t) - c_1(t) \quad (3.5)$$

⑤ 将 $r_1(t)$ 看作新的初始信号重复执行步骤①至④, Huang 将这一处理过程称为筛选过程。筛选过程一直持续直到余项 $r_n(t)$ 为单调函数 (不可再被分解) 为止。这样原信号即可由各阶 IMF 分量以及余项 $r_n(t)$ 表示:

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (3.6)$$

3.2.2 EEMD

随着研究的深入, 虽然 EMD 自问世以来已经被各领域学者成功地用于处理多种非语音信号。但是 EMD 仍存在一些问题, 最严重的缺点就是模态混叠现象会导致分解出的 IMF 失真, 严重影响信号的特征提取效果。

所谓模态混叠是指一个 IMF 中包含有多个频率尺度的信号分量, 或者不同的 IMF 中包含了同一频率尺度的信号分量, 结果表现为相邻两个 IMF 的波形互相混叠, 相互影响以至难以辨别, 模态混叠是信号间歇性的表现^[36]。这种间歇性往往由噪声干扰引起, 有可能导致信号时频分布的严重偏移, 或导致无法明确 IMF 的物理意义。为了解决这一问题, Huang 的小组在用 EMD 分解白噪声信号时得到一些规律, 他们发现白噪声中各种频率是均匀分布的, 分解结果中的每个 IMF 的平均频率近似等于后一个的两倍并且其功率谱都呈现相同的带通特性。而用 EMD 分解非噪声信号时, 往往一些尺度会缺失, 这时就会发生模态混叠。在此基础上, Huang 提出了一种改进的 EMD 方法——EEMD 方法^[37]。

EEMD 是一种与众不同的数据分析方法, 该方法每次将原始信号与不同的随机白噪声混合, 并对重构信号进行若干次 EMD 分解, 对得到的所有 IMF 分量求总体平均, 得到最终的 IMF 作为原始信号的基函数。相比于 EMD, EEMD 做到了消除模式混合问题, 使得 IMF 的物理意义更明确。EEMD 的前提是以下两个观点:

① 在多次加噪分解试验后, 原始信号中有用的特征被完整保留下来, 而人为添加的随机高斯白噪声通过总体平均的方式相互抵消;

② 由于 EMD 是自适应的, 多次添加适度随机白噪声就可以将尽可能多的尺

度分析引入原始信号并改变信号中原有时间跨度，EMD 能够根据这些变化重新感知信号^[38]。

经过总体平均得到的 IMF 能更好的体现信号中各组分的内涵，即 EEMD 输出的 IMF 较真实的反映了对象内在本质过程，更客观更有意义。具体分 5 步完成：

- ① 初始化 EMD 执行的总次数 M ，加入白噪声信号；
- ② 重构信号 $x^m(t) = x(t) + w^m(t)$ ，式中 $w^m(t)(m=1, \dots, M)$ 是与 $x^m(t)$ 同等长度的随机高斯白噪声；
- ③ 将添加了白噪声的信号进行 EMD 分解，即对每个 $x^m(t)(m=1, \dots, M)$ 做经验模态分解，得到所有的本征模态函数 $IMF_k^m(t)(m=1, \dots, M)$ 其中 k 是 IMF 的阶数；
- ④ 若 $m < M$ ，返回步骤②、③， $m = m + 1$ ，重复执行上述步骤①、②，随机白噪声序列不断变化，直至进行 M 次分解为止，得到所有的 IMF；
- ⑤ 对 M 次 EMD 分解得到的所有 IMF 求整体平均，即对 $IMF_k^m(t)$ 求总体平均值，得到 \overline{IMF}_k ，将其作为原信号的最终 IMF：

$$IMF(t) = \frac{1}{M} \sum_{m=1}^M IMF_k^m(t) \quad (3.7)$$

3.3 公共场所枪声特征描述建模及特征提取方法

3.3.1 公共场所环境下枪声信号重构

EEMD 方法理论基础是假设噪声模型符合高斯分布，而通过公共场所噪声特性分析本文假设 $S\alpha S$ 分布更符合实际公共场所的背景噪声，由于 $S\alpha S$ 分布也符合广义中心极限定理，高斯分布与 $S\alpha S$ 分布序列的本征模态函数的总体均值都为零，因此原理上与 EEMD 方法一致，即借鉴 EEMD 的思路结合公共场所噪声模型用 $S\alpha S$ 分布替代高斯分布对枪声信号重构。具体方法：

找出真实公共场所噪声特征指数 α 的最大似然估计值， $S\alpha S$ 分布的密度函数可以写为：

$$f_\alpha(x) = \frac{\alpha}{|1-\alpha|\pi} x^{1/(\alpha-1)} \int_0^{\pi/2} v(\theta) \exp[-x^{\alpha/(\alpha-1)} v(\theta)] d\theta, (\alpha \neq 1, x > 0) \quad (3.8)$$

式中，

$$f_1(x) = \frac{1}{\pi(1+x^2)}, \quad f_2(x) = \frac{1}{2\sqrt{\pi}} e^{-x^2/4}, \quad v(\theta) = \frac{1}{\sin(\alpha\theta)^{\alpha/(\alpha-1)}} \cos[(\alpha-1)\theta] \cos(\theta)^{1/(\alpha-1)}$$

因此，可以通过观测值 x_1, x_2, \dots, x_N 利用求下一对数似然函数的最大值的方法来估计参数：

$$\begin{aligned} \sum_{i=1}^N \log[f_{\alpha}(z_i)] &= N \log \alpha - N \log(\alpha - 1) \\ &+ \sum_{i=1}^N (\log z_i) / (\alpha - 1) + \sum_{i=1}^N \log \int_0^{\pi/2} v(\theta) \exp[-z_i^{\alpha/(\alpha-1)} v(\theta)] d\theta \end{aligned} \quad (3.9)$$

式中 $z_i = |x_i - \alpha|/c$ ，由此产生 $\alpha(0 < \alpha \leq 2)$, $a = 0$, $-1 < \beta < 1$ 的稳定分布序列 X 。

若 $\alpha = 1$ ，则定义 $\beta_A = \beta$ ；

若 $\alpha \neq 1$ ，则定义 $k(\alpha) = 1 - |1 - \alpha|$ ， $\beta_A = 2 \tan^{-1}(\beta / \cot(\pi\alpha/2)) / (\pi k(\alpha))$ ， $\Phi_0 = -0.5\pi\beta_A(k(\alpha)/\alpha)$ 。

分别产生两个独立的随机变量 \mathbf{V} 和 \mathbf{W} ，其中 \mathbf{V} 满足 $(-\pi/2, \pi/2)$ 范围内均匀分布， \mathbf{W} 满足均值为 1 的指数分布。定义以下变量：

$$\begin{aligned} \varepsilon &= 1 - \alpha, \quad \tau = -\varepsilon \tan(\alpha\Phi_0), \quad \mathbf{u} = \tan(0.5\mathbf{V}), \quad \mathbf{b} = \tan(0.5\varepsilon\mathbf{V}), \quad \mathbf{B} = \mathbf{b} / (0.5\varepsilon\mathbf{V}), \\ \mathbf{z} &= \frac{\cos(\varepsilon\mathbf{V}) - \tan(\alpha\Phi_0) \sin(\varepsilon\mathbf{V})}{\mathbf{W} \cos(\mathbf{V})}, \quad \mathbf{d} = \frac{\mathbf{z}^{\varepsilon/\alpha} - 1}{\varepsilon} \end{aligned}$$

$S\alpha S$ 稳定分布序列 X 由下式给出：

$$X = \frac{2(\mathbf{u} - \mathbf{b})(1 + \mathbf{u}\mathbf{b}) - \Phi_0 \tau \mathbf{B}(\mathbf{b}(1 - \mathbf{u}^2) - 2\mathbf{u})}{(1 - \mathbf{u}^2)(1 + \mathbf{b}^2)} (1 + \varepsilon \mathbf{d}) + \tau \mathbf{d} \quad (3.10)$$

对上述过程进行循环即可得到已给满足给定条件的 $S\alpha S$ 分布序列。如此，则重构信号可用下式表示：

$$x^i(t) = x(t) + \varepsilon_k X \quad (3.11)$$

式中 $x(t)$ 为原始信号， $x^i(t)$ 为重构信号。

3.3.2 改进 EEMD 的枪声特征描述建模

本文提出了一种改进 EEMD 的枪声特征描述建模方法。该方法的特点是由于 $S\alpha S$ 分布符合广义中心极限定理，所有阶 $S\alpha S$ 分布序列的本征模态函数的总体均值为零。每次将随机的 $S\alpha S$ 分布序列人为添加到被分析的信号中以重构信号，并对重构信号进行若干次 EMD 分解，将得到的所有 IMF 分量求平均，得到最终的 IMF 作为被分析信号的基函数。

改进 EEMD 的枪声特征描述建模的流程图如图 3.3 所示，基本步骤如下：

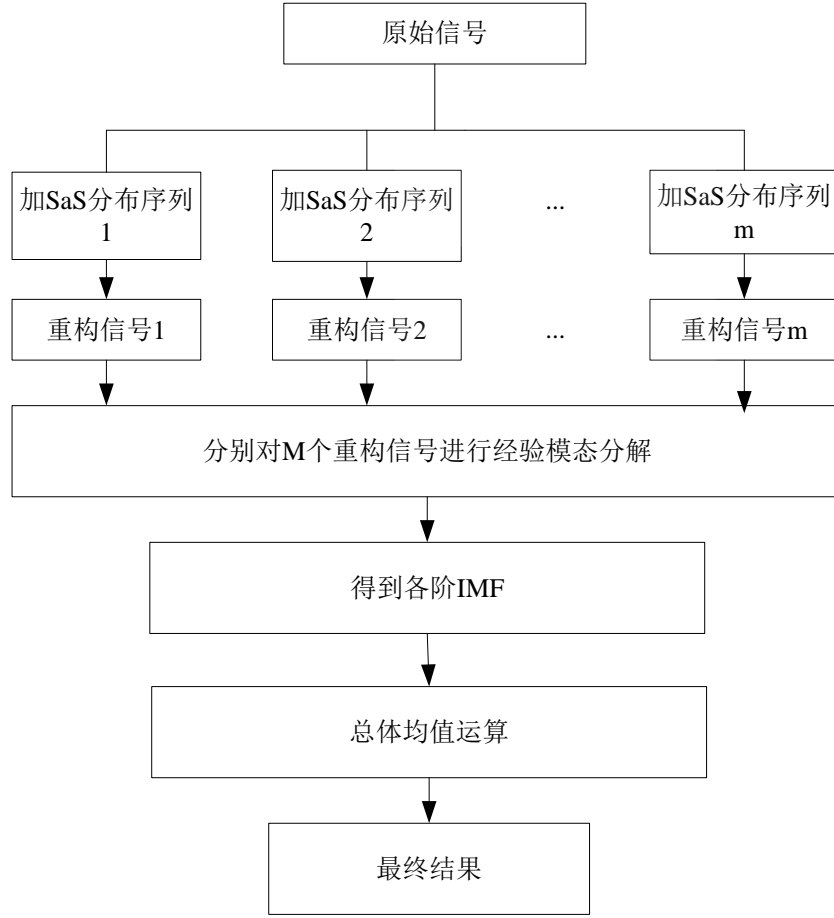


图 3.3 改进 EEMD 的枪声特征描述建模流程图

Fig.3.3 Flow chart of advanced EEMD gunshot feature description model

1 阶 IMF 与 EEMD 方法相似，可用下式表示：

$$\overline{\overline{IMF}}_1(t) = \frac{1}{I} \sum_{i=1}^I E_1[x^i(t)] \quad (3.12)$$

2 阶 IMF 对 $S\alpha S$ 分布序列进行独立的 EMD 分解，并用下式表示：

$$\overline{\overline{IMF}}_2(t) = \frac{1}{I} \sum_{i=1}^I E_1\{r_1(t) + \epsilon_1 E_1[X]\} \quad (3.13)$$

$k = 2, \dots, K$ 时， k 阶余项的表达式为 $k-1$ 阶余项与 k 阶 IMF 之差：

$$r_k(t) = r_{k-1}(t) - \overline{\overline{IMF}}_k(t) \quad (3.14)$$

k 阶 IMF 的表达式：

$$\overline{\overline{IMF}}_k(t) = \frac{1}{I} \sum_{i=1}^I E_1\{r_{k-1}(t) + \epsilon_{k-1} E_{k-1}[X]\} \quad (3.15)$$

K 为最终分解的阶数，如此则重构信号可由各阶 IMF 与余项的和来表达：

$$x^i(t) = \sum_{k=1}^K \overline{\overline{IMF}}_k + r_k^i(t) \quad (3.16)$$

图 3.4-3.6 为经过本文的方法分解不同场景下枪声信号的结果。

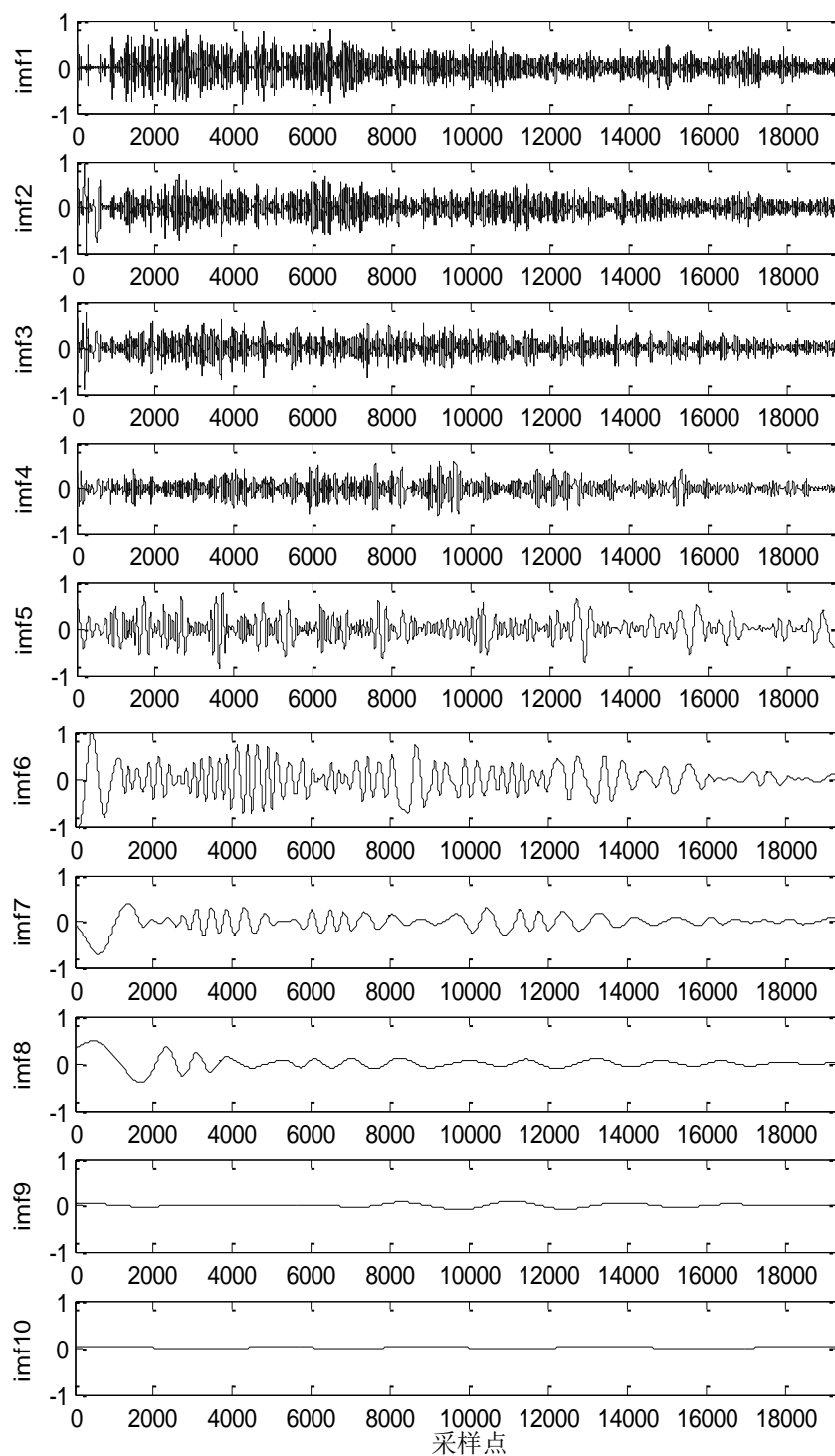


图 3.4 本文的方法分解三峡广场环境下手枪信号得到各阶 IMF

Fig. 3.4 Our method decomposed pistol signal under three gorges square get each order IMF

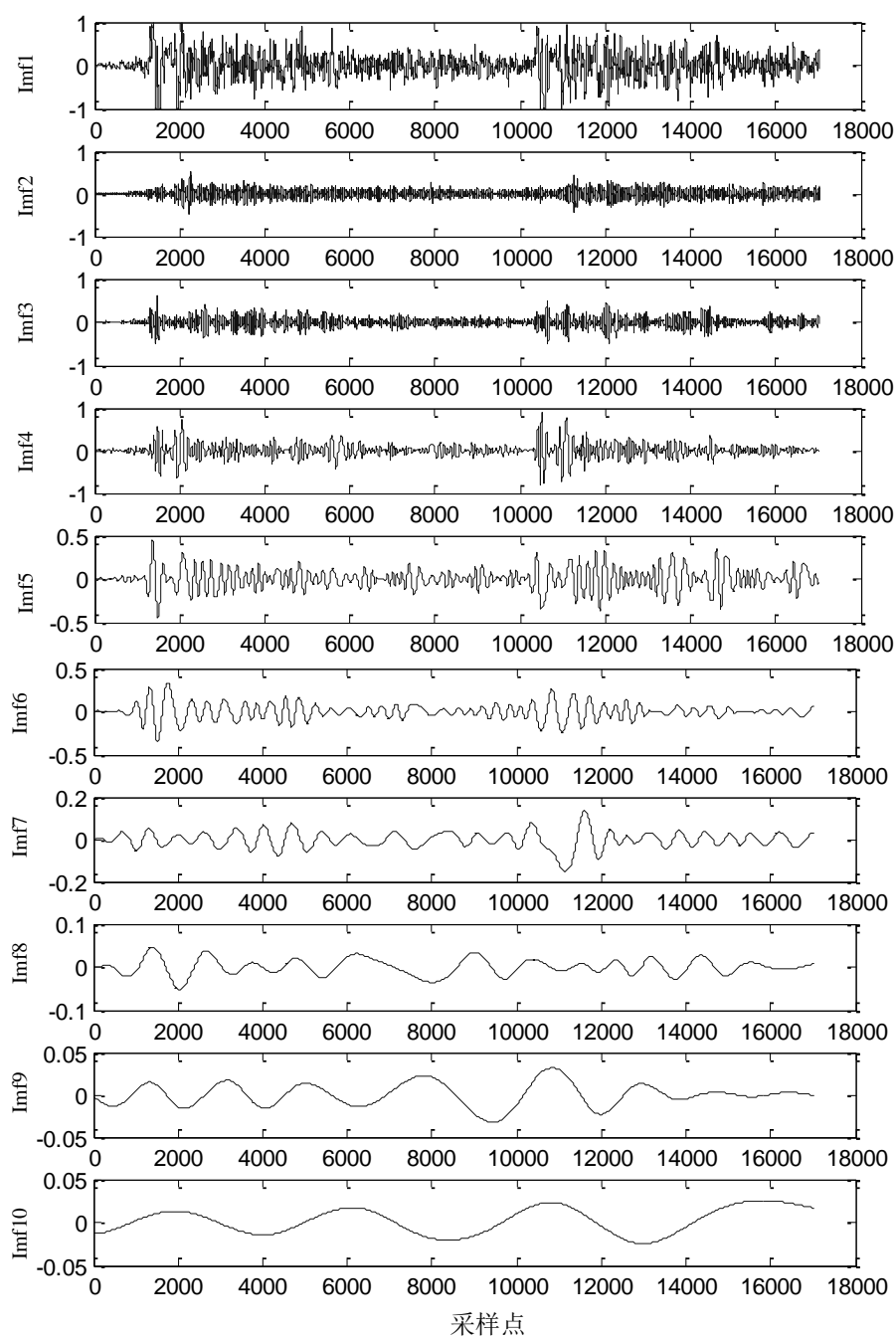


图 3.5 本文的方法分解城市道路环境下步枪信号得到各阶 IMF

Fig. 3.5 Our method decomposed noise rifle signal under city road get each order IMF

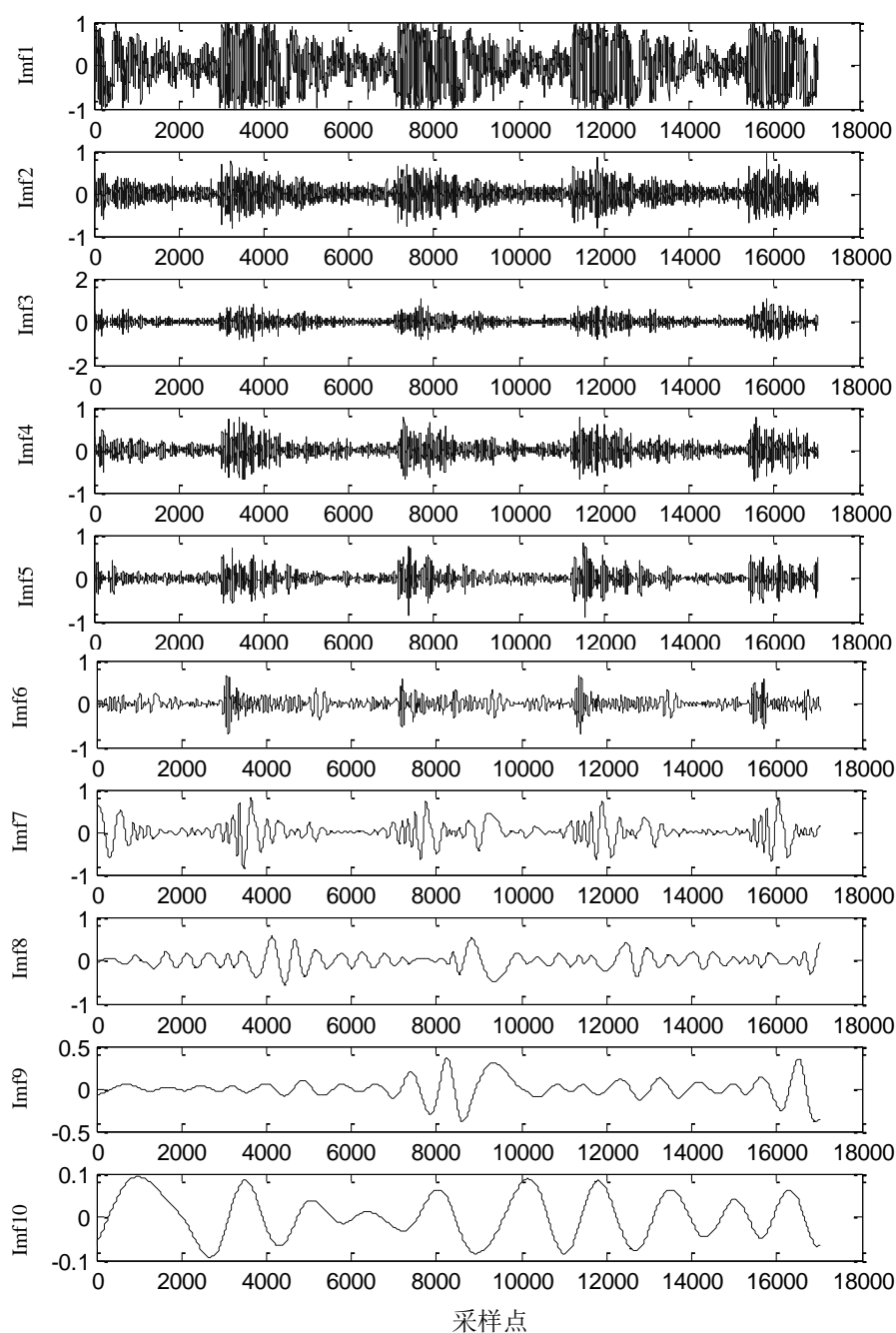


图 3.6 本文的方法分解火车站环境下机枪信号得到各阶 IMF

Fig. 3.6 Our method decomposed machine guns signal under railway station get each order IMF

3.3.3 公共场所枪声特征提取方法

特征提取的目的就是从许多特征中找出一组对识别最有效的特征以构成维数最优的特征向量。如果选择的是不同类目标差别很大的特征，就比较容易设计出性能较好的分类器。

枪声信号经过 EMD 分解后，会生成有限个 IMF 分量，每个 IMF 分量均代表着枪声某一过程，其中包括火药爆炸、子弹突破音障等。扣动扳机之后，火药爆

炸是个不稳定过程，其频率并不固定。当子弹出膛之后，子弹突破音障的频率并不是恒定不变的，而是有一定变化的。另外，由于撞针敲击子弹底部后不规则振动、内部其他振动产生的噪声、多普勒效应等种种原因，采集到的枪声信号中所包含的频率也有一定的变化，而且各个频率之间很难建立一种直接的联系。

因此，求取信号的瞬时频率意义不是很大。本文假设 IMF 分量的能量与重构信号的能量之间关系密切。先对 IMF 做 FFT 得到信号的幅值谱，然后利用 Parseval 定理分别求出重构的能量 (E) 以及各个 IMF 分量的能量 (E_i)，故以各个 IMF 分量的能量相对于原信号能量的能量比作为特征向量。即：

$$E_i = \frac{1}{N} \sum_{i=0}^{N-1} A_i^2 \quad (3.17)$$

$$\mathbf{k}_i = \frac{E_i}{E} \quad (3.18)$$

式中 A_i 为幅值谱中的幅值， N 为信号的长度， i 为信号经过生成 IMF 分量的个数。这样重构信号经过本文的方法分解生成了 N 个 IMF 分量和一个余项。本文将归一化的 \mathbf{k}_i 作为特征向量，输入分类器进行目标的分类与识别。本文分别采用 BP 神经网络以及支持向量机作为分类器并进行对比实验。如此则特征提取过程可用以下 5 步表示

Step1: 结合公共场所噪声模型重构原始信号。

Step2: 建立改进 EEMD 的枪声特征描述模型。

Step3: 采用本文的方法分解重构信号，得到所有 IMF 求整体平均，将其作为原枪声信号的最终 IMF。

Step4: 取各阶 IMF 与重构信号的能量之比 $\mathbf{k}_i = \frac{E_i}{E}$ 为特征向量。

Step5: 将特征向量 \mathbf{k}_i 输入分类器得到识别结果。

3.5 分类器

BP 神经网络是目前应用最广泛的神经网络模型之一^[39-42]。而支持向量机是目前性能最好的分类器之一。由于分类器的设计不是研究的重点，因此本节只简要阐述这两种本文用到的分类器。

3.4.1 BP 神经网络

BP 神经网络作为一种多层前向神经网络是单向传播的，除了输入输出节点网络一般具有一层以上的隐含层节点，而且这些节点同层不相耦合。输入数据从输入层节点进入，途经各个隐层节点，最后由输出层节点输出结果。

BP 网络的基本思想是：整个学习过程由信号的正向传播与误差的反向传播组成。输入层传入数据，隐层负责处理，最后由输出层输出结果，这是正向传播过

程。当输出层输出的结果与希望的不同，就会进入误差的反向传播阶段。即根据各单元分摊误差信号作为修正权值通过隐含层将误差返回到输入层。网络的训练过程会一直进行，直到进行到预先设定的学习次数或网络输出的误差达到预设范围之内为止。加权系数计算步骤如下：

① 加权系数初始化：用较小的随机数为 BP 神经网络所有加权系数赋初值准备训练样本，给出 N 组训练信号矢量组 $\mathbf{X}=(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m)^T$ 和 $\mathbf{D}=(\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_l)^T$ 取 $n=1$ ；

② 取 \mathbf{X}_n 和 \mathbf{D}_n ，计算隐含层和输出层各个神经元的输出；

③ 计算网络输出和期望输出的误差函数；

④ 计算输出层网络加权系数的调整量，并修正加权系数；

⑤ 取 $n=n+1$ 返回到步骤②，直到 E_p 进入设定的范围结束。

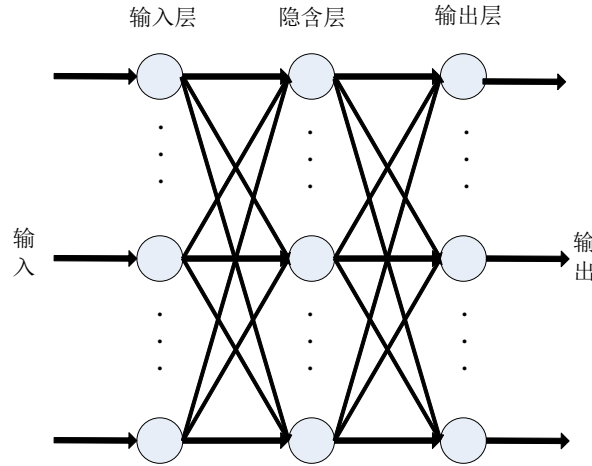


图 3.7 BP 神经网络结构图

Fig.3.7 BP neural network structure

3.4.2 支持向量机

支持向量机 SVM 是依据统计学习理论中的结构风险最小化原则提出的。本质上来看，支持向量机避开了从归纳到演绎的传统过程，大大简化了分类和回归问题。

SVM 的基本思想：首先将原始输入空间通过核函数变换到 Hilbert 空间。然后寻找最优化线性分类面。最优线性分类面在线性可分情况下会获得最大 Margin 以成功分类。也就是说 SVM 在保证经验风险最小的同时使界中置信范围最小。枪声识别是一个多分类问题，需要构造一个多分类器。

① 构造 n 个两分类规则，式中规则 $f_k(\mathbf{x})$ ， $k=1, \cdots, n$ 将 k 类的训练样本与其他样本分开（若向量 x_i 属于第 k 类，则 $\text{sgn}[f_k(\mathbf{x})]=1$ ，否则 $\text{sgn}[f_k(\mathbf{x})]=-1$ ）；

② 通过选取函数 $f_k(\mathbf{x})$, $k=1, \dots, n$ 中最大值的类别 $m = \arg \max \{f_1(\mathbf{x}), \dots, f_n(\mathbf{x})\}$ 这样就构造出一个 n 分类器。

设已知训练数据 $\mathbf{x}_1^1, \dots, \mathbf{x}_{l_1}^1, \dots, \mathbf{x}_1^n, \dots, \mathbf{x}_{l_n}^n$ 式中, \mathbf{x}_i^k 中的上标 k 表示向量属于 k 类。考虑线性函数集 $f_k(\mathbf{x}) = (\mathbf{x} * \mathbf{w}^k) + \mathbf{b}_k$, $k=1, \dots, n$ 本文的目标是构造 n 个函数(n 对 $(\mathbf{w}^k, \mathbf{b}_k)$), 使得规则 $m = \arg \max \{[(\mathbf{x} * \mathbf{w}^1 + \mathbf{b}_1), \dots, (\mathbf{x} * \mathbf{w}^n + \mathbf{b}_n)]\}$ 将训练数据无错误的分开, 则不等式 $(\mathbf{x}_i^k * \mathbf{w}^k) + \mathbf{b}_k - (\mathbf{x}_i^k * \mathbf{w}^m) - \mathbf{b}_m \geq 1$ 对所有 $k=1, \dots, n$, $m \neq k$ 和 $i=1, \dots, l_k$ 成立。

如果有这样一种解法, 本文要选择一系列点对 $(\mathbf{w}^k, \mathbf{b}_k)$, $k=1, \dots, n$ 使得泛函数 $\sum_{k=1}^n (\mathbf{w}^k * \mathbf{w}^k)$ 取最小值。

总而言之, 支持向量机是目前性能最好的分类器之一。

3.5 本章小结

本章是本文的核心, 首先介绍作为理论基础的 EMD 和 EEMD 方法。借鉴了 EEMD 方法假设环境噪声为高斯分布重构原始信号从而改进 EMD 的思路, 本文在此基础上做了如下改进: 本文假设 $S\alpha S$ 分布更符合公共场所背景噪声并替代高斯分布重构原始信号。其次, 重新推导公式, 提出改进 EEMD 的枪声特征描述模型, 对多次 EMD 分解得到的 IMF 求整体平均, 将其作为原枪声信号的最终 IMF, 将各阶 IMF 分量进行 FFT, 再根据 Parseval 定理求出原信号的能量以及各个 IMF 分量的能量, 最后以各个 IMF 分量的能量相对于原信号能量的能量比作为特征向量。

4 实验结果与分析

4.1 引言

本章根据第 3 章提出的改进 EEMD 的枪声特征提取及识别方法，设计并实施了一系列验证性实验。目前，业界并没有标准的公共场所枪声数据库，需要自己整理建库。而受环境所限无法在公共场所直接录制枪声，因此本文从网上收集整理了 3 大类（手枪、步枪、机枪）共 82 种枪声的模拟数据建立了原始枪声库。公共场所噪声库是通过在三峡广场、重庆市内道路两旁以及火车站录制而建立的。通过将原始枪声库与公共场所噪声库进行混合得到实验用的公共场所枪声信号。实验证明，本文提出的改进 EEMD 相比传统的语音处理方法中 MFCC、LPCC 等特征描述参数能够更准确的描述枪声特性。同时，相比原始的 EMD 与 EEMD 方法亦有一定程度的提高，而且鲁棒性更好。

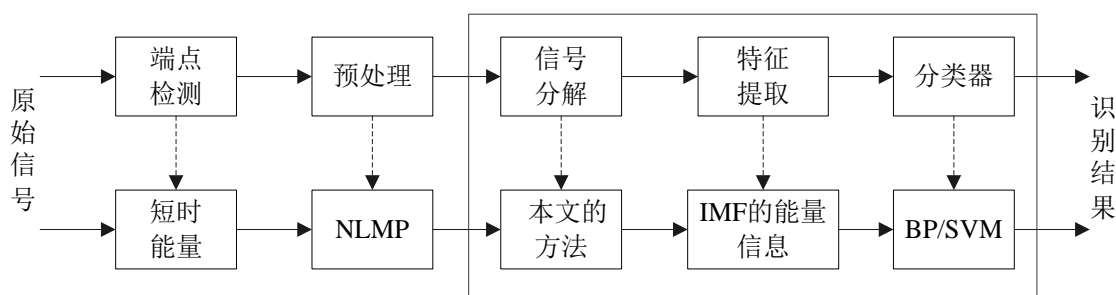


图 4.1 公共场所枪声的特征提取及识别流程框图

Fig.4.1 Flow chart of extraction and classification of gunshot in public place

4.2 预处理

对于音频信号处理，信号的特性分析和预处理都是十分重要的环节。信号的特性分析是信号特征提取的前提，只有对信号特性进行深入分析，才有可能得到既可以表征信号本质又具有良好鉴别性的特征。预处理的目的是减小噪声干扰，消除无用信息，增强信号的有用信息，从而提高特征提取、模式匹配及识别的可靠性。因此，特征提取的优良，识别率的高低与预处理的方法有直接的关系。

4.2.1 端点检测

要在公共场所这样的复杂环境下识别出来是何种枪声，首先需要将枪声信号的有效片段从环境噪声中检测出来，即确定枪声信号发生的起止点和终止点。这是枪声信号特征提取的基础，枪击事件识别的前提。本节针对枪声信号在公共场所噪声背景下的端点检测做了研究。由第 2 章对典型枪声的时域分析本文知道通

过短时能量双阈值算法可以实现公共场所背景噪声下的枪声信号端点检测。

端点检测算法的实现步骤如下：首先通过实验观察及统计，根据背景噪声和枪声信号的短时能量设置高低两个门限，门限值分别为 $0.8\overline{E_{\max}}$ 和 $0.4\overline{E_{\max}}$ ， $\overline{E_{\max}}$ 表示各类枪声信号短时能量最大值再取均值。低门限值数值较小，易被超过，而高门限值数值较大，需要较大的短时能量值才能通过。整个检测过程分为四个状态：背景噪声、过渡段、目标信号、终止点。在背景噪声状态阶段，若信号短时能量高于低门限则标记该点为起始点，同时进入过渡段。过渡段的设定是因为在实际的公共场所中并非突兀的信号都是本文“感兴趣”的信号。比如步行街广场、小孩子的欢笑声、口哨声、汽车的鸣笛声等通常在时域波形表现为瞬时尖峰脉冲，因此会干扰到检测。于是在过渡段中，当高门限阈值被超越，被标记为目标信号，否则回到背景噪声状态。对于一些瞬时的尖峰脉冲，通过设置持续时间门限来判别，设置 4800 个采样点，约为 0.3s。当检测状态处于目标信号时，信号短时能量强度又回落到低门限以下，并且没有越过持续时间门限值，则假设是干扰尖峰脉冲。否则就标记为目标信号的终止点。

图 4.2-4.4 是端点检测的结果，对于检测到的目标信号采用两条线段将端点位置标注了出来。

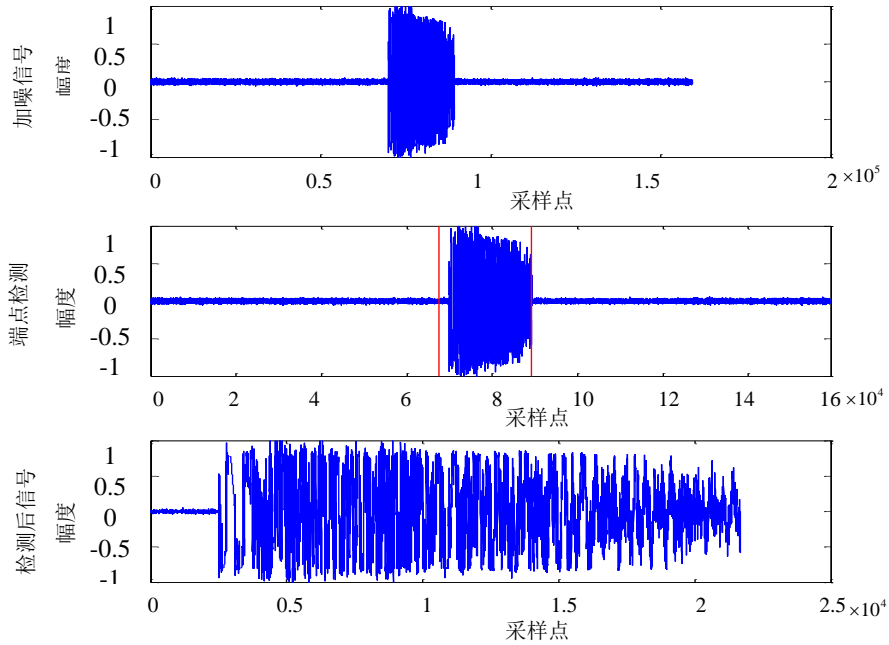


图 4.2 三峡广场环境下手枪信号的端点检测

Fig. 4.2 Endpoint detection of pistol signal under three gorges square environment

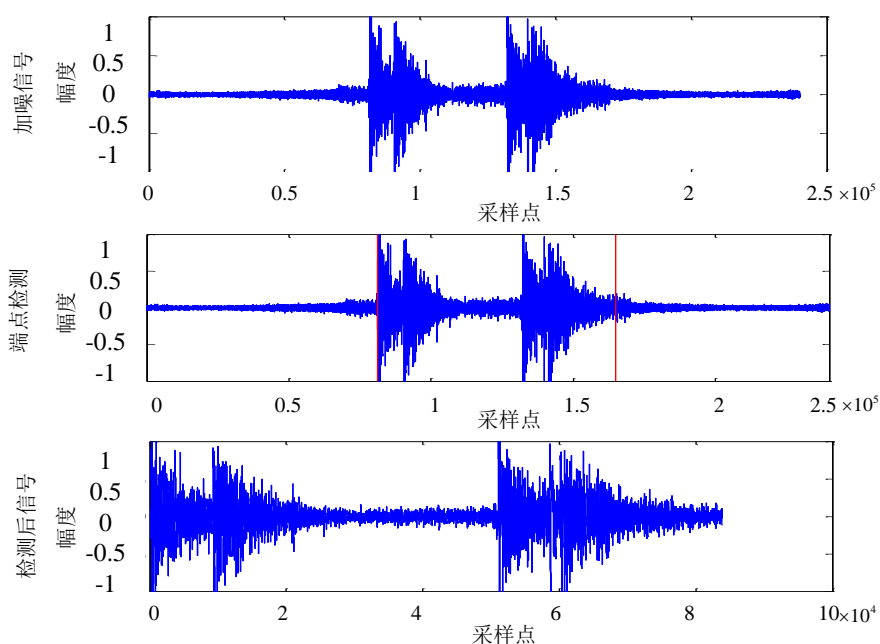


图 4.3 城市道路环境下步枪信号的端点检测

Fig. 4.3 Endpoint detection of rifle signal under city road environment

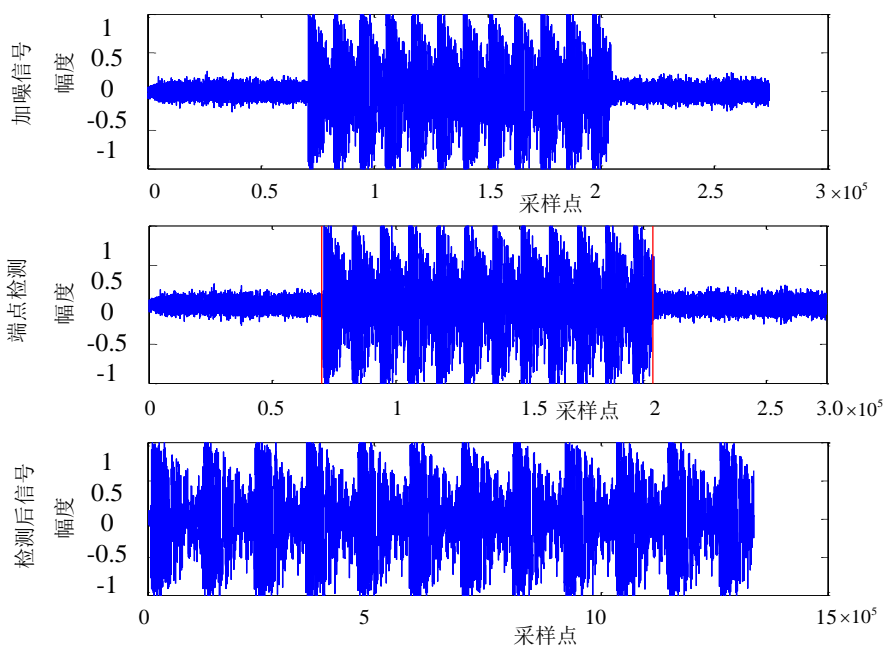


图 4.4 火车站环境下机枪信号的端点检测

Fig. 4.4 Endpoint detection of machine guns signal under railway station environment

接下来结合实验结果对算法的有效性以及存在问题进行讨论。图 4.3 和 4.4 分别显示了步枪信号和机枪信号的端点检测结果。由图 4.3 可以看出，步枪声发生之前，虽然受到尖峰脉冲噪声的干扰，不过起始点仍然被正确地检测了出来。表

明算法通过设置双能量阈值门限可以很好地避开噪声中瞬时尖峰脉冲的影响。由于枪声是较宽的脉冲信号，有一定的时间持续性，而当噪声脉冲的能量超过低门限阈值，在一定的时间内没能再超过高能量阈值，那么就假设是噪声干扰。从图 4.4 中可以看出，噪声中的尖峰脉冲短时能量很大，但没有引起误检，这是由于在过渡段设置的时间门限所起的作用。算法在检测到有疑似目标信号的起点时，开始计时，若被检测到的信号短时能量值没能在设定的时间门限范围内持续，就假设是干扰的尖峰脉冲噪声。

前面讨论了双阈值能量门限及过渡段的时间门限对抗尖峰脉冲干扰的效果，不过算法也存在一些问题：① 端点检测不精确：对于各类长度不一的枪声，会出现端点检测的不精确问题。比如机枪声明显比步枪声持续时间久，而步枪会比左轮手枪尾声拖得长。这就使得较长的声音可能还没有结束，就判断为终止；而较短的声音已经进入噪声，背景噪声却归入了枪声片段。② 信号淹没：对于低信噪比情况，各类枪声尾声的短时能量通常不会超过背景噪声，就可能会淹没在噪声中，那就意味着终止点不能够准确的被检测到。由图 4.3 步枪声的检测结果可以看出，在步枪声没有结束的时刻已经被判为终止。③ 误检测：误检测也是端点检测尤为突出的一个问题。公共场所背景噪声不能保证任何时刻背景噪声都是低频段、小能量。这样就会导致算法误检测，如图 4.2 的起始点并未精确地检测出。能量阈值设置过高会出现漏检，过低会将尖峰脉冲噪声错检测为目标信号。经多次实验确定在 0dB 信噪比的情况下，该算法的误检率为 6.1%，也就意味着在喧闹的环境下该算法也能够比较有效的检测出公共场所的枪声。

综上所述，短时能量的双阈值门限算法对于公共场所典型枪声的端点检测行之有效，但也存在不足。误检总不可避免，但对于端点检测不精确和信号淹没的问题仍可以补救，因此下一节将对淹没在背景噪声中的信号进行自适应滤波处理。

4.2.2 自适应滤波

传统自适应滤波器是基于最小均方误差准则进行滤波的，其性能在公共场所噪声背景下会严重退化，导致结果不可靠。为了解决这一问题，通常用最小分散系数准则代替基于最小均方误差准则。用误差函数的 p 范数取代平方误差函数可以很好的测量估计误差在零附近的波动范围。

基于这些思想，本文采用最小平均 p 范数(Least mean p -norm, LMP)。为了提高稳定性和算法的收敛速度，改进原算法，采用归一化的 LMP(normalized LMP, NLMP)算法，该算法可以更准确地估计梯度，改变自适应滤波器权向量增量的方向，从而获得了更快的收敛速度。

设 c 为 M 阶 FIR 自适应滤波器在 t 时刻的权系数矢量，表示为：

$$\mathbf{W}(t) = [\mathbf{w}_0(t), \mathbf{w}_1(t), \dots, \mathbf{w}_{M-1}(t)]^T \quad (4.1)$$

在 t 时刻的输入信号矢量 $\mathbf{x}_M(t)$ 可表示为:

$$\mathbf{x}_M(t) = [\mathbf{x}(t), \mathbf{x}(t-1), \dots, \mathbf{x}(t-M+1)]^T \quad (4.2)$$

在 t 时刻的输出信号 $y(t)$ 可表示为:

$$y(t) = \mathbf{x}_M(t)^T \mathbf{W}(t) \quad (4.3)$$

将 t 时刻的输入信号矢量及其前 $m-1$ 个时刻的输入信号矢量构成的矩阵表示为:

$$\mathbf{X}_{M,m}(t) = [\mathbf{x}_M(t), \mathbf{x}_M(t-1), \dots, \mathbf{x}_M(t-M+1)] \quad (4.4)$$

相应的输出信号矢量 $\mathbf{Y}(t)$ 可表示为:

$$\mathbf{Y}(t) = \mathbf{X}_{M,m}(t)^T \mathbf{W}(t) \quad (4.5)$$

式中, $\mathbf{Y}(t) = [\mathbf{Y}(t), \mathbf{Y}(t-1), \dots, \mathbf{Y}(t-M+1)]^T$ 。 $\mathbf{D}(t)$ 表示由 n 时刻及其前 $m-1$ 时刻构成的输出矢量而 $\mathbf{E}(t)$ 表示相应的误差矢量:

$$\mathbf{D}(t) = [\mathbf{D}(t), \mathbf{D}(t-1), \dots, \mathbf{D}(t-M+1)]^T \quad (4.6)$$

$$\mathbf{E}(t) = [\mathbf{E}(t), \mathbf{E}(t-1), \dots, \mathbf{E}(t-M+1)]^T \quad (4.7)$$

三者之间的关系为:

$$\mathbf{E}(t) = \mathbf{D}(t) - \mathbf{Y}(t) \quad (4.8)$$

代价函数可表示为:

$$\mathbf{J} = \mathbf{E} \left[|\mathbf{E}(t)|^p \right] \quad (4.9)$$

利用梯度下降法, 并以误差信号的瞬时值代替其统计平均, 得到梯度估计为:

$$\begin{aligned} \hat{\nabla}_{\mathbf{W}}(t) &= \frac{\partial \mathbf{J}(t)}{\partial \mathbf{W}(t)} = p \frac{\partial \mathbf{E}(t)}{\partial \mathbf{W}(t)} \left\{ |\mathbf{E}(t)|^{p-1} \Theta \text{sign}[\mathbf{E}(t)] \right\} \\ &= p[-\mathbf{X}_{M,m}(t)] \left\{ |\mathbf{E}(t)|^{p-1} \Theta \text{sign}[\mathbf{E}(t)] \right\} = -p \mathbf{X}_{M,m}(t) \mathbf{E}(t)^{<p-1>} \end{aligned} \quad (4.10)$$

式中: $|\mathbf{E}(t)|^{p-1} = [\mathbf{e}(t)^{p-1}, \mathbf{e}(t-1)^{p-1}, \dots, \mathbf{e}(t-m+1)^{p-1}]^T$,

$\text{sign}[\mathbf{E}(t)] = [\text{sign}(\mathbf{e}(t)), \text{sign}(\mathbf{e}(t-1)), \dots, \text{sign}(\mathbf{e}(t-m+1))]^T$,

$\mathbf{E}^{<p-1>}(t) = |\mathbf{E}(t)|^{p-1} \Theta \text{sign}(\mathbf{E}(t)) = [\mathbf{e}(t)^{<p-1>}, \mathbf{e}(t-1)^{<p-1>}, \dots, \mathbf{e}(t-m+1)^{<p-1>}]^T$,

$(\bullet)^{<p-1>} = |(\bullet)|^{p-1} \text{sign}(\bullet)$, $\text{sign}(\bullet)$ 表示符号函数。

当 $m=1$ 时, 得到归一化最小平均 p 范数(NLMP) 算法的权系数调整公式为:

$$\mathbf{W}(t+1) = \mathbf{W}(t) + p \frac{|\mathbf{e}(t)|^{p-1} \text{sign}(\mathbf{e}(t))}{\lambda + \|\mathbf{x}_M(t)\|_p^p} \mathbf{x}_M(t) \quad (4.11)$$

NLMP 算法中步长因子 μ 取 0.05, 阶数 1.05。图 4.5-4.7 是自适应滤波的结果。

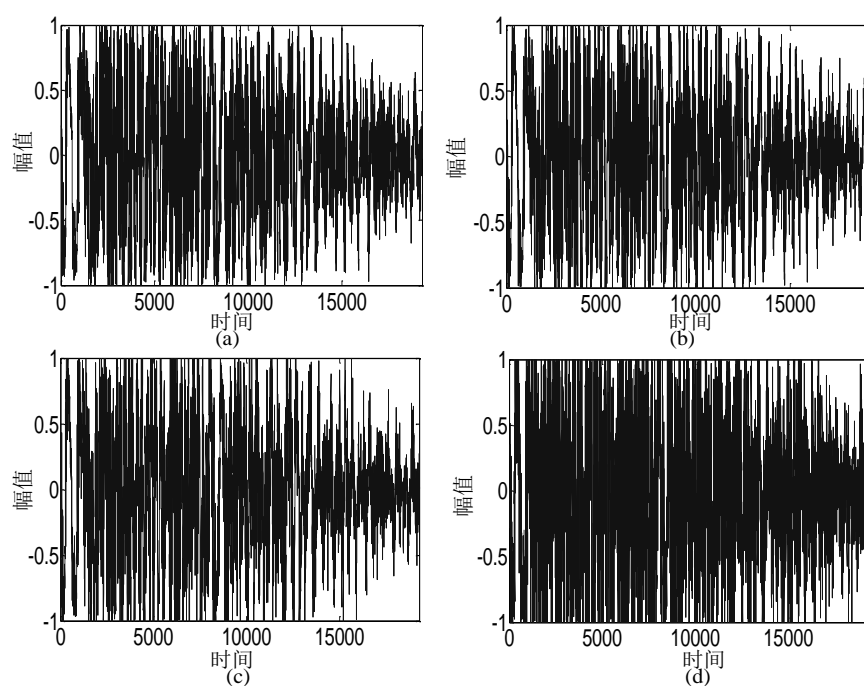


图 4.5 不同信噪比三峡广场环境下手枪信号自适应滤波：(a)原始信号自适应滤波；(b) SNR=30db 自适应滤波；(c) SNR=25db 自适应滤波；(d) SNR=20db 自适应滤波

Fig. 4.5 Different signal to noise ratio of pistol signal under three gorges square environment after adaptive filter(a) Adaptive filtering of original signal;(b) Adaptive filtering of SNR=30db;(c) Adaptive filtering of SNR=25db;(d) Adaptive filtering of SNR=20db

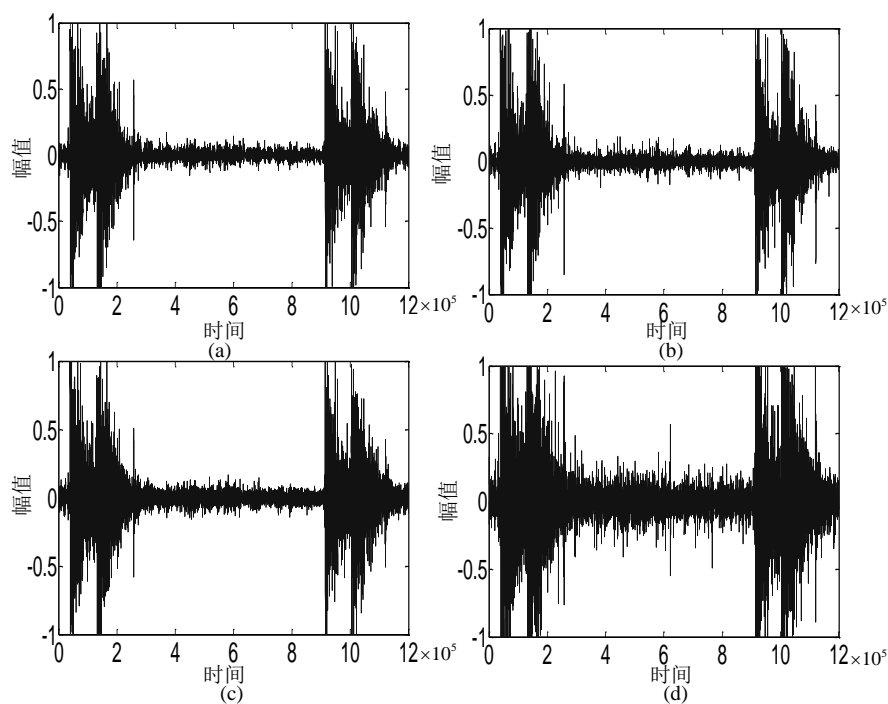


图 4.6 不同信噪比城市道路环境下步枪信号自适应滤波：(a)原始信号自适应滤波；(b) SNR=30db 自适应滤波；(c) SNR=25db 自适应滤波；(d) SNR=20db 自适应滤波

Fig. 4.6 Different signal to noise ratio of rifle signal under city road environment after adaptive filter: (a) Adaptive filtering of original signal;(b) Adaptive filtering of SNR=30db;(c) Adaptive filtering of SNR=25db;(d) Adaptive filtering of SNR=20db

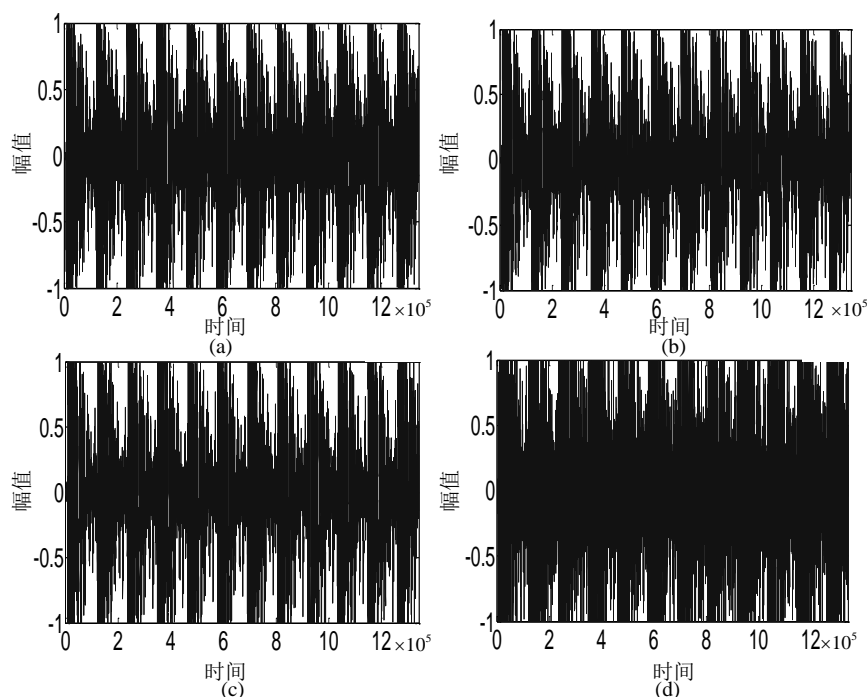


图 4.7 不同信噪比火车站环境下机枪信号自适应滤波：(a)原始信号自适应滤波；(b) SNR=30db 自适应滤波；(c) SNR=25db 自适应滤波；(d) SNR=20db 自适应滤波

Fig. 4.7 Different signal to noise ratio of machine guns signal under railway station environment after adaptive filter: (a) Adaptive filtering of original signal;(b) Adaptive filtering of SNR=30db;(c) Adaptive filtering of SNR=25db;(d) Adaptive filtering of SNR=20db

通过 NLMP 算法自适应滤波对比发现：工频噪声所引起的干扰被明显去除，枪声信号的波形特征基本上得到还原，使得信号细节得以重现。NLMP 算法在自适应滤波过程中，有效地去除了工频噪声干扰的同时，对于原始信号的整体和细节信息都得到了很好的保留。

综上所述，最小平均 p 范数算法针对工频噪声去除行之有效。虽然通过去噪，目标信号有耗损，但作为研究目标的枪声信号还是较好保留了整体波形和信号特征。

4.3 实验结果与分析

目前，业界并没有标准的公共场所枪声数据库，需要自己整理建库。而受环境所限无法在公共场所直接录制枪声，因此本文从网上收集整理了3大类（手枪、步枪、机枪）共82种枪声的模拟数据建立了原始枪声库。该库中共有3类枪声，每类有80种样本，单个样本的采样率为44.1KHz，位速为705Kbps，位数为16位，样本长度为60-120s。公共场所噪声库是通过在三峡广场、重庆市内道路两旁以及火车站录制而建立的。通过将原始枪声库与公共场所噪声库进行混合得到实验用的公共场所枪声信号。其流程框图如图4.8所示。实验中，3类典型的枪声，每类80个样本。从每类枪声库中随机抽取40个样本训练，剩余的40个为待测样本。实验采用相同的训练、测试样本集连续进行十次实验，识别率结果均为十次实验的均值±标准差。

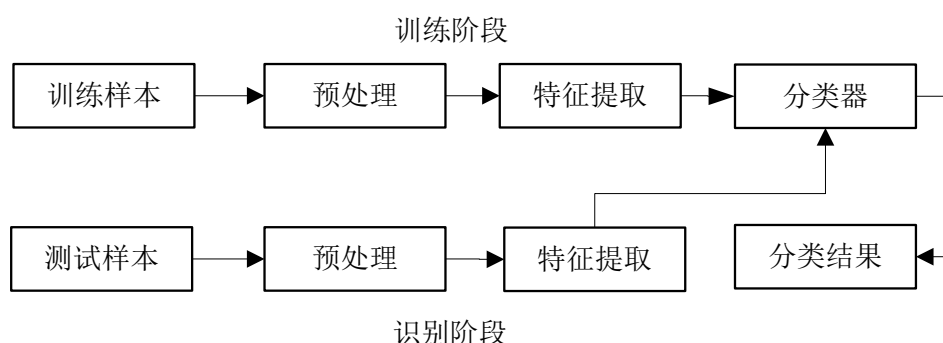


图 4.8 公共场所典型枪声特征提取及识别流程框图

Fig.4.8 Flow chart of typical gunshot feature extraction and classification in public places

4.3.1 不同特征描述方法对比实验

为了检验特征描述算法的性能，在训练样本 N 分别为 40, 30, 20 的条件下采用 MFCC、LPCC、EMD、EEMD 以及本文的方法 5 种方法，分别重复 10 次实验取平均值；实验中，3 类典型的枪声，每类 80 个样本，去除训练样本其余作为测试样本，分类器为 BP 神经网络分类器。结果如表 4.1 所示。特征描述参数设置：稳定分布特征系数 $\gamma=1.0, \alpha=1.8$ ；噪声幅值限定系数 0.2；最大迭代次数 100；IMF 最大阶数 10。BP 神经网络分类器参数设置：隐层节点数 10；学习步长 0.3；动向量系数 0.95；最小均方误差 $1e-8$ ；最小梯度 $1e-20$ 。

表 4.1 不同训练样本数无噪声情况下枪声识别的正确率(%)比较

Tab.4.1 Comparisons of gunshot recognition accuracy under different training samples without noise

特征描述方法	(%)			用时/s
	N=40	N=30	N=20	
MFCC	62.41±3.54	60.97±4.31	59.73±3.95	0.047
LPCC	72.33±4.24	71.22±5.10	70.57±4.65	0.088
EMD	85.93±2.77	84.39±3.01	83.25±3.17	0.982
EEMD	87.31±3.98	86.74±4.22	85.33±4.57	10.924
本文的方法	93.33±2.63	92.18±2.95	91.65±3.03	13.282

从表 4.1 中可以看出, 与两种传统的特征描述参数 MFCC 和 LPCC 相比, 本文的方法提高识别率高达 19.32-30.92%。显然对于枪声这种非语音信号, 原本用于语音信号识别的 MFCC 与 LPCC 等特征描述参数已不再适用, 本文的方法对于非语音信号能更为有效的提取其特征。而与原始的 EMD、EEMD 方法, 本文的方法提高识别率 6.02-10.08%。这是由于本文的方法经过特性分析重新建立了公共场所噪声模型, 用 $S\alpha S$ 分布代替高斯分布, 从原理上改进了原始的 EMD、EEMD 方法。

为了检验特征描述算法在真实噪声环境下的性能, 在信噪比(SNR)为 30db、25db、20db 的白噪声、三峡广场、城市道路以及火车站噪声环境下分别采用 MFCC、LPCC、EMD、EEMD 以及本文的方法 5 种方法重复 10 次实验取平均值; 实验中, 每类枪声 80 个样本, 分类器为 BP 神经网络分类器。结果如表 4.2-4.5 所示。

从表 4.2-4.5 中可以看出, 与两种传统的特征描述参数 MFCC 和 LPCC 相比, 本文的方法提高识别率高达 14.25-29.88%。这里可以看出相比于传统的 MFCC 与 LPCC 等特征描述参数相比本文的方法在模拟真实噪声环境下依然有较高的识别率, 鲁棒性更好。而与原始的 EMD、EEMD 方法, 本文的方法提高识别率 2.96-7.56%。

表 4.2 不同信噪比白噪声环境下枪声识别的正确率(%)比较

Tab.4.2 Comparisons of gunshot recognition accuracy under different SNR conditions with white

特征描述方法	noise (%)			用时/s
	SNR=30db	SNR=25db	SNR=20db	
MFCC	61.67±3.11	60.52±4.01	58.48±4.22	0.048
LPCC	71.67±6.01	70.56±5.72	69.81±5.03	0.087
EMD	85.26±4.24	82.96±2.97	80.53±4.30	0.993

特征描述方法	SNR=30db	SNR=25db	SNR=20db	用时/s
EEMD	85.96 ± 3.44	82.66 ± 3.05	81.69 ± 3.53	10.725
本文的方法	89.46 ± 3.44	88.89 ± 2.88	83.33 ± 4.57	13.109

表 4.3 不同信噪比三峡广场环境下枪声识别的正确率(%)比较

Tab.4.3 Comparisons of gunshot recognition accuracy under different SNR conditions under three gorges square environment (%)

特征描述方法	SNR=30db	SNR=25db	SNR=20db	用时/s
MFCC	60.27 ± 6.37	57.52 ± 5.70	54.48 ± 5.01	0.049
LPCC	70.34 ± 5.51	68.33 ± 3.53	65.29 ± 4.62	0.089
EMD	83.22 ± 3.09	81.48 ± 2.42	80.18 ± 4.27	0.998
EEMD	84.44 ± 2.32	81.77 ± 3.05	80.45 ± 4.79	10.476
本文的方法	88.77 ± 2.05	86.18 ± 3.65	81.85 ± 4.16	13.203

表 4.4 不同信噪比城市道路环境下枪声识别的正确率(%)比较

Tab.4.4 Comparisons of gunshot recognition accuracy under different SNR conditions under city road environment (%)

特征描述方法	SNR=30db	SNR=25db	SNR=20db	用时/s
MFCC	60.67 ± 5.62	58.35 ± 4.97	56.21 ± 5.35	0.048
LPCC	70.19 ± 5.31	69.34 ± 5.17	68.82 ± 4.29	0.091
EMD	85.14 ± 4.17	84.22 ± 3.87	81.57 ± 3.59	0.992
EEMD	86.92 ± 3.19	85.71 ± 4.19	82.62 ± 3.75	10.386
本文的方法	89.23 ± 2.53	88.57 ± 3.05	83.19 ± 3.13	13.413

表 4.5 不同信噪比火车站环境下枪声识别的正确率(%)比较

Tab.4.5 Comparisons of gunshot recognition accuracy under different SNR conditions under railway station environment (%)

特征描述方法	SNR=30db	SNR=25db	SNR=20db	用时/s
MFCC	59.35 ± 7.28	56.82 ± 6.39	53.34 ± 5.83	0.051
LPCC	69.73 ± 5.70	65.07 ± 6.27	64.21 ± 4.53	0.092
EMD	81.67 ± 5.02	80.63 ± 5.32	79.07 ± 4.88	0.986
EEMD	82.96 ± 4.97	81.22 ± 4.44	80.18 ± 4.11	10.582

特征描述方法	SNR=30db	SNR=25db	SNR=20db	用时/s
本文的方法	85.92±2.37	84.62±3.62	84.07±3.57	13.534

$$\text{识别率}(\%) = \text{识别正确的样本数} / \text{测试样本总数} \times 100\%$$

这里可以看出相比于原始的 EMD、EEMD 方法相比本文的方法在模拟真实噪声环境下依然有较高的识别率。图 4.9 为五种方法的受试者工作特征曲线(Receiver Operating Characteristic, ROC), 是一种描述统计特征的参数, 得此名的原因在于曲线上各点反映着相同的感受性, 它们都是对同一信号刺激的反应, 只不过是在几种不同的判定标准下所得的结果而已。接受者操作特性曲线就是以虚报概率为横轴, 击中概率为纵轴所组成的坐标图, 和被试在特定刺激条件下由于采用不同的判断标准得出的不同结果画出的曲线。ROC 曲线, 是反映敏感性和特异性连续变量的综合指标, 是用构图法揭示敏感性和特异性的相互关系, 它通过将连续变量设定出多个不同的临界值, 从而计算出一系列敏感性和特异性, 再以敏感性为纵坐标、特异性为横坐标绘制成曲线, 曲线下面积(Area Under Curve, AUC)越大, 识别准确性越高。本文的方法曲线下面积(AUC)高达 0.994 比其他方法在统计学意义上都更有优势。

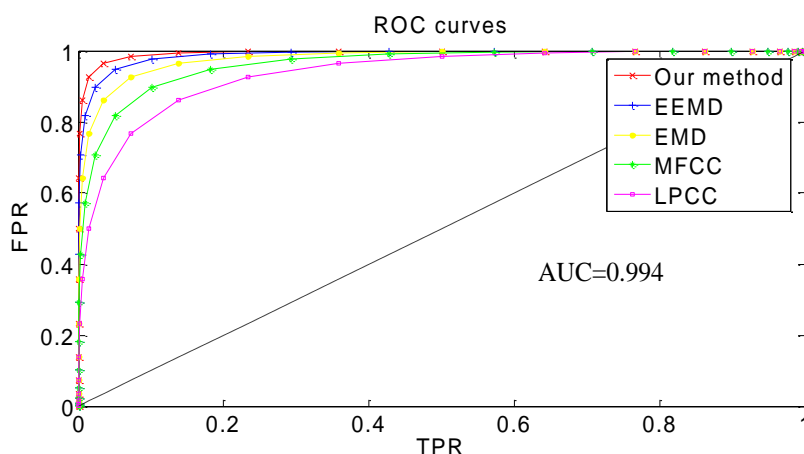


图 4.9 五种方法的受试者工作特征 (ROC) 曲线

Fig.4.9 Receiver operating characteristic curves results of five methods

4.3.2 不同分类器对比实验

为了检验特征分类器的性能, 在训练样本 N 为 40、信噪比 SNR 为 30db 的白噪声、三峡广场、城市道路以及火车站噪声环境下采用本文的方法, 分别重复 10 次实验取平均值; 实验中, 3 类典型的枪声, 每类 80 个样本, 去除训练样本其余作为测试样本, 分类器为 BP 神经网络分类器和 SVM 分类器。分类器参数设置:

稳定分布特征系数 $\gamma=1.0, \alpha=1.8$ ；噪声幅值限定系数 0.2；最大迭代次数 100；IMF 最大阶数 10。BP 神经网络分类器参数设置：隐层节点数 10；学习步长 0.3；动向量系数 0.95；最小均方误差 $1e-8$ ；最小梯度 $1e-20$ 。SVM 分类器采用高斯径向基核函数(RBF)， $K(x, x') = \exp(-\|x - x'\|^2 / \sigma^2)$ 。模型选择 k 遍交叉验证法，利用遗传算法找最优参数：惩罚参数 $C=100$ ；核参数 $\sigma = 0.235$ 。

表 4.6 本文的方法在不同分类器情况下分类识别率(%)的比较

Table 4.6 Comparisons of our method recognition accuracy under different classifications (%)

分类器	无噪声环境	白噪声环境	广场	道路	车站	用时/s
BP	93.33 ± 2.64	89.46 ± 3.44	88.77 ± 2.05	89.23 ± 2.53	85.92 ± 2.37	7.836
SVM	95.15 ± 2.37	93.72 ± 2.74	91.34 ± 1.98	92.15 ± 2.25	88.73 ± 2.30	3.144

由表 4.6 的实验结果可知，在本文的方法作为特征描述方法的情况下，SVM 作为分类器较 BP 识别效果更佳。BP 神经网络分类器在无噪声情况下，对公共场所典型枪声的识别率为 93.33%，较 SVM 低 1.8 个百分点。这是由神经网络分类器的分类原理所致，如果 BP 神经网络要提高识别率就需要更复杂的计算，更长的学习过程。

4.4 本章小结

本章对第 3 章提出的改进 EEMD 的枪声特征提取及识别的方法进行了实验验证。首先，对枪声数据进行预处理，由于无法直接在公共场所录制枪声，因此采用模拟枪声信号与实际公共场所环境噪声混合的方法得到原始信号，并进行端点检测和自适应滤波。其次，用第 3 章提出的改进 EEMD 的特征提取方法得到特征向量。最后，将特征向量输入分类器得到识别结果。实验证明，本文提出的改进 EEMD 相比传统的语音处理方法中 MFCC、LPCC 等特征描述参数能够更准确的描述枪声特性。同时，相比原始的 EMD 与 EEMD 方法亦有一定程度的提高，而且鲁棒性更好。

5 总结展望

5.1 工作总结

公共场所的安全防护对社会安定和谐, 人民生命财产安全意义重大。为解决这个日益突出的社会问题, 各地不同程度地建设了视频监控平台, 为人们打造平安和谐的生活环境发挥了积极的作用。基于音频监控的公共场所枪声识别可以有效揭示枪击事件的状况, 还可以补充视频监控侵犯隐私, 夜晚效果不佳等等不足。因此, 对于公共场所枪声信号的研究具有重大的实用价值和社会意义。本文调研了相关的国内外文献, 立足于公共场所环境, 针对其中频率多发的枪声信号, 对其特征提取及识别方法进行了积极有益的探索和研究, 其主要工作内容总结为以下 3 个方面:

① 公共场所枪声信号及环境背景噪声特性分析。对典型的几种枪声信号及公共环境下背景噪声进行了特性分析。并通过对时域、频域、时频域、倒谱域上枪声信号相关特性的统计分析, 得到一些枪声信号具有非线性、非平稳性的共性。公共场所环境如广场、车站、码头等背景噪声非常嘈杂, 常会淹没原始枪声信号的特征, 使得枪声信号变得很不规则, 甚至破坏原始枪声信号的时频及其他相关特征。经过对公共场所噪声的特性分析, 本文假设公共场所噪声更符合 $S\alpha S$ 分布, 并以此替代高斯分布。

② 公共场所枪声特征描述建模。EEMD 通过每次将随机的高斯分布序列人为添加到被分析的信号中, 用以解决 EMD 存在的模态混叠问题。受此启发, 本文充分考虑公共场所环境下背景噪声分布情况, 对 EEMD 进行改进, 提出在重构枪声信号时用 $S\alpha S$ 分布代替高斯分布, 将重构信号进行 EMD 分解, 对多次 EMD 分解得到的 IMF 求整体平均, 将其作为原枪声信号的最终 IMF。

③ 公共场所枪声特征提取方法研究。枪声信号经过 EMD 分解后, 会生成有限个 IMF 分量, 而每个 IMF 分量的频率并不固定。为此, 需要寻找新的特征。为此, 本文将经过 EMD 分解生成的各阶 IMF 分量进行 FFT, 再根据 Parseval 定理求出原信号的能量以及各个 IMF 分量的能量, 最终以各个 IMF 分量的能量相对于原信号能量的能量比作为特征向量。

5.2 工作展望

本文针对公共场所枪声信号的特征提取和识别两个方面进行了较为深入的理论分析, 并进行了相关的实验, 取得了些许成果。但由于对枪声识别的研究刚刚起步, 在追求低计算成本、高识别率兼顾对现实环境噪声的强鲁棒性等方面仍需

要探索，单本人的工作还远远不够，因此，在总结前期所做工作的同时，本文还考虑了课题下一步的具体研究工作，包括以下几个方面：

① 公共场所枪声特征提取及识别研究与公共场所背景噪声密不可分，因而背景噪声的模型还有待深入研究。包括在不同时间段（如早上、中午、晚上、深夜等）不同场景（如广场、车站、码头、校园等公共场所）全天候实时采集背景噪声数据，建立更为完善更符合真实噪声分布的模型。

② 在此基础上，利用建立的真实噪声分布模型优化 EMD 或新的非语音信号分析方法构建新的公共场所枪声特征描述模型。此外，能量比也并非唯一可以作为特征向量的参数，可以考虑引入其他系数以完善特征提取的全面性。

③ 本文采用了 BP 神经网络和支持向量机两种分类器，得到了比较好的识别精度。下一步还可以考虑采用特征融合的方法进一步提高识别精度。

致 谢

“天行健，君子以自强不息”，三年的研究生生活转瞬即逝，感伤之余，要庆幸的是自己所在的 IITLAB 是个快乐的大家庭，使得我在生活、学习和工作各个方面都有所提高。因此，在这个毕业季，论文完成之时，我要向曾经教导、关心、帮助和支持过我的 IITLAB 里所有可爱的人们献以我最诚挚的敬意！

首先我要对我的导师李伟红副教授表示最衷心的感谢。本文能够顺利完成，离不开李老师的悉心指导和严格要求，李老师渊博的专业知识，严谨的治学态度，精益求精的工作作风，诲人不倦的高尚师德，严以律己、宽以待人的崇高风范，朴实无华、平易近人的人格魅力对我影响深远，不仅使我树立了自己的学术目标、掌握了基本的研究方法，还使我明白了许多待人接物与为人处世的道理。其次感谢龚卫国教授在学习和生活上给予我的关怀和帮助，在科研工作上从方法论的高度指引我前进的方向。另外，龚卫国老师经常强调做事先做人的道理，未来社会团队合作的思想，我都铭记在心。还要感谢我的副导师蔡金老师，他广博的知识面与丰富的人生阅历都使我收益颇丰，在此对三位老师在我学习和生活中给予无微不至的关怀表示感谢。

感谢李正浩、杨利平老师对我的指导和帮助，在课题研究中，两位老师给了我很多指导，提出了许多宝贵的意见。感谢在研究生学习期间，研究室的各位兄弟姐妹对我的无私帮助，他们热情大方，踏实勤奋，使我的研究生生活充满色彩和快乐，与他们的相处，使我学会了很多，感受到了读研的乐趣。

衷心感谢我的家人，勤劳朴实的父母在我的成长中给予了无限的关怀和鼓励，你们是我不断前进的坚实后盾，你们的默默付出让我不怕困难，专心学习和工作。在此，祝你们身体健康、幸福快乐！

最后，衷心地感谢在百忙之中评阅论文、参加答辩并提出宝贵意见的各位专家、教授！

张智龙

二〇一三年四月 于重庆

参考文献

- [1] Izabela L. Freire, Jose A. Apolinario Jr. Gunshot detection in noisy environments[C]. The 7th International Telecommunications Symposium ITS 2010,13(5): 455 - 459.
- [2] R. O. Duda, P. E. Hart, D. Stork. Pattern Classification [M].Wiley Inter science, 2000, 8(4):239-244.
- [3] 陆燕芳, 何巧, 罗晓松. 火炮声探测技术研究报告 [J]. 电声技术, 1993, 93(3): 2-6.
- [4] K.H. Davis. Automatic Recognition of Spoken Digits [J]. Journal of the Acoustical Society of America, 1952, 3(24):637-642.
- [5] F. Itakura. Minimum prediction residual principle applied to speech recognition [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1975, 23(1): 67-72.
- [6] 郭伟. 基于听觉神经原理的语音信号处理 [D]. 上海, 上海交通大学硕士学位论文, 2009.
- [7] S. Davis, P. Mermelstein. 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.
- [8] 赵力. 语音信号处理 [M]. 北京, 机械工程出版社, 2003.
- [9] D. Hoiem, Y. Ke, R. Sukthankar. SOLAR: Sound Object Localization and Retrieval in Complex Audio Environments [C]. Acoustics, Speech, and Signal Processing. Proceedings. IEEE International Conference, 2005, 3(5): 18-23.
- [10] I. Akinori, A. Akihito, I. Masashi, S. Makino. Detection of Abnormal Sound Using Multi-stage GMM for Surveillance Microphone [C]. 2009 Fifth International Conference on Information Assurance and Security, 2009, 1: 733-736.
- [11] P. K. Atrey, N. C. Maddage. Audio Based Event Detection for Multimedia Surveillance [C]. Acoustics, Speech, and Signal Processing, IEEE International Conference, 2006, 5: 14-19.
- [12] Woohyun Choi, Jinsang Rho, Han D.K., Hanseok Ko. Selective Background Adaptation Based Abnormal Acoustic Event Recognition for Audio Surveillance [C]. IEEE Conference on Advanced Video and Signal Based Surveillance, 2012, 6(1): 118-123.
- [13] I. L. Freire, J. A. Apolinário. Gunshot Detection in Noisy Environments [C]. The 7th International Telecommunications Symposium. Manaus, Brazil, September 2010: 1-4.
- [14] L. Gerosa, G. Valenzise, M. Tagliasacchi. Scream and Gunshot Detection and Localization for Audio-Surveillance Systems [C]. IEEE Conference on Advanced Video and Signal Based Surveillance, 2007, 9(5): 21-26.
- [15] 栾少文. 公共场所典型枪声的特征提取 [J]. 计算机工程, 2010, 36(7): 208-210.

- [16] 吕霄云. 基于 MFCC 和 GMM 的枪声识别算法 [J]. 计算机应用, 2010, 卷(3): 796-798.
- [17] Stg'phane Mallat, Multifrequency Channel. Decompositions of Images and Wavelet Models. IEEE Transaction in Acoustic Speech and Signal Processing, 1989, 37(7): 2091-2098.
- [18] Semion Kizhner, Thomas P. Flatley, Norden E. Huang. On the hilbert-huang transform data processing system development[C]. IEEE Aerospace Conference Proceedings, 2004, (4): 263 – 266.
- [19] Lei Hong. Decomposition and forecast for financial time series with high-frequency based on empirical mode decomposition [J]. Control Systems Technology, IEEE 2011, 19(3): 507– 512.
- [20] Nikolaos Tsakalozos, Konstantinos Drakakis, Scott Rickard. A formal study of the nonlinearity and consistency of the Empirical Mode Decomposition [J]. Signal Processing, IEEE 2011, 58(6): 2473– 2481.
- [21] Zhan Renjun, Wu Husheng. Feature extraction and application of engineering bon-stationary signals based on EMD-AR model and SVD[C]. IEEE International Conference on Information Technology, 2010, 10(1): 825– 827.
- [22] J. R. Jimenez, H. R. Wu. Empirical mode decomposition as a tool for data[C]. The 6th IEEE Conference on Industrial Electronics and Applications, 2011, 8(5): 246– 249.
- [23] Yannis Kopsinis, Stephen McLaughlin. Development of EMD-based denoising methods inspired by wavelet thresholding[C]. IEEE Conference on Advanced Video and Signal Based Surveillance, 2009, 8(1): 21– 26.
- [24] A. Pikrakis, T. Giannakopoulos, and S. Theodoridis. Gunshot detection in audio streams from movies by means of dynamic programming and Baysean networks[J]. IEEE 2008, 20(13): 2201– 2206.
- [25] Dexiang Zhang, Xiaopei Wu, Yixiang Lu. Application of EMD in fault analysis of automobile gearbox vibration [J]. Journal of Anhui University: Natural Science Edition, 2009, 3(2): 35– 38.
- [26] Raymond C. Wong, Arvind Kumar Sinha, Malcolm Mahadevan, Tiong Cheng Yeo. Diagnostic utility, safety, and cost-effectiveness of emergency department-initiated early scheduled technetium-99m single photon emission computed tomography imaging followed by expedited outpatient cardiac clinic visits in acute chest pain syndromes[J]. Emerg Radiol , 2011, 17(15): 375– 380.
- [27] Yunchao Gao, Enfang Sang, Zhengyan Shen. Comparison of EMD and complex EMD in signal processing [J]. Image and Signal Processing, 2008, 5(1): 141 – 145.
- [28] Xilin Gan, Weigen Huang , Jingsong Yang , Bin Fu. Internal wave packet characterization from SAR images using empirical mode decomposition (EMD) [J]. Image and Signal Processing, 2008, 4(4): 750 – 753.

- [29] Z. Wu and N. E. Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method[J]. IEEE Signal Processing Medical Signal Processing & Medical Imaging, 2009, 1(1): 1 - 41.
- [30] Maria E. Torres , Marcelo A. Colominas ,Gaston Schlott thauer , Patrick Flandrin. A complete ensemble empirical mode decomposition with adaptive noise[C]. ICASSP 2011,22(18):4144-4147.
- [31] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Publishing house of electronics industry, 2004, 9(3):258-266.
- [32] 芦涛, 王成儒. 基于 SVM 的汉语语音情感识别的研究[M]. 燕山大学硕士学位论文, 2007.
- [33] Skowronski D. Harris J G. Increased MFCC filter bandwidth for noise robust phoneme recognition[C].IEEE International Conference on acoustics Speech and Signal Processing,2002,10(7):436-442.
- [34] 陈卫东, 王晓亚, 解静. 基于 LPCC 的多语种识别算法[J]. 信号与信息处理, 2009, 39(9): 16-20.
- [35] 谭皇, 杨巨龙, 张玉军, 高勇. 基于 EMD 和 2.5 维谱的直升机声信号特征提取[J]. 舰船电子工程, 2010, (11):41-44.
- [36] 林婉如, 熊盛武, 谢哮虎. 局部经验模态分解算法[J]. 计算机工程与应用, 2011, 47(13):123-126.
- [37] 焦彦军, 胡春. 基于改进 EEMD 方法的数字滤波器[J]. 电力自动化设备, 2011, 31(11):64-68.
- [38] 曹机, 杨世锡, 杨将新. 大型旋转机械非平稳振动信号的 EEMD 降噪方法[J]. 振动与冲击, 2009, 28(9):33-38.
- [39] Venkat, Venkatasubramanian, Kng Chan. A Neural Network Methodology for Process Fault Diagnosis[J]. Journal of American Institute of Chemical Engineers, 1989, 35(12): 1993-2002.
- [40] 何永勇, 钟秉林, 黄仁. 基于椭圆单元网络的旋转机械多故障同时性诊断[J]. 振动工程学报, 1997, 10(2):131-138.
- [41] Zhou Jiemin, Lin Gang, Gong Shuli, et al. Application of multi-sensor data fusion based on fuzzy neural network in rotating mechanical failure diagnosis[J].Transactions of Nanjing University of Aeronautics Astronautics,2001,18(1):91-83.
- [42] Pan Dan, Zheng Qilun, Huang Qia et al. Fuzzy classification of fault diagnosis using ANN[J]. Journal of South China University of Technology, 1999, 27(4): 92-98
- [43] 李伟红, 汤海兵, 龚卫国. 公共场所异常声源定位中时延估计方法研究[J]. 仪器仪表学报, 2012, 8(4): 126-130.

附 录

A. 作者在攻读硕士学位期间发表的论文目录

- [1] **Zhilong Zhang**, Weihong Li, Weiguo Gong, Jianhua Zhong. An improved EEMD model for feature extraction and classification of gunshot in public places[C]. International Conference on Pattern Recognition (ICPR), 2012: 1517 – 1520.(EI 检索号: 20131016091884)
- [2] 李伟红, 刘鹏, 龚卫国, 汤海兵, 张智龙, 仲建华. 一种公共场所异常声音的识别与定位方法. (发明专利, 专利号: 201110444850)