****

**《检测技术及系统设计》**

**研讨报告**

**研讨题目：**

**院 系： 仪器科学与工程学院**

**专 业： 测控技术与仪器**

**小组成员： 学 号：**

本科生课程考试成绩单

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 院 系 | 仪器科学与工程学院 | 专业 | | 测控技术与仪器 | | |
| 学生姓名 |  | 学号 | |  | | |
| 课程名称 | 检测技术与系统设计（Seminar） | | | | | |
| 授课时间 | 2020年2月 — 2020年6月 | | 周学时 | 32 | 学分 | 2 |
| 简  要  评  语 |  | | | | | |
| 研讨题目 |  | | | | | |
| 成绩 |  | | | | | |
| 备注 |  | | | | | |

任课教师签名：

日期：

**摘要**

**关键词：**

**目录**

1. **绪论**

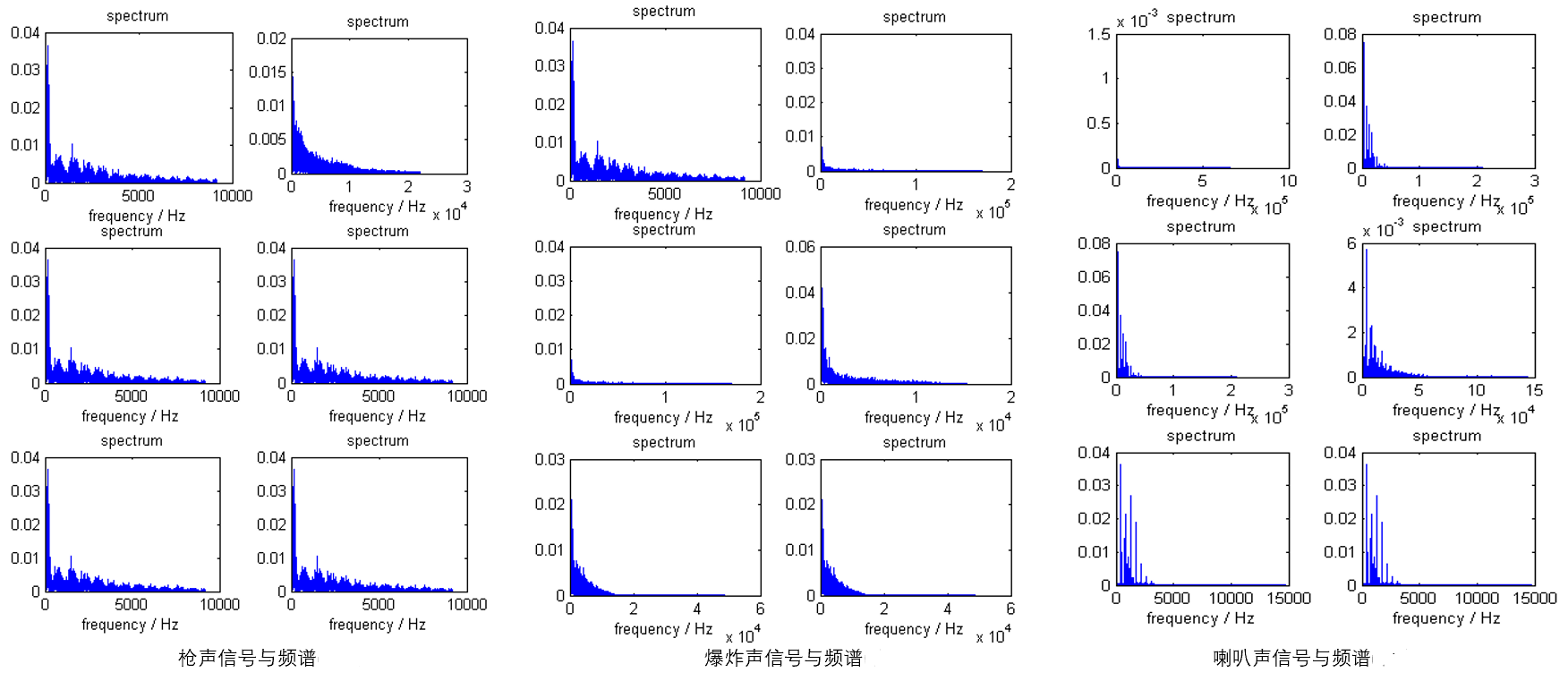
**正文字号小四，宋体，行间距1.2倍，图表需要标号**

1. **XXX**
2. **XXX**
3. **XXX**
4. **总结与心得**
5. **信号处理前端**

**声学事件检测系统的信号处理前端负责对所获取的声信号进行信号处理，并对目标信号片段进行提取。本章将结合TUT Acoustic Scene[克]和Freesound[克]两个声学事件数据库的部分信号样本，介绍信号处理前端的两个环节：滤波降噪和端点检测，以及对应的Matlab算法仿真结果。**

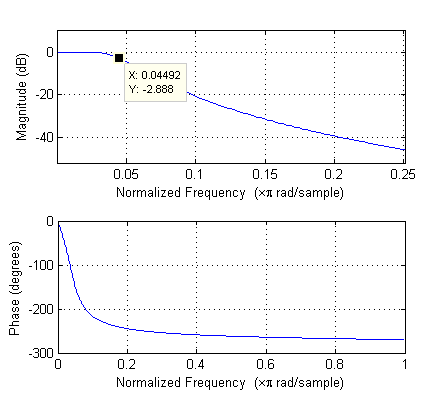
**X.1 滤波降噪**

**为了对所采集的声信号进行初步的去噪处理，以减小噪声信号对目标信号片段的提取和分类的干扰，信号处理前端设置滤波降噪环节。为设计合适的滤波器，首先需要对目标信号片段的特性进行分析，从图X（a）（b）（c）可以看到，枪声、爆炸声和汽车喇叭声信号的主要成分都在1000Hz以下。取1000Hz作为低通滤波的截值频率，能够去除高频噪声，同时不会损失目标信号的主要成分，不会使目标信号失真。**



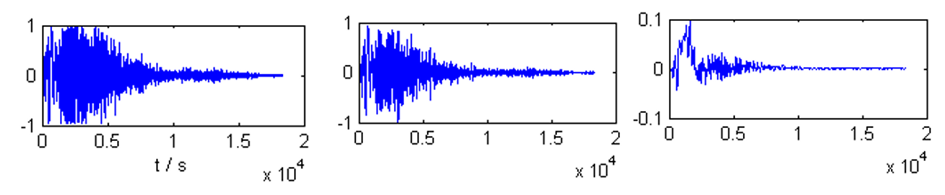
**图X（a） 图X（b） 图X（c）**

**对于滤波器的具体设计，采用3阶Butterworth滤波器进行低通滤波，并取截止频率为1000Hz, 如图X所示。**

****

**图X**

**通过低通滤波去除高频噪声后，目标信号的大致波形往往能显现出来。为了进一步获取目标信号的波形轮廓以便识别，采用高阶均值滤波进行处理。如图X所示，目标信号依次通过Butterworth低通滤波和1000阶均值滤波后，枪声信号的信号特征已经非常明显。**



**图X**

**对于相对纯净的信号片段，通过滤波、谱减法（Spectral Substraction）等前端信号处理手段的确能够获取目标信号的波形特征。尤其是通过低通滤波与谱减法相组合的信号处理方法，对于单发、背景噪声小、现场干扰少的情景，能够有效的获取枪声信号的波形。若再做进一步的波形分析、相关性分析或是波形匹配等基于波形的检测，一定程度上也能够实现目标信号的检测识别。对于靶场、测试基地、郊外等声源少、背景噪声小、干扰少的场景而言，这是一种可行的解决方案[蒋小]。然而，考虑到本系统的应用场景公共场所，对于现场信号而言，相对纯净的信号片段是几乎不可能得到的，相反，现场传声器所采集到的数据必然是不同类型声音混叠、背景嘈杂，甚至会有多个或多种目标信号并发的情况。另外，考虑到公共场所的场景，尤其是室内和街道场景，声音传播的多径效应明显，对于大功率的目标信号，这个问题更为显著。无论是多声源、多目标还是多径效应，都必然会造成波形的混叠。多个波形一旦混叠，基于波形的检测便无从谈起。**

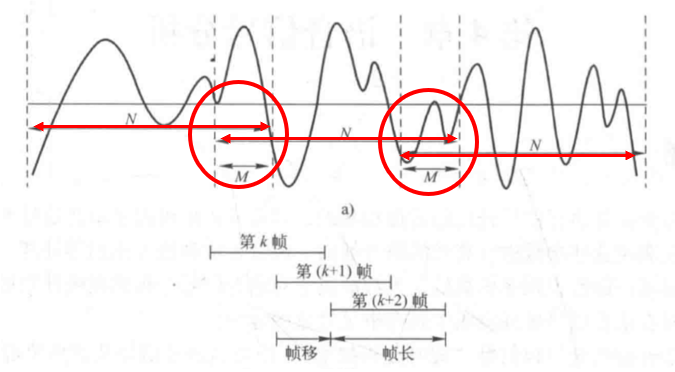
**结合上述分析，传统的基于波形的检测方法的应用场景与本系统的应用场景有较大差异，因此不采用基于波形的检测方法，而是通过目标信号的特性，将目标信号从背景中分离出来，再提取信号片段的声学特征，通过分类器对声信号进行检测。**

**X.2 端点检测**

**在X.1节中已经提到，本系统的声学检测算法首先分析目标信号的特点，按其特点将目标信号从背景中分离，再提取该信号片段的特征，最后使用分类器对声信号进行分类，从而达到对目标信号的检测。而将目标信号从背景中分离，可以转化为一个端点检测问题[张克]。本系统所针对的突发公共安全事件的声信号（枪声、爆炸等）均为大功率信号，前背景分离也就简化为在现场信号中，确定大功率信号端点的问题。**

**端点检测（Endpoint Detection）是指，从一段声音信号中准确找出目标信号的起始点和结束点，它的目的是为了使有效的目标信号和无用的背景声信号得以分离。端点检测的方法大体分为两类，一类是基于阈值的方法，该方法根据声信号和背景声的不同特征，把这些特征与设定的阈值进行比较，从而达到目标声音断点检测的目的。这种方法原理简单，运算方便，被广泛使用；另一类方法是基于模式识别的方法，需要估计目标信号和背景声的模型参数进行检测。基于模式识别的方法复杂度高，运算量大，往往很难被应用到现场实时声信号检测中[赵力]。因此，本系统采用基于阈值的端点检测方法。**

**在基于阈值的端点检测方法中，有两类最简单、运算复杂度最小、使用最广泛的方法：短时过零率分析和短时能量分析。短时分析贯穿了声学信号分析的全过程，这是因为有很多声学信号，从整体来看，其特性及表征其本质特征的参数均是随时间而变化的，因此它是一个非平稳过程，不能用处理平稳信号的数字信号处理技术对其进行分析出来。但是，虽然这些信号具有时变特性，但是在一个短时间范围内，其特性基本保持不变，即相对稳定，因而可以看作是一个准稳态过程，即具有“短时平稳性”。所以很多声学信号的分析和处理，往往建立在“短时”的基础上，即进行“短时分析”，将信号分为一段一段分析其特征参数。其中每一段称为一帧（frame）[赵力]。**



**图X**

**分帧（framing）可以采用连续分段的方法，但一般都采用如图X所示的交叠分段的方法，这是为了使帧与帧之间平滑过渡，保持其连续性。后一帧与前一帧的相对移动称为帧移。帧移与帧长的壁纸一般去0~0.5。分帧使用可移动的有限长度窗口进行加权的方法实现的，即用一定的窗函数来乘，从而形成加窗信号：**

**在声学信号数字处理中，常用的窗函数包括矩形窗、Hamming窗，他们的表达式如下：**

**矩形窗：**

**Hamming窗：**

**取帧长300，帧移100，Hamming窗进行分帧加窗[张克]，信号被分成很多个帧进行短时分析。需要特别指出的是，分帧加窗本质是利用了信号的短时平稳性，也就是说，一帧的长度相当的短，以至于信号在这一帧内是平稳的，其特性可以认为是不变的。相应的，其能量特征在帧内的各处没有显著变化，即一帧内，信号的能量分布近乎均匀，而不是像分帧前的整个信号（比如枪声信号）存在能量集中分布。**

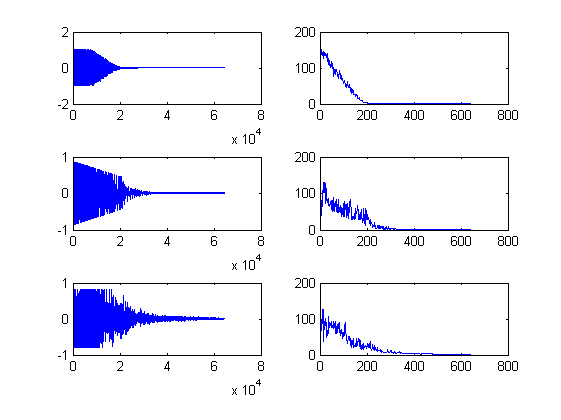
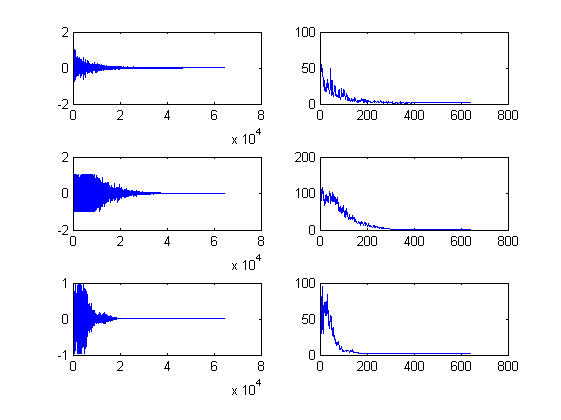
**过零率（Zero Crossing Rate，ZCR）标识一帧信号波形穿过横轴（零电平）的次数。对于连续信号，过零即意味着时域波形通过时间轴；而对于离散信号，如果相邻的采样值改变符号则称过零。因此，过零率就是样本改变符号的次数。声信号的短时过零率为[赵力]：**

**式中，是符号函数，即：**

**而对于短时能量分析，短时能量的定义如下：**

**图X**

**图X是对部分带噪声的枪声信号进行短时能量分析的结果。显然，短时能量是一个度量信号幅度值和功率变化的指标。考虑到本系统目标信号的功率特征明显，采用短时能量分析进行端点检测。**

****

**图X**

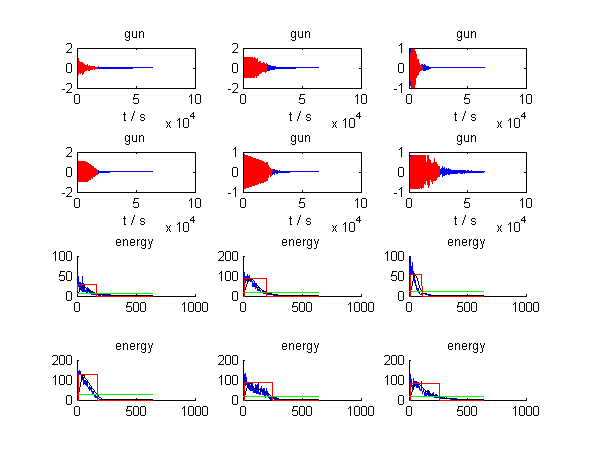
**对现场声信号进行短时能量分析后，通过合理的阈值划分，大功率的疑似信号就能从背景中被分离出来。对于阈值划分，采用自适应的短时能量阈值[张克]：**

**其中，为比例系数，为帧n的短时能量。仿真结果指出，取0.4左右时，枪声、爆炸声、汽车喇叭声三类大功率信号的端点检测准确率较高。短时能量阈值的意义在于，对于任意帧n，若其短时能量大于阈值，则认为该帧所对应的信号片段属于疑似目标信号的一部分，该帧对应的信号片段被保留；反之，若其短时能量小于阈值，则认为该帧所对应的信号片段不是疑似信号的一部分，该帧及对应的信号片段被抛弃。根据仿真结果，为了达到更稳定的阈值分割效果，可以在分割前先经过一个50阶的均值滤波，以提高稳定性。**

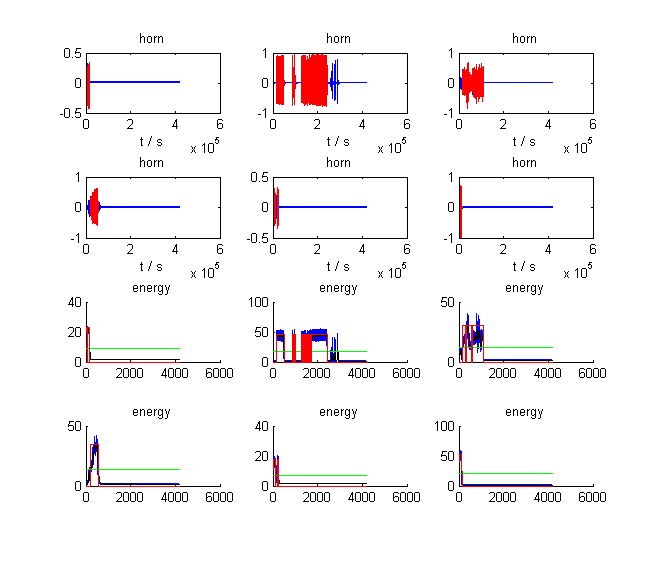
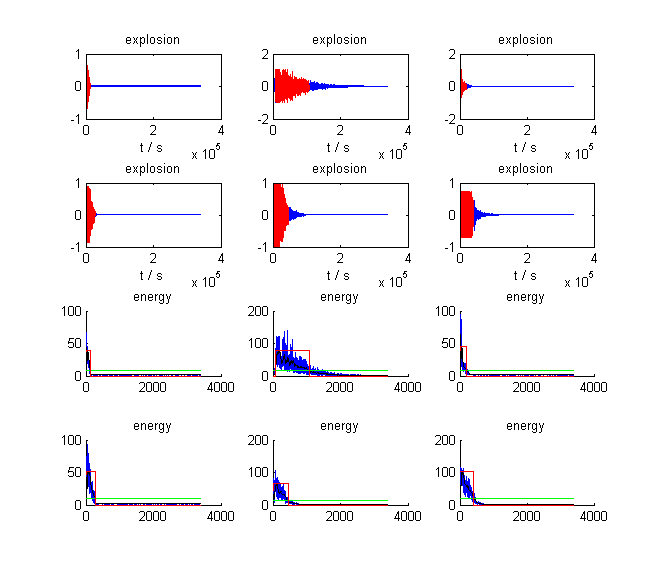
**在短时能量阈值分割中，有一个需要关注的问题是次要片段的影响。所谓次要片段是相对主要片段的概念。主要片段是指，产生目标信号的声学事件发生时，直接产生的声音，在传声器处采集到的的信号；而次要片段是指，在目标声学事件发生后，声音在空气中传播而出现的回响、多径等传播效应，被传声器采集到的二次信号。由于部分次要片段仍然具有较大功率，短时能量阈值分割会把这些片段当作疑似目标信号片段提取出来。但是，次要片段高度碎片化、持续时间短，并且难以作为判断声学事件是否发生的标准。因此，需要在阈值分割后将其滤除。**

**一种比较使用的方法是做持续时间处理[张克]。上文提到，次要片段持续时间短、高度碎片化，与主要片段相比，如果用合适的阈值对持续时间做分割，便能区分主要片段和次要片段。**

**其中，为经过阈值处理的能量的逻辑值，T为a的持续正值时间，为设定的持续时间阈值。仿真结果表明，针对已有的枪声、爆炸声、汽车喇叭声，取30个采样点做持续时间处理，能实现较好的次要片段滤除。如图X所示，(a)(b)(c)依次为部分枪声、爆炸声、汽车喇叭声的端点检测仿真结果。上方6个子图，蓝色为原始信号，红色为端点检测结果；下方6个子图，蓝色为短时能量分析结果，黑色为均值滤波结果，绿色为短时能量分割阈值，红色为持续时间分析后的端点检测结果。**

****

**图X(a)**

****

**图X(b) 图X(c)**

**X.3 特征工程**

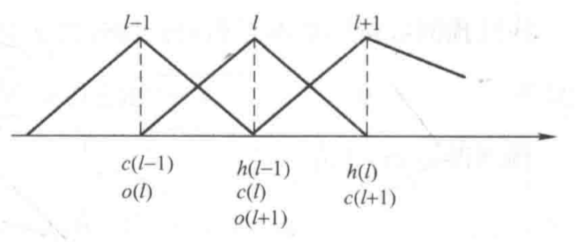
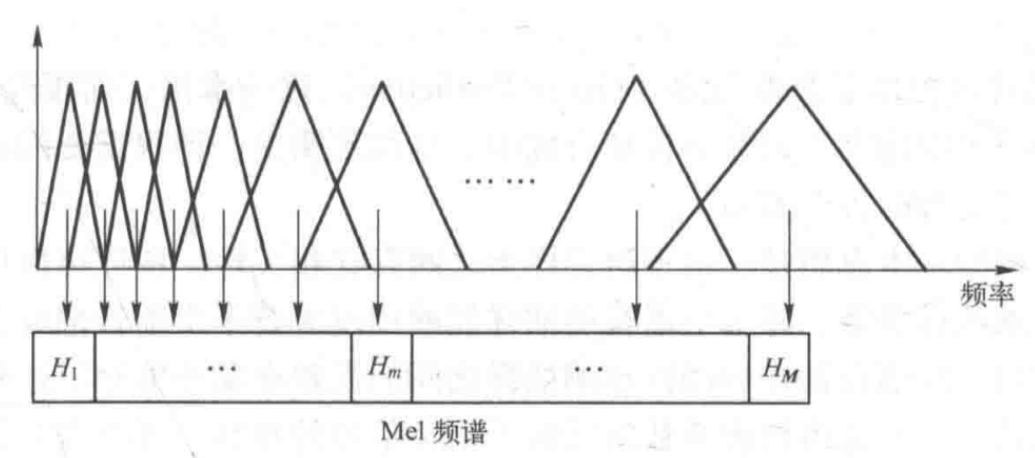
**在X.1、X.2节中，已经能够提取出疑似的目标信号，这些疑似的目标信号成为待检信号。将待检信号输入分类器做检测之前，需要先对其特征化。**

**心理学研究发现，针对不同的频率，人耳会存在不同的敏感度，因此线性的频谱并不能很好的反映人耳听觉的特性，其次由于掩蔽效应的存在，使得不同频率之间也不是相互无关的。由此人们提出了结合听觉特性的感知域的特征。这里我们主要采用Mel频率倒谱系数（MelFrequency Cepstral Coefficients，MFCC）特征[1940]。**

**研究表明，人耳在低频上的感知特性成线性关系，而对于高频信号的感**

**知则近似成对数关系[1940]。为了描述人耳对不同频率的特性，人们提出了可以描绘人耳感知域特性的Mel频率的概念，并得到从线性频率转换到Mel频率的表达式[朱强]：**

**其中，表示Mel频率，表示物理频率。根据人耳的听觉特性，所听到的声音的高低与声音的频率并不成线性关系，用Mel频率而符合人耳的听觉特性。临界频率的贷款随着频率的变化而变化，并于Mel频率的增长一致，在1000Hz以下，大致呈线性分布，带宽为100Hz左右；在1000Hz以上呈对数增长。类似于临界频带的划分，可以将声信号频率划分为一些列三角形的滤波器序列，即Mel滤波器组，如图X所示。**

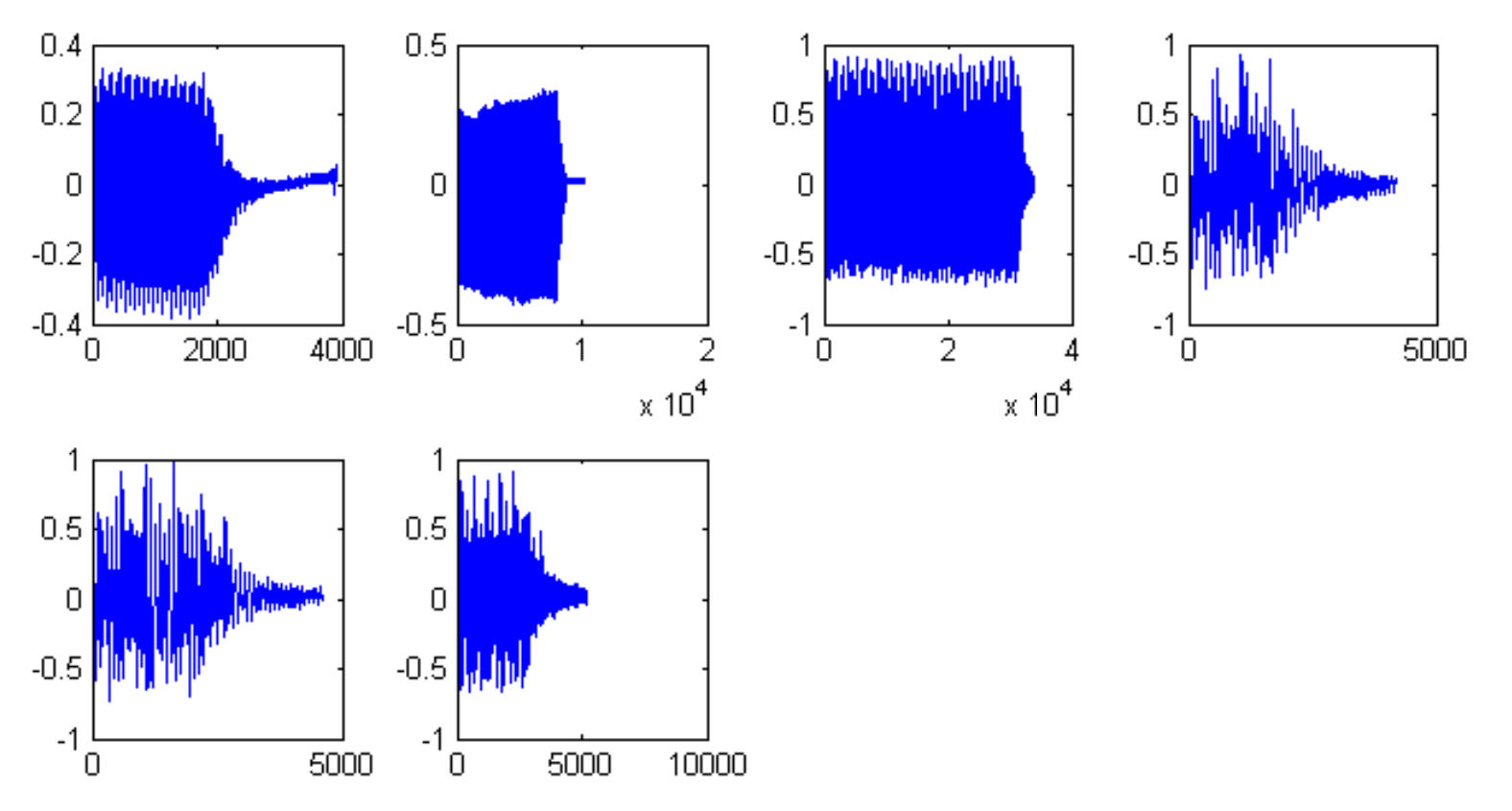


**图X(a) 图X(b)**

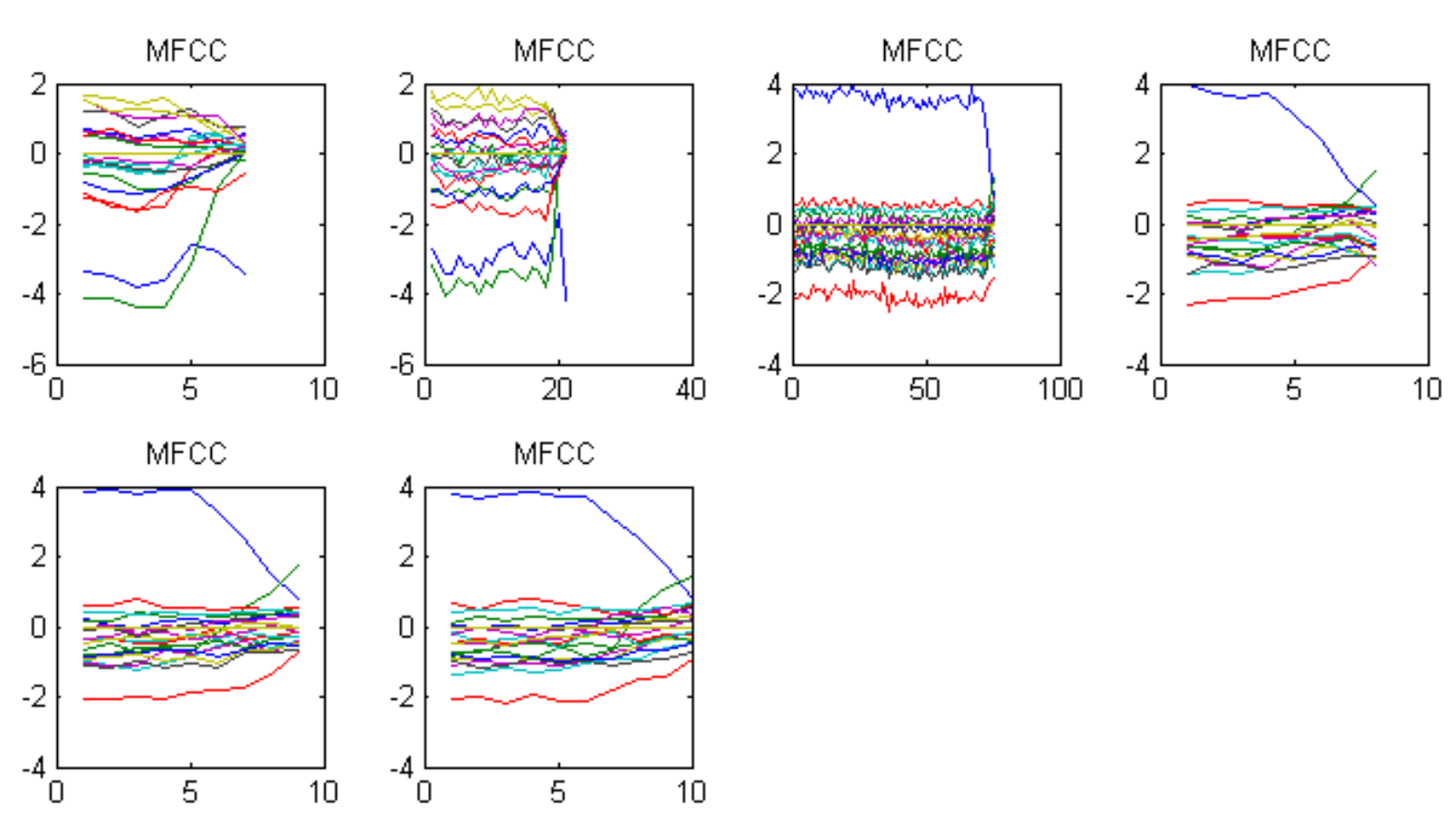
**取每个三角滤波器频率带宽内所有信号幅度加权和作为某个带通滤波器的输出，然后对所有滤波器输出做对数运算，再进一步做离散余弦变换即得到MFCC。MFCC参数计算过程的具体步骤如下[赵力]：**

1. **根据式（XXX），将实际频率尺度转换为Mel频率尺度。**
2. **在Mel频率轴上配置个通道的三角形滤波器，的个数由信号的截止频率决定。每一个三角滤波器的中心频率在Mel频率轴上等间隔分配，设、和分别是第个三角滤波器的下限、中心和上限频率，则相邻的三角形滤波器之间的下限、中心和上限频率如图X，并由如下关系成立：**
3. **根据语音信号幅度谱求每一个三角形滤波器的输出**
4. **对所有滤波器输出做对数运算，再进一步做离散余弦变换即可得到MFCC**

**此处取20，并选用三角形Mel滤波器组，对疑似目标信号片段进行MFCC特征的提取。图X(a)是6个通过端点检测从背景中提取出来的汽车喇叭声疑似信号，图X(b)是这6个疑似片段对应的MFCC仿真结果，不难看出其中模式的相似性。**

****

**图X(a)**

****

**图X(b)**

**X.4 分类器**

**经过滤波去噪和端点检测后，获得了高信噪比的目标信号前景，而经过MFCC特征化后，目标信号片段的信息维度被大大降低，有效的信息被进一步提取，无效的、不敏感的信息被抛弃。可以说，前面三个环节，都是在为分类器提供信噪比尽可能高、特征尽可能明显、维度尽可能合适的分类器输入。那么最后一个环节，便是设计一个合适的分类器，实现对目标信号的分类，实现对枪声、爆炸声、汽车喇叭声的声学检测。**

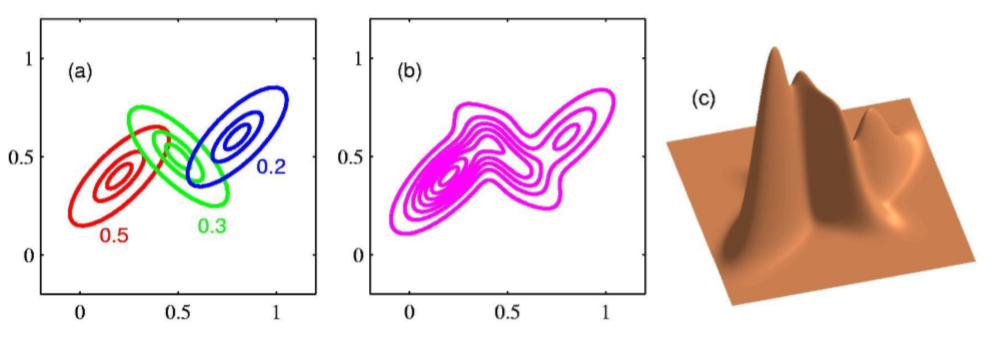
**分类（Classification）是一个根据给定数据样本进行类别预测的过程。分类的目标是，给当一个输入x，在K个类别（离散值）中指定它所对应的类别Ck。分类往往是很多数据处理与分析的最后一步。对应的，分类器（Classifier）则是实现分类任务的数学模型。目前，在统计学习、模式识别、智能感知等领域，分类器模型及其设计获得了广泛的关注。**

**目前声学事件检测的主流分类模型大多基于统计学习，其中应用较广泛的包括HMM、SVM和GMM[朱]。HMM对时序结构有良好的表示，其中的隐状态可以很好的刻画发音机制，使其在早期的语音识别等领域收到很大重视，也取得了很好的效果。但在枪声检测中，由于枪声发声机制相对简单，且具有突然性，另外，HMM（Hidden Markov Model）模型相对复杂，参数较多，使得HMM在本文所考虑的声学检测中效果并不突出；SVM（Support Vector Machine）是目前应用最广泛的分类器之一，基于判别式将样本空间进行划分。然而，一方面，在线性非线性、核函数设计、大规模样本训练等需要有较多考究，另一方面，使用单独的SVM并不能实现本环节所需要的多分类任务。GMM（Gaussian Mixture Model）即混合高斯模型，如图X所示，通过采用多个高斯分量的线性组合，可以近似任意概率密度分布，并且对数据没有特殊限制。GMM可以表示为式子（XXX）：**

**其中，是输入，是高斯分量数目，是高斯分量的权重，是高斯分量的概率密度。权重之间满足归一化条件：**

**多维高斯分布的概率密度：**

**其中，是维数，是协方差矩阵，是均值[bi]。GMM模型非常简单，只有高斯分量数目一个参数，且通过计算WSS能够很容易确定合适的值[meta]。模型训练好后，易于在检测系统嵌入式微机终端部署，能实现一站式检测[duf]。**

****

**图X**

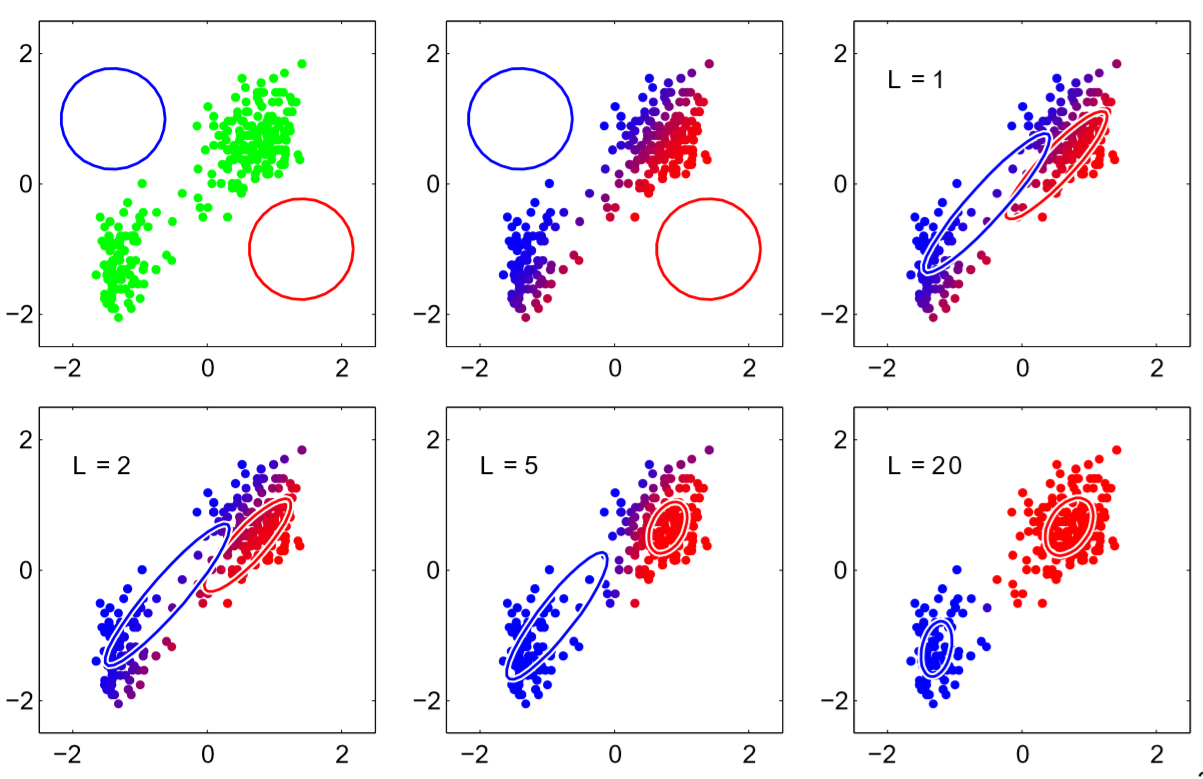
**训练GMM模型的关键在于聚类（Clustering）。对于共有N个样本，每个样本维度D的数据集，聚类的任务是将这N个样本划分到K个聚落中。一种广泛使用的算法是K-Means算法，分为如下几步[bishop]：**

1. **初始化，任意挑取K个聚类中心**
2. **将每个样本划归到最近的聚类中心**
3. **对各个聚落，调整聚类中心为划归到该聚落的所有样本的平均点**
4. **重复步骤（2）直到收敛**

**K-Means具有实现简单、收敛性可保证等优点。但在许多GMM模型的训练中，K-Means往往并不直接作为训练算法，而是用于对EM算法的参数做初始化[赵力][bi]。EM算法（Expetation-Maximization Algorithm）是训练GMM的常用方法。首先，对每个样本进行软划分（soft assign），然后基于软划分重新估计每个高斯分量的权重、每个高斯分量所占样本数、每个高斯分量的高斯分布参数。如图X所示，迭代循环直到收敛，步骤如下：**

1. **E环节：对样本点进行软划分**
2. **M环节：基于软划分对参数进行重新估计**

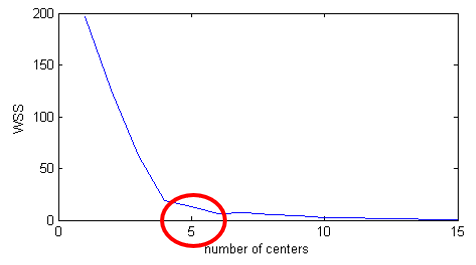
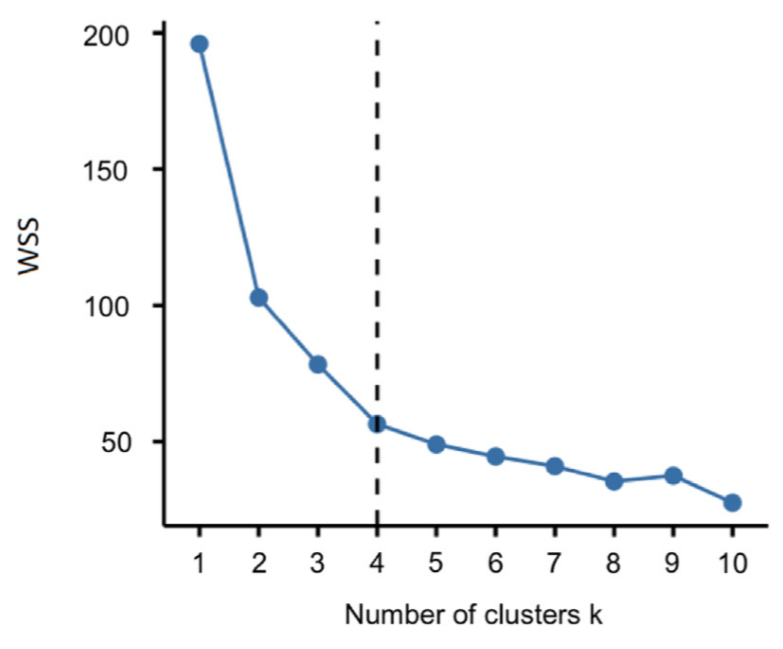
**GMM通过对不同类型的声音建模，使得每种声音在特征空间中都有一个独特的区域，由此可以根据测试数据在不同模型的似然度来判断当前数据的归属。通过EM算法训练GMM简单高效，对不同的数据规模没有特别要求。同比而言，GMM能够较好的覆盖系统设计所提出的要求。**



**图X**

**还有一个尚待解决的问题便是对于高斯分量数目的设计。正如前文已经介绍，高斯分量的数目本质上就是聚类中心的数目问题。直到目前，选择最优的聚类核心数目仍是聚类分析的核心研究话题之一。一种常规而典型的方法是比较不同聚类核心数目的有效性指数（validity index）。有效性指数通过聚类算法的结果计算得到，具有最优有效性指数的数目可以作为一个参数选择的方案。对于有效性指数，目前最简单而又广泛应用的是Elbow方法[met]。Elbow方法是基于类内方差WSS（Within-cluster Sum of Square）进行有效性指数的计算，WSS衡量了聚落的紧凑程度。显然，WSS越小，聚类效果越好。如图X(a)所示，Elbow方法的目标便是寻找一个聚类核心数目，使得新增一个聚类核心并不会再明显降低WSS。**

**如图X(b)所示，根据仿真结果可以发现，高斯分量数目达到5左右时，WSS几乎不会再随高斯分量数目的增加而显著减小，因此取5作为分类器的高斯分量数目。**

**图X(a) 图X(b)**

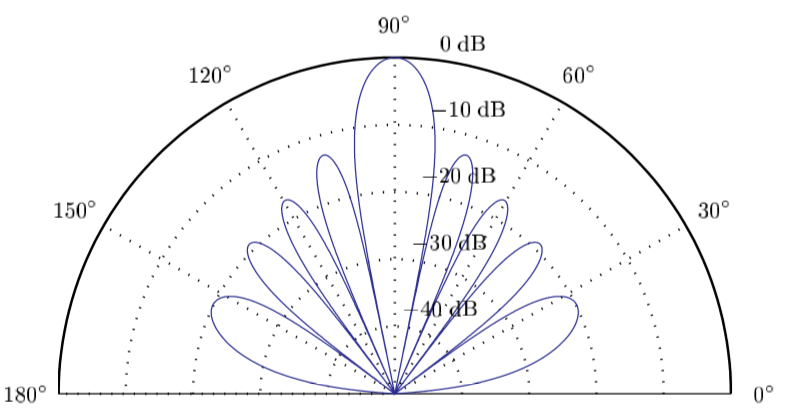
**对3类目标信号，分别训练GMM。对于每类目标信号，首先通过K-Means初始化训练聚类参数，然后通过EM算法训练GMM。由于总共有3个GMM分类器，每个GMM的输出都是一个描述待检信号是否属于该类别的似然概率，因此最后通过极大似然估计（Maximum Likelihood Estimation）来做分类：**

**基于Matlab平台展开算法的仿真验证。从TUT Acoustic Scene[克]和Freesound[克]两个声学事件数据库获取街道背景声、枪声、爆炸声、汽车喇叭声等数据，将目标信号与背景声混合后依次进行滤波去噪、端点检测、特征提取和分类器检测，仿真结果表明方案可行。**

**X.4 进一步的研究方向**

**关于“突发公共安全事件声学检测系统”的系统设计已经在前文全部介绍完毕。前文介绍的主要是一套经过算法验证、可操作性强、易于部署、成本低廉的解决方案。限于理论水平、实现复杂度和设计工程量，在鲁棒性、可拓展性、功能丰富性、场景适配性等问题上，目前的设计方案存在一些不足。本节就目前设计方案存在的问题与可优化的方案做简要介绍。**

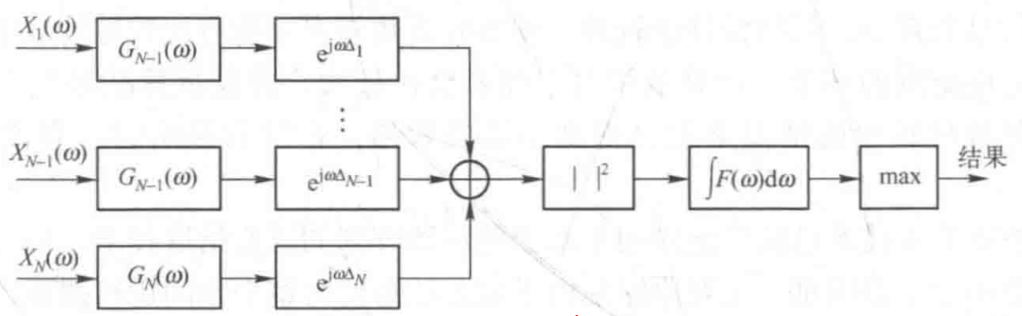
**麦克风阵列是目前声学信号处理常用的手段。基于麦克风阵列可以构建波束成形（Beamforming）算法。如图X所示，波束成形器（Beamformer）是一个空间滤波器（Spatial Filter），对传感器阵列输出进行组合运算来形成特定的、有指向性的特征。换句话说，波束形成就是利用传感器阵列和阵列信号处理方法，来对特定方位的信号进行增强和对准，使得信号处理具有指向性[mic]。**



**图X**

**一种对于目标信号提取的优化方案在于，使用掩膜数据训练基于BLSTM和全连接神经网络的信噪分离器[mic]，使用这个信噪分离器，可以分别得到目标信号及其功率谱、背景信号及其功率谱[mic]，再用信、噪功率谱设计波束成形器，进一步提高输入信号的信噪比[mic]。这能提高目标信号提取的正确率。**

**基于麦克风阵列，一个可以拓展的应用是声源定位。声源定位能够测定声学事件的发生方位，为检测系统提供更丰富的检测信息，例如并发检测的实现。基于麦克风阵列的声源定位方法包括：基于可控波束成形器的声源定位、基于到达时间差TDOA（Time Difference Of Arrival）的定位 [赵力]。基于可控波束的定位是最早的一种定位算法，采用波束成形方法，调整传声器阵列的对准方向，在整个接受空间内扫描，得到能量最大的方向即为声源方位。该方法是在满足极大似然准则的前提下，以搜索整个空间的方式，使传声器阵列所形成的波束能够对准信源；基于到达时间差TDOA的定位又称时延估计。所谓时延就是传声器阵列中不同位置单元接收到同一信源由于传输距离差异而产生的时间差。其中应用最广泛的时延估计方法是广义互相关GCC（Generalized Cross Correlation），算法流程如图X所示。得到时延估计后再做几何解算即可得到声源位置。**



**图X**

**（终端与数据中心/监控中心的通信，数据在中心的计算、存储，对应提问问题4）**

1. **问题解答**
2. **枪声识别的准确还是比较重要的，建议考虑多个传声器，以及相应的多数据**

**输入算法？**

**这是一个很好的建议。在多传声器设计基础上，波束成形、声源定位等更优质的检测方法和更多的升学检测功能得以开展。考虑到设计复杂度和理论水平深度，本文在X.4中针对多传声器设计的探究做了简要介绍。**

1. **软件算法的计算复杂度如何，可以在DSP芯片上进行计算吗？能否使用云**

**计算等技术进行改进？**

**问题16对类似问题做出了详细解答。**

**7. 滤波器可不可以考虑贝塞尔滤波器，使用高阶的巴特沃斯滤波器是否会使声音失真？**

**首先，X.1节滤波降噪中已经介绍，选用合适截止频率和合适阶数的Butterworth滤波器不会导致目标信号的主要成分造成失真。使用Bezier等其他低通滤波器，在不造成主要低频成分失真的前提下，也是可行的，并且对后续处理不会造成显著区别。检测的核心在于，对于提取出来的前进型号进行MFCC特征提取，依据MFCC 特征进行检测分类。可见，只要滤波器不显著性影响目标信号的功率特性、不显著改变目标信号的MFCC特征，各种低通滤波其实并无显著差异，起到的主要是一个初步处理和清洗的作用。这一点其实在课堂报告中已经提到过，但限于时间没能深挖做进一步的解释，可能引起了困惑。**

**10. 一般公共场合较为嘈杂，怎样排除同频或过大噪音的影响？**

**这是一个很好的问题。所谓“同频”问题没有太理解是想表达什么意思，姑且认位是干扰信号的频率与目标信号的低频主成分接近。的确，这种干扰不会被低通滤波去除。这个问题的解答需要分2类：首先，如果这个干扰是小功率干扰，那么无论他是同频还是不同频，都会被端点检测部分所排除，而不会影响到后续的分类器检测；其次，如果这个所谓“同频”干扰是一个大功率干扰，那么很不幸，端点检测很可能会把它当作疑似目标信号而从背景中提取出来，送到后面的特征化和分类器部分。解决这个问题的核心在于，尽管这个是一个“同频”干扰，但其声学特征（本系统采用了MFCC）与目标信号存在有显著差别，在后续分类器对其进行检测时，自然能分辨出它不是目标信号。课堂报告中对MFCC做了详细的解释，其实只要再做进一步推理思考就能得到这个结论。**

1. **采用GMM分类的稳定性如何，是否采用一定量且具有代表性的测试集（真**

**实数据）进行测试，测试的准确率如何？**

**不是很理解所谓“GMM的稳定性”，姑且认为是问训练算法的收敛性吧。X.4节中对训练GMM的K-Means和EM算法做了详细的介绍，这两个聚类算法的收敛性都是可以严格证明的，证明方法网上有很多，此处不再赘述。关于算法测试，X章中多次提及，在算法设计的时候引入了TUT Acoustic Scene[TUT]权威声学数据库的数据进行算法验证，此外也用到了较流行的共享数据库Freesound。限于个人电脑容量和算力，没有对大规模数据进行测试，对约60个合成样本进行了测试，正确率100%。关于这一点，课堂报告中软件架构综述就着重介绍了所引入的权威数据库做测试，部分测试结果也在课堂报告的最后一部分-分类器中给出了。**

1. **信号的采集、处理已经需要较贵的硬件资源，而且在后期的信号处理需要**

**进行机器学习和模型训练，那么更需要一个高算力的服务器，这样会进一步增加整个系统的成本，那么整个成本与系统成效是否能相互匹配？**

**在微机终端部署训练好的分类器模型并不是什么很复杂或者“算力要求很高”的难题。这里或许是把机器学习中的经典分类器与深度学习神经网络混淆了。再次强调，正如X.4中做出了详细的说明，本系统设计使用的分类器是GMM，不是感知机，更不是computation-consuming的深度神经网络，对计算设备的算力要求不高，但参数存储会有相当的开销。关于分类器参数引起的存储空间开销在硬件架构中已纳入考虑并做了适配的存储扩展。**

1. **公共场所的噪声非常复杂，如何从复杂的背景下提取出异常声音信号？**

**首先基于功率特征进行前背景分离，然后通过基于MFCC 特征的分类器进**

**行分类检测。依靠功率特征和声学特征的检测思想贯穿整个系统设计，不是简单的利用频率或单一特性，这个在在课堂报告、X.1、X.4、问题7、问题10都进行了解释，此处不再重复说明。**

1. **对于异常声音的特征表述，MFCC只反映了声音参数的静态特性，动态特**

**性怎么解决？**

**所采用的MFCC由连续T个时刻，所有I个余弦变换结果组成，最后输入GMM分类器，具有一定动态特性。**

**19.当环境中存在与枪声（或爆炸声等）频率相似但不具有危险性的声音时会不会造成错误检测？**

**课堂报告、问题10、问题17、问题21都在问一个所谓“频率”问题，问题大同小异，都是对基于MFCC分类与前端去噪有相当大的误解。前端去噪只是尽其所能去除一些不必要存在的高频干扰，不是通过前端滤波来滤除全部干扰，无论干扰是否存在，最后是依靠基于MFCC的分类实现分类检测，只要目标信号与非目标信号的功率和MFCC存在一定差异，就能区别开。**

1. **单点枪声识别效果较好，密集枪声是否可以很好检测？**

**密集枪声也能实现检测、分类、计数，考虑具体的枪声密度/射击频率，需要适当调整采样率和前背景分割阈值。**

**21. 所检测的目标信号频率处于低频段，而检测器所身处的场景中可能也存在较多的同样处于低频段的行人噪声，是否可以做到很好的区分**

**再次强调，检测系统不是依据频率进行前背景分离和目标信号分类，而是基于功率特征和MFCC特征。再次强调，低通滤波作用在于初步消除干扰而不是彻底消除干扰，低频段干扰只要其听觉特性与目标信号存在一定差异，基于MFCC的GMM分类能够区分识别。在课堂报告、X.1、X.4、问题7、问题10都对这个问题做了解释，此处不再赘述。**

1. **并发信号对系统的影响如何处理？**

**目前设计主要考虑单点信号的检测。X.4节介绍了基于麦克风阵列的声源定位的方法，可以实现并发检测。**

1. **在室内与室外是否需要训练不同的模型针对不同的声学环境？**

**必然是需要的，这涉及到一个模型的迁移问题。针对不同的场景，必然需要用不同的数据集进行训练，比如针对街道和交通道路场景，那么训练集中的背景数据便使用街道、交通道路的场景数据；如果要应用于室内场景，那么必然用商场、旅店、餐馆等室内场景数据训练背景的GMM，这个无需过多解释，毕竟一个系统不会同时在室外做检测同时又在室内做检测，退一步讲，即使这样一个系统需要既适配室外场景又适配室内场景，那自然需要两类场景都加入到数据集中，并且在场景切换时候需要适当的标定。需要强调的是，这并不意味着需要大幅调整系统设计，只是调整训练数据以及标定前背景分离所使用到的一些阈值参数（见X.X节）**

**参考文献**

[xx] 蒋小为. 枪声信号分析与预处理[C]. 中国声学学会第十一届青年学术会议会议论文集, 2015:509-512.

[xx] 卢慧洋. 枪声定位系统的研究与设计[D].西安科技大学,2016.

[xx] 佘大鹏. 基于多组麦克风阵列的枪声定位算法研究[D].国防科学技术大学,2015.

[1] 张克刚,叶湘滨. 基于短时能量和小波去噪的枪声信号检测方法[J]. 电测与仪表, 2015,52(S1):130-132+138.

[1] 赵力. 语音信号处理. 机械工业出版社

[1] 韩纪庆. 声学事件检测技术的发展历程与研究进展[J]. 数据采集与处理, 2016,31(02):231-241.

[] C. Clavel, T. Ehrette and G. Richard, "Events Detection for an Audio-Based Surveillance System," *2005 IEEE International Conference on Multimedia and Expo*, Amsterdam, 2005, pp. 1306-1309, doi: 10.1109/ICME.2005.1521669.

[] A. Mesaros, T. Heittola and T. Virtanen, "TUT database for acoustic scene classification and sound event detection," *2016 24th European Signal Processing Conference (EUSIPCO)*, Budapest, 2016, pp. 1128-1132, doi: 10.1109/EUSIPCO.2016.7760424.

[] Stevens, S. S., and J. Volkmann. “The Relation of Pitch to Frequency: A Revised Scale.” The American Journal of Psychology, vol. 53, no. 3, 1940, pp. 329–353. JSTOR

[1] 朱强强. 公共场所下的枪声检测研究[D].哈尔滨工业大学,2017.

[] C.M [Bishop,](https://cds.cern.ch/search?f=author&p=Bishop%2C%20Christopher%20M&ln=zh_CN) Pattern Recognition and Machine Learning, New York, NY : Springer, 2006. - 738 p.

[] Bernhard Scholkopf and Alexander J. Smola. 2001. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA.

[] Burges, C.J. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2, 121–167 (1998).

[] A. Dufaux, L. Besacier, M. Ansorge and F. Pellandini, "Automatic sound detection and recognition for noisy environment," 2000 10th European Signal Processing Conference, Tampere, 2000, pp. 1-4.

[] Pimentel, B. A., & de Carvalho, A. C. P. L. F. (2020). A Meta-learning approach for recommending the number of clusters for clustering algorithms. Knowledge-Based Systems, 105682.

[] Jacob Benesty, Jingdong Chen,Yiteng Huang, Microphone Array Signal Processing, Springer