突发公共安全事件声学检测系统

Acoustic Detecting System for Public Safety Emergency

报告人：招梓枫 林涵

东南大学仪器科学与工程学院
Instrument Science and Engineering

选题背景
Research Background

- 公共安全问题成为近年来聚焦的话题之一

- 目前的公共场所监控以视频方法为主，存在视野盲区、易受光照影响等问题。对于事件检测，还可能存在语义不明的问题，监控手段不够全面

- 对于突发公共安全事件（以枪击、爆炸为例），声学方法具有敏度高、成本低、全天候等特点，但目前缺乏声学检测以及视频-声学联动的监控方法，监控手段不够完善

研究思路
Research Roadmap

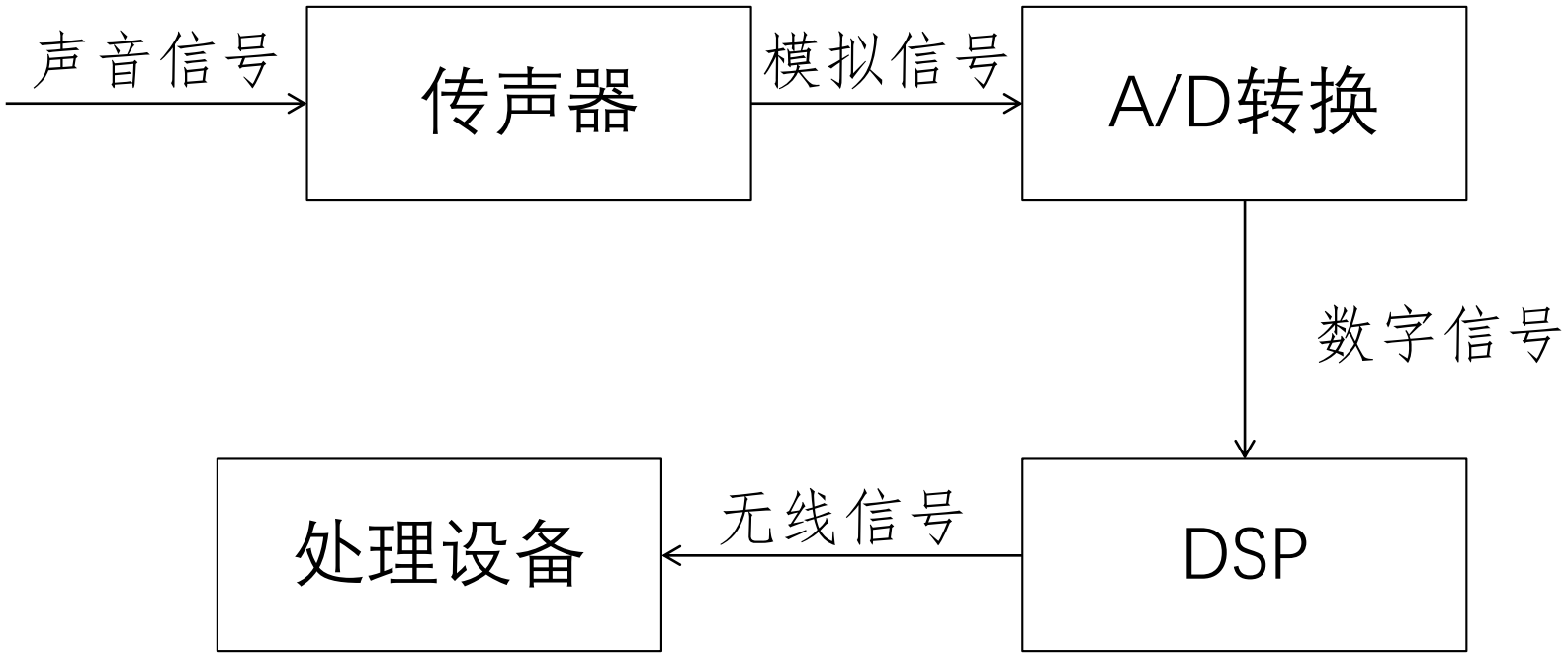
- 结合系统要求，分析了各类型传声器，并确定了具体选型
- 综合性能指标和成本因素，确定了ADC和DSP的选型，并设计了声学检测算法所依赖的硬件系统
- 结合突发公共安全事件和街道场景，设计了从降噪、端点检测、特征提取到分类器分类的成套软件解决方案
- 基于已有数据集对所设计的算法进行了测试

东南大学仪器科学与工程学院
Instrument Science and Engineering

目录
Category

硬件选型

概述
Abstract



传声器选型要求
Requirement for Microphone

- 低频性能好（放大、不失真）
- 大面积使用，价格不能过高
- 能耗尽量低
- 收音范围合适
- 在外界复杂环境中使用，必须受温湿度影响尽可能小
- 体积不能特别大
- 产品的质量尽量高、使用寿命尽量长、安装和维修成本低
- 承受声压尽可能大，满足使用需求
- 收录声压较高、脉冲较大的声源必须使用较低灵敏度麦克风

传声器选型原则
Criterion for Microphone

- 必要参数是否达标>稳定性>价格>其他性能参数

- 必要参数：最大声压级（AOP）、频率响应、瞬时响应

东南大学仪器科学与工程学院
Instrument Science and Engineering

传声器指标简介
Index for Microphone

- 分为三类来概述
- 技术指标
- 声学指标
- 市场指标

东南大学仪器科学与工程学院
Instrument Science and Engineering

传声器技术指标
Technical Index

- 灵敏度
- 方向性
- 信噪比（SNR）
- 最大声压级（AOP）
- 一致性
- 瞬时响应
- 电源抑制比（PSRR）
- 频率响应
- 总谐波失真（THD）
- 阻抗
- 动态范围
- 等效输入噪声（EIN）

传声器技术指标
Technical Index

- 灵敏度
- 灵敏度是指其输出端对于给定标准声学输入的电气响应。
- 单位声压的输出电压值

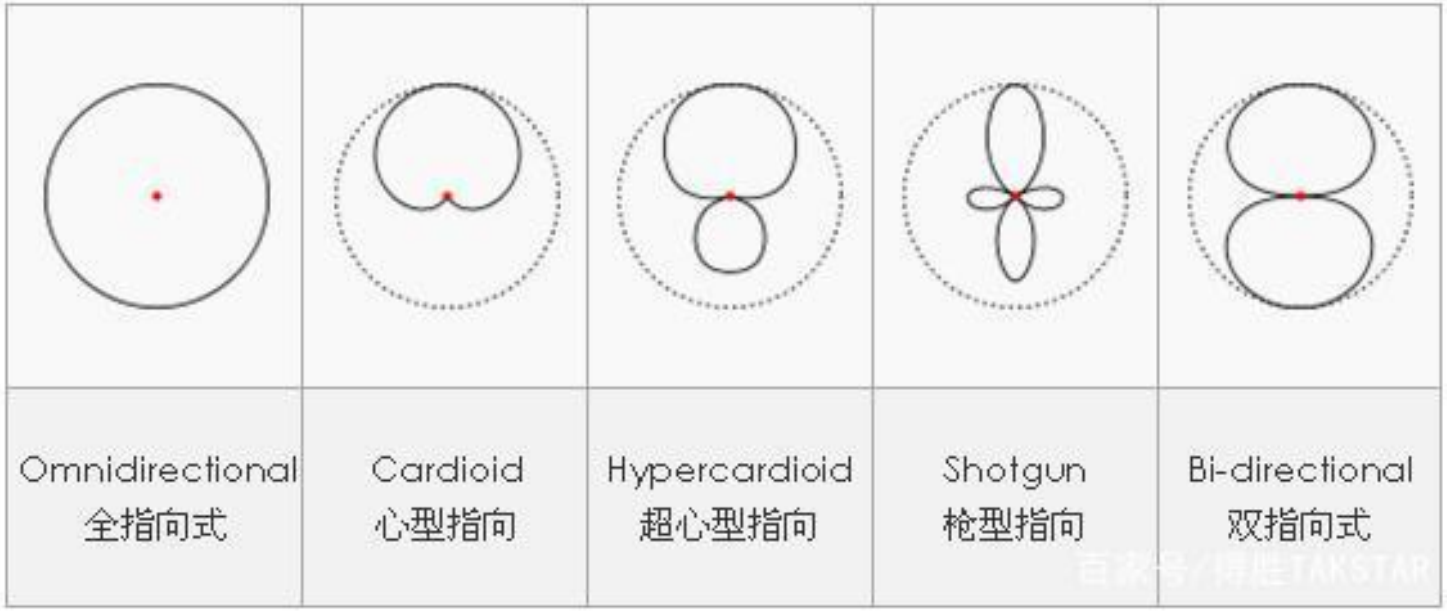
$$Sensitivity_{dBV} = 20 \times \log_{10}\left(\frac{Sensitivity_{mV/Pa}}{Output_{REF}}\right)$$

$$Sensitivity_{dBFS} = 20 \times \log_{10}\left(\frac{Sensitivity_{\%FS}}{Output_{REF}}\right)$$

东南大学仪器科学与工程学院
Instrument Science and Engineering

传声器技术指标
Technical Index

- 方向性
- 方向性描述麦克风的灵敏度随声源空间位置的改变而变化的模式。



传声器技术指标
Technical Index

- 信噪比（SNR）表示参考信号与麦克风输出的噪声水平的比值。
- 最大声压级（AOP）指的是麦克风输出THD等于10%时输入的声压大小（SPL）
- 一致性是麦克风在焊接后能否保持原有性能的指标

传声器技术指标
Technical Index

- 瞬时响应即麦克风对瞬态输入的电学反应
- 电源抑制比是麦克风输出对于电源输入噪声抑制能力的参数。
- 频率响应描述麦克风在整个频谱上的输出水平。

传声器技术指标
Technical Index

- 总谐波失真（THD）
- 阻抗
- 动态范围
- 等效输入噪声（EIN）

传声器声学指标
Acoustic Index

- 声学指标
  - 拾音轴内响应
  - 扩散声场频响
  - 离轴响应
  - 极性响应
  - 通道隔离度
  - 声反馈前增益
  - 离轴声染色
  - 极性图

传声器市场指标
Market Index

- 市场指标
  - 价格
  - 能耗
  - 稳定性
  - 良品率
  - 使用寿命
  - 供货能力

传声器具体要求
Requirement Details

- 技术指标要求

- 枪声在1m处声压级在130-155dB之间，根据声压的距离衰减公式每增加一倍距离衰减6dB，8m处大约在106-131dB，因此对传声器AOP要求至少在135以上
- 枪声爆炸等都是瞬时声波，需要瞬时响应性能好
- 对低频要求敏感，所以选用低灵敏度，大振膜传声器且无变压器输出
- 在300-7000频段范围内频响较好
- 全指向与一致性好

东南大学仪器科学与工程学院
Instrument Science and Engineering

传声器具体要求
Requirement Details

- 市场指标要求

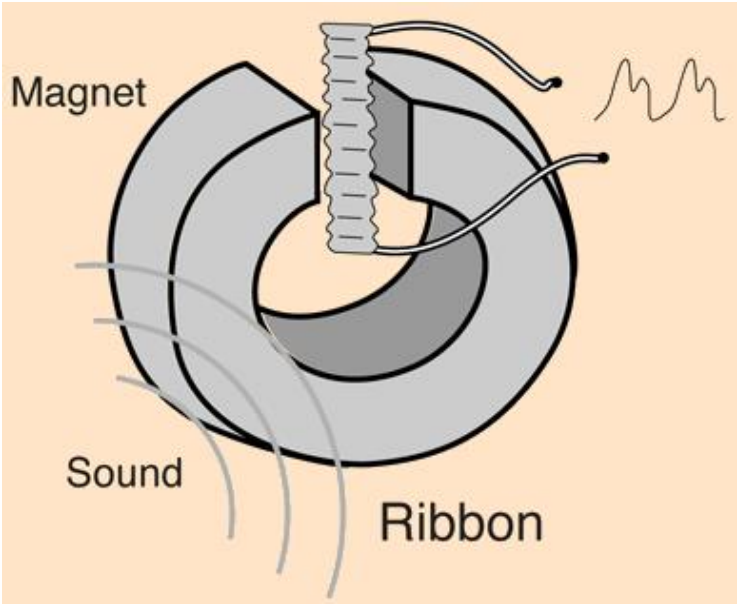
- 价格尽量中低、稳定性要求高、能耗尽可能低、使用寿命有保障、供货能力强

传声器种类选型
Selection on Types

- 根据声电转换分类

- 电动式（动圈式、铝带式），电容式（ECM、MEMS）、压电式（晶体式、陶瓷式、MEMS）、碳粒式、激光式、光纤式、矢量麦克风

传声器种类选型
Selection on Types

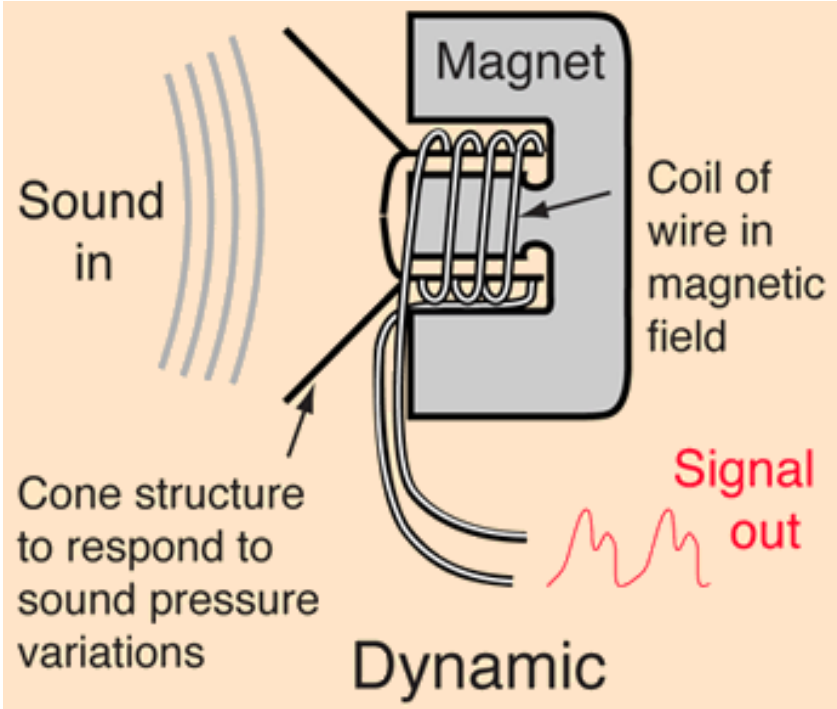


- 铝带式
- 优点：音质效果好、双向响应效果好、瞬态响应好
- 致命缺点:价格昂贵且铝片易受损伤、维修成本高、高声压会造成损坏
- 不考虑选用

传声器种类选型
Selection on Types

- 动圈式

- 优点：简单紧固、易于小型化、
  不需要额外供电、不易过载（失真）、
  指向性好
- 致命缺点：频响和瞬态响应不够好
- 不考虑选用

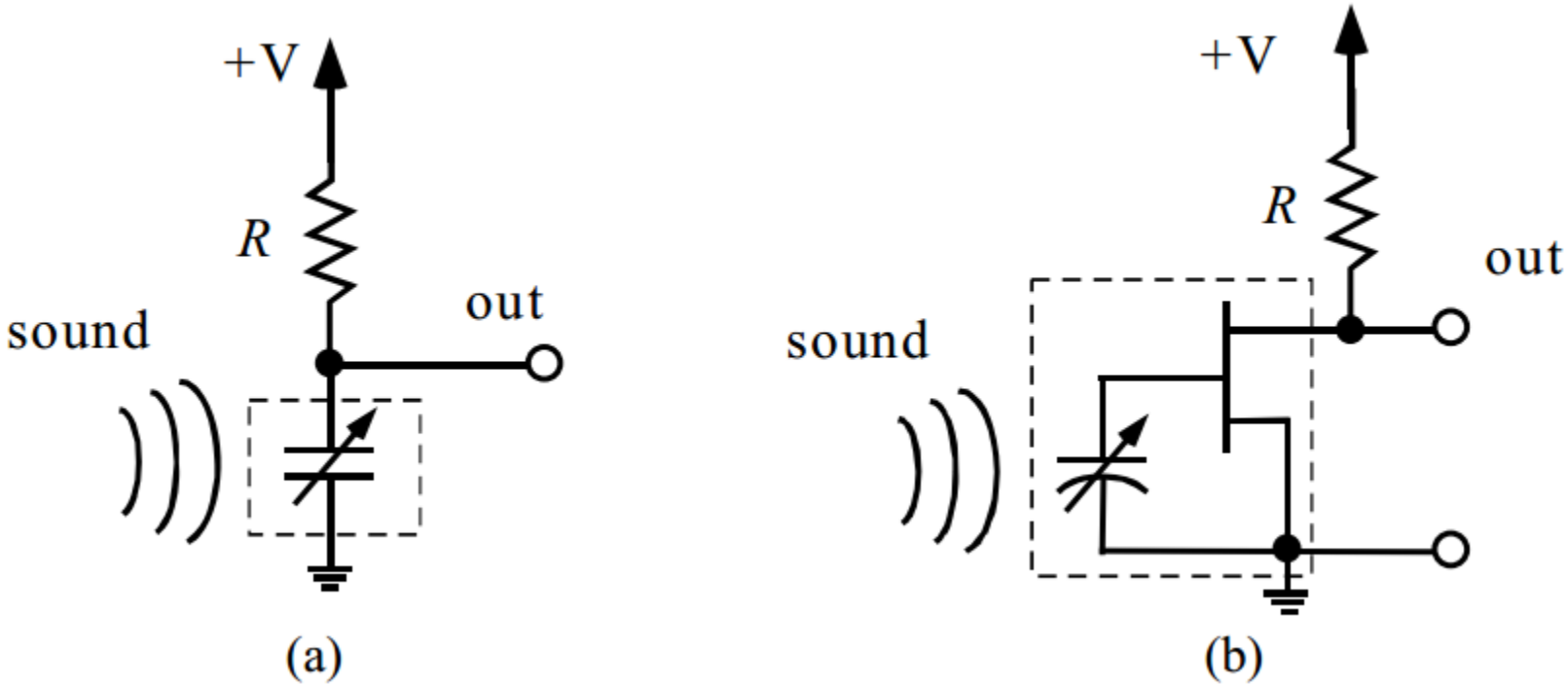


传声器种类选型
Selection on Types

- 电容式
- 优点：频响特性与瞬态响应好
- 缺点：价格较高、需要外部供电、受湿度影响

- 驻极体式（ECM）
- 优点：结构简单，体积小，价格低，瞬态性能好、频响特性好
- 缺点：受湿度影响大、一致性差、内部可能过载（失真）、灵敏度高

传声器种类选型
Selection on Types

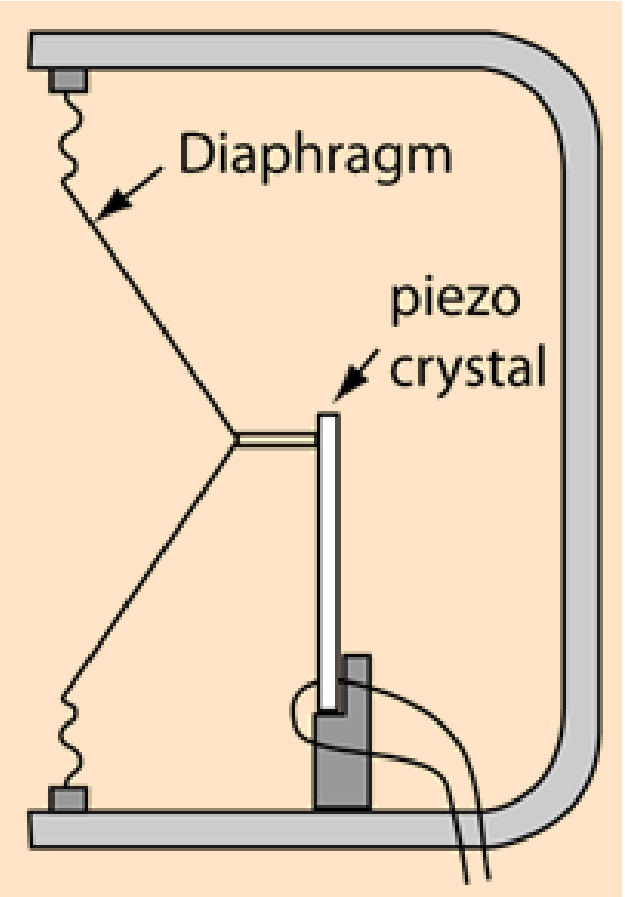


东南大学仪器科学与工程学院
Instrument Science and Engineering

传声器种类选型
Selection on Types

- 压电式

- 优点：输出电平高、价格低
- 缺点：频率响应较差、稳定性差



东南大学仪器科学与工程学院
Instrument Science and Engineering

传声器种类选型
Selection on Types

- MEMS式

- 优点：体积小、可SMT、产品稳定性好、不怕温湿度变化、一致性好
- 缺点：价格较高

传声器种类选型
Selection on Types

- 最终种类选型：MEMS压电式麦克风

- 优点：
- 1. 信噪比高
- 2. 受湿度、尘土、温度影响小
- 3. 一致性好
- 4. 支持单端与差分输出
- 5. 电源抑制比（PSRR）比传统的高30dB
- 6. 声学过载点（AOP）可以达到150dB的最大声压级

- 缺点：
- 价格高
- 瞬时响应与低频频响比驻极体差

传声器产品选型
Selection on Products

- Vesper公司的VM2020

- 超高声学过载点（AOP）
- 差分模拟输出
- 零件间差异小
- 耐用的压电MEMS构造

- 价格2.6美元

传声器产品参数
Product Parameters

SPECIFICATIONS

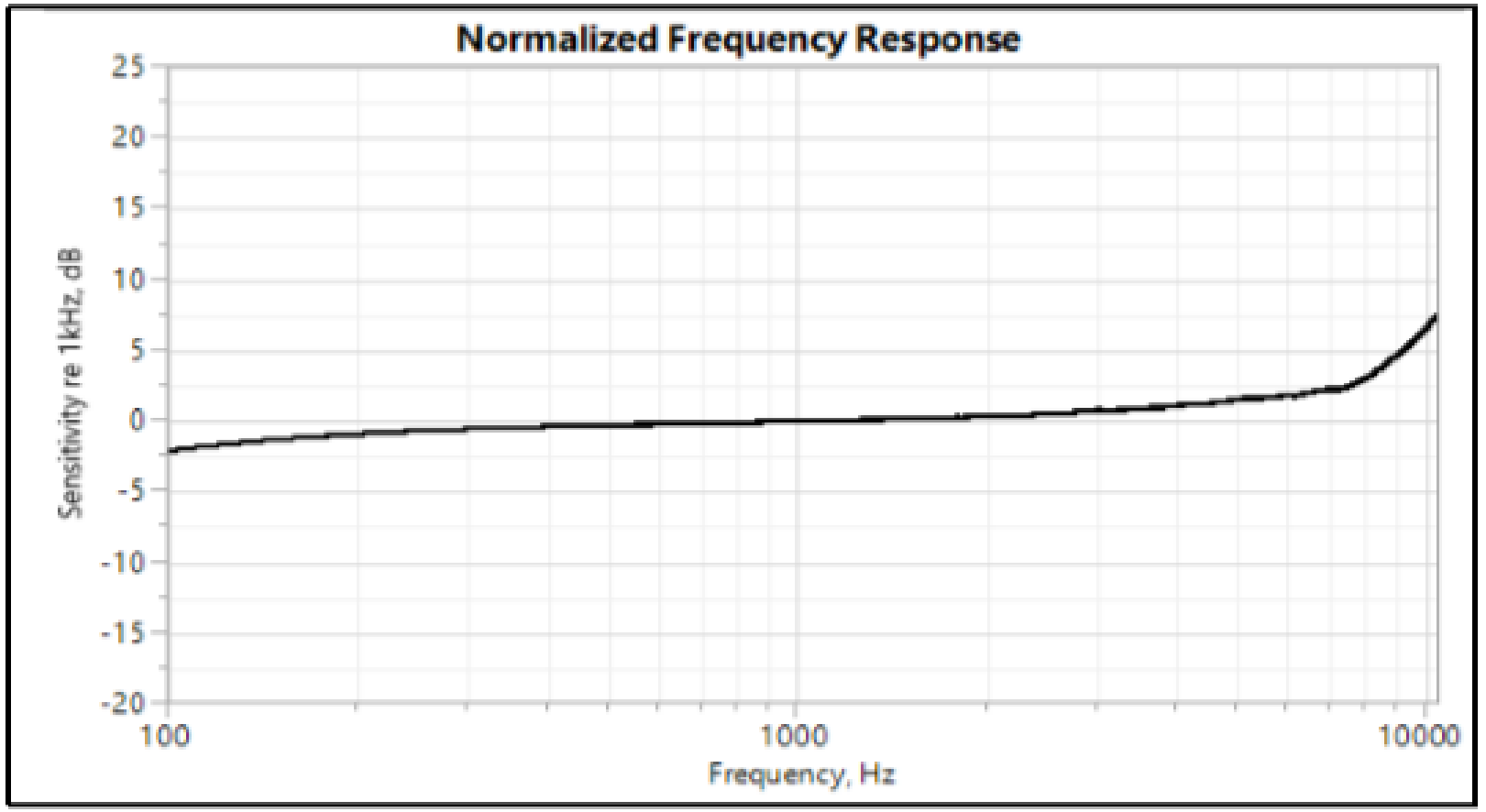
All specifications are at 25°C, VDD = 1.8 V unless otherwise noted

Parameter	Symbol	Conditions	Min.	Typ.	Max.	Units
Acoustic Specifications						
Sensitivity		1 kHz, 94 dB SPL	-66	-63	-60	dBV
Signal-to-Noise Ratio	SNR	94 dB SPL at 1 kHz signal, 20Hz to 20kHz, A-weighted Noise		50		dB(A)
Total Harmonic Distortion	THD	94 dB SPL		0.1		%
Total Harmonic Distortion	THD	149 dB SPL		1		%
Acoustic Overload Point	AOP	10.0% THD		152		dB SPL
Roll Off Frequency		-3dB at 1KHz			80	Hz
Directivity			Omni			
Polarity		Increase in sound pressure	Increase in output voltage			
Electrical Specifications						
Supply Voltage			1.6	1.8	3.6	V
Supply Current		$V_{Supply} \leq 3.6$ V		248		µA
Power Supply Rejection Ratio	PSRR	VDD = 1.8, 1kHz, 200mV$_{PP}$ Sine wave		90		dB
Power Supply Rejection	PSR	VDD = 1.8, 217Hz, 100mV$_{PP}$ square wave, 20 Hz – 20kHz, A-weighted		-112		dB(A)
Output Impedance	Z_{OUT}			1100		Ω
Output DC Offset		Both Vout+ and Vout-		0.8		V
Startup Time		Within ±0.5dB of actual sensitivity		200		µS

- 灵敏度-63dBV 较低
- 信噪比50dB(A) 较低
- AOP 152dB SPL高
- PSRR 90dB 高
- 响应时间200us 标准
- 阻抗1100Ω
- 指向性 全指向

东南大学仪器科学与工程学院
Instrument Science and Engineering

传声器产品参数
Product Parameters



Normalized Frequency Response

东南大学仪器科学与工程学院
Instrument Science and Engineering

DSP要求
Requirement for DSP

- 精度满足要求
- 处理速度满足要求
- 足够的外设资源

DSP种类
DSP Types

- 按数据格式分为

- 定点式
- 优点：体积小、功耗低、价格低、接口多、结构简单

- 浮点式
- 优点：运算精度高、动态范围大、地址总线宽（寻址空间广）

DSP选型
Selection on DSP

- TMS320F2812
- 定点 32 位
- 性价比高（几十人民币）
- 处理性能可达150MIPS
- IO口丰富，两个串口
- 两个独立的采样保持电路
- 哈佛总线结构，快速中断响应
- 片内128k*16位的片内FLASH，18k*16位的SRAM。
- 4M 线性程序与数据寻址空间

DSP拓展
Expansion on DSP

- 分类器参数所占空间较大
- 选用一片SRAM　　IS61LV25616AL
- 选用两篇FLASH　　SST39VF800

- 电源匹配
- AMS1117将5V转化为3.3V IO口电压及1.9V MIC及内核电压

ADC选型
Selection on ADC

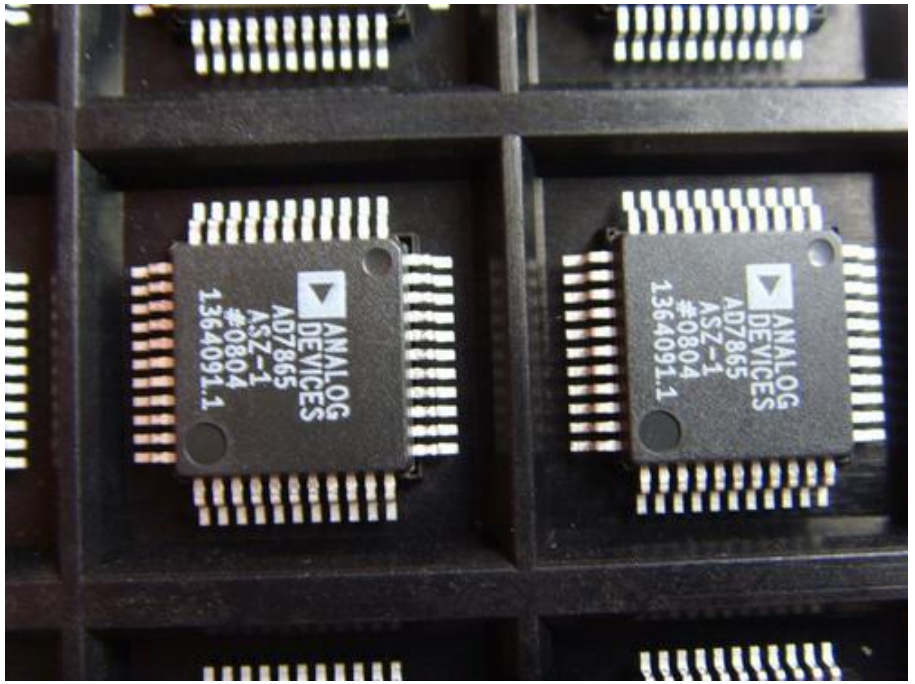
- 精度
- 速度
- 成本
- 性能

低速A/D	高速A/D
线性误差	线性误差
微分误差	微分误差
电源电流	电源电流
功 耗	功 耗
转换时间	转换率
失调误差	失调误差
增益误差	增益误差
	信 噪 比
	信噪失真比
	无杂散动态范围
	总谐波失真
	二次谐波
	三次谐波

东南大学仪器科学与工程学院
Instrument Science and Engineering

ADC选型
Selection on ADC

- AD7865-1
- 高速、4通道、14位
- 采集速度 350ksps
- 高输入范围（10V）
- 低功耗
- 价格相对较低

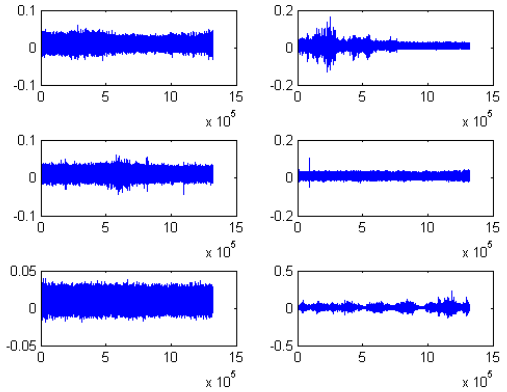


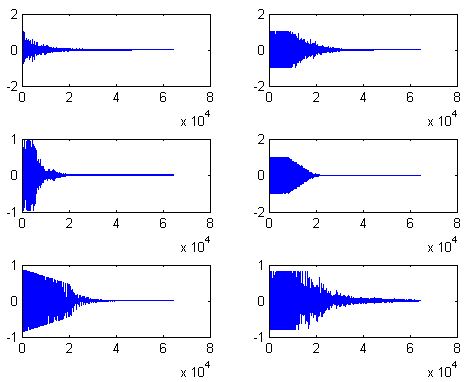
软件架构及算法仿真

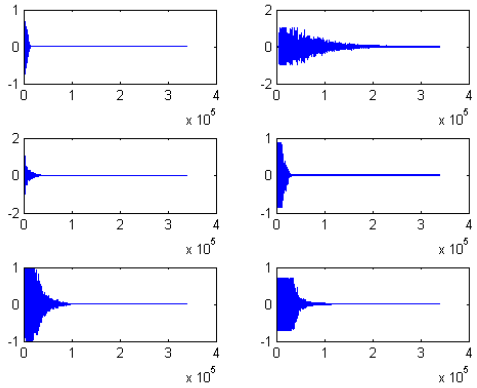
软件架构 Software Architecture
综述 Abstract

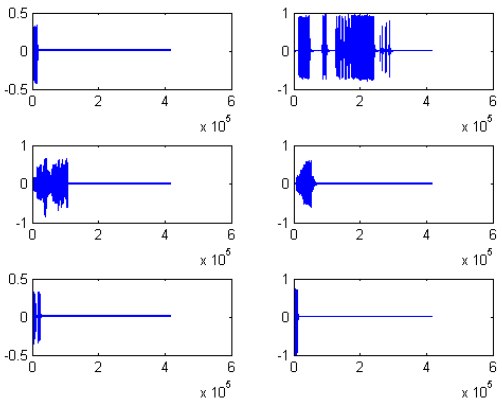
实现目标
- 针对街道、交通道路等公共场所场景，基于所设计硬件系统，设计相应的突发事件声学检测解决方案

综述
- 综合已有的枪声检测、声学事件检测、语音识别等领域的研究，结合项目背景，设计了从滤波去噪到端点检测的前端信号处理算法，确定了特征工程和分类器的选用方案并进行了相应测试
- 基于Matlab平台进行信号分析并设计了4个模块对应的仿真程序（检测系统部署时移植到DSP上）
- 从TUT Acoustic scenes[12]，Freesound[13]等声学事件数据库获取信号样本，加噪并混叠后对所设计的检测算法进行了测试


部分加噪后的街道测试信号


部分加噪后的枪声测试信号


部分加噪后的爆炸测试信号


部分加噪后的鸣笛测试信号

[12] TUT database for acoustic scene classification and sound event detection, EUSIPCO, A Mesaros
[13] Freesound technical demo, ACM international conference on Multimedia, F Font

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
综述 Abstract

高频段信号为噪声，用滤波器对信号进行初步清洗

根据疑似信号的特点，检测出疑似信号片段，并从声音序列中抽取出来

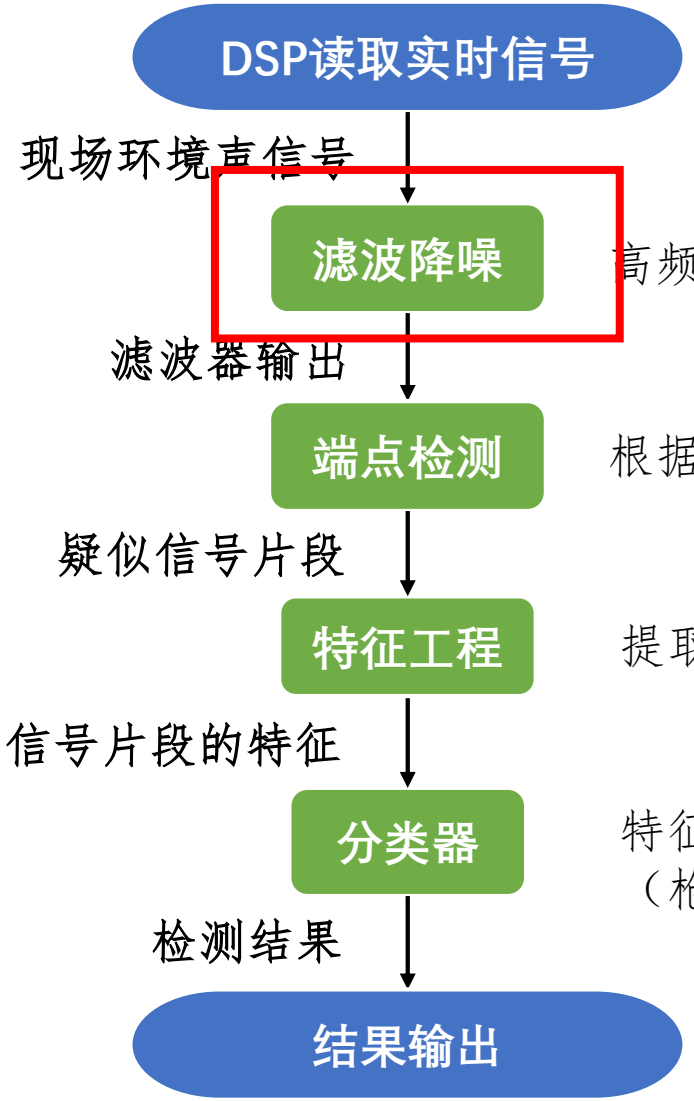
提取每个疑似信号片段的特征

特征作为分类器的输入，用事先训练好的分类器进行分类
（枪声/爆炸声/车喇叭），从而实现目标信号的检测

- 滤波降噪
- 端点检测
- 特征工程
- 分类器

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
滤波降噪 Filtering & Denoising



高频段信号为噪声，用滤波器对信号进行初步清洗

根据疑似信号的特点，检测出疑似信号片段，并从声音序列中抽取出来

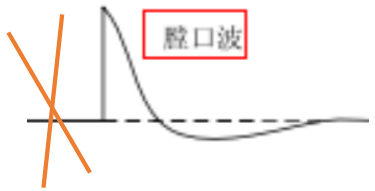
提取每个疑似信号片段的特征

特征作为分类器的输入，用事先训练好的分类器进行分类（枪声/爆炸声/车喇叭），从而实现目标信号的检测

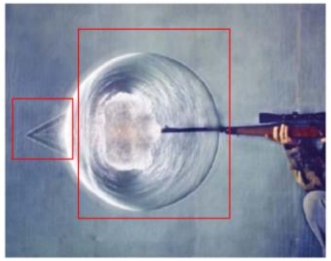
报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

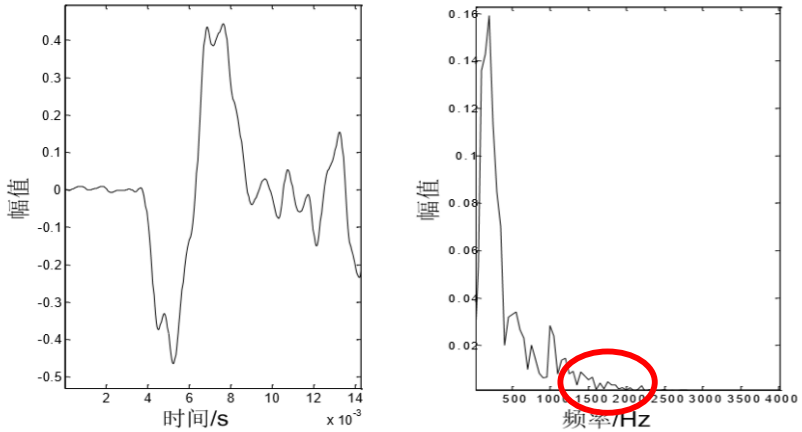
软件架构 Software Architecture
滤波降噪 Filtering & Denoising

- 典型的枪声信号
  仅考虑**膛口波(muzzle blast)**，典型的枪声信号是一个**负压-正压**的过程
  理论波形的的频率集中在**低频**[1][2]，若要在检测的基础上做进精确定位可以综合**膛口波**与**马赫波(shock wave)**做分析[12]

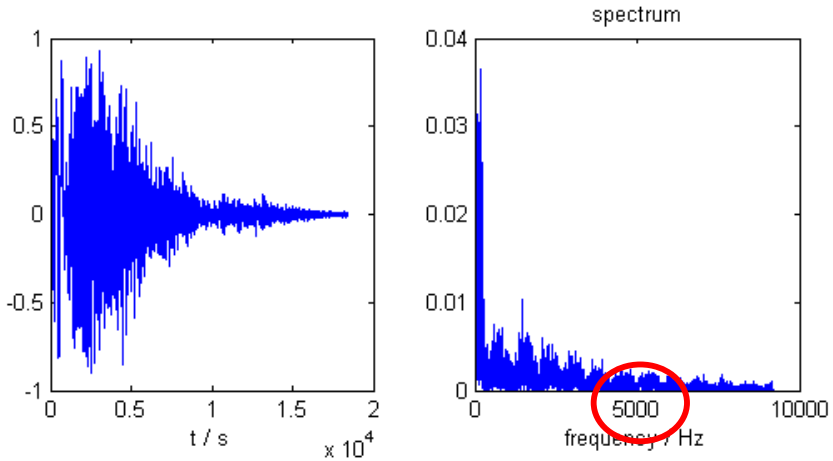


膛口波波形[2]（有误）


膛口波与马赫波[2]



低噪声下膛口波的波形模式与频谱[1]

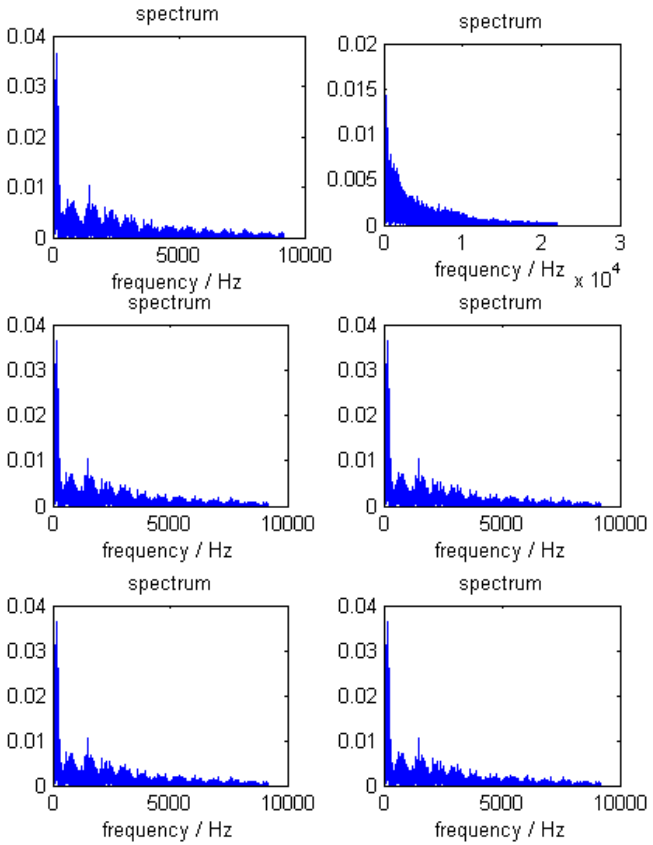


枪声信号数据

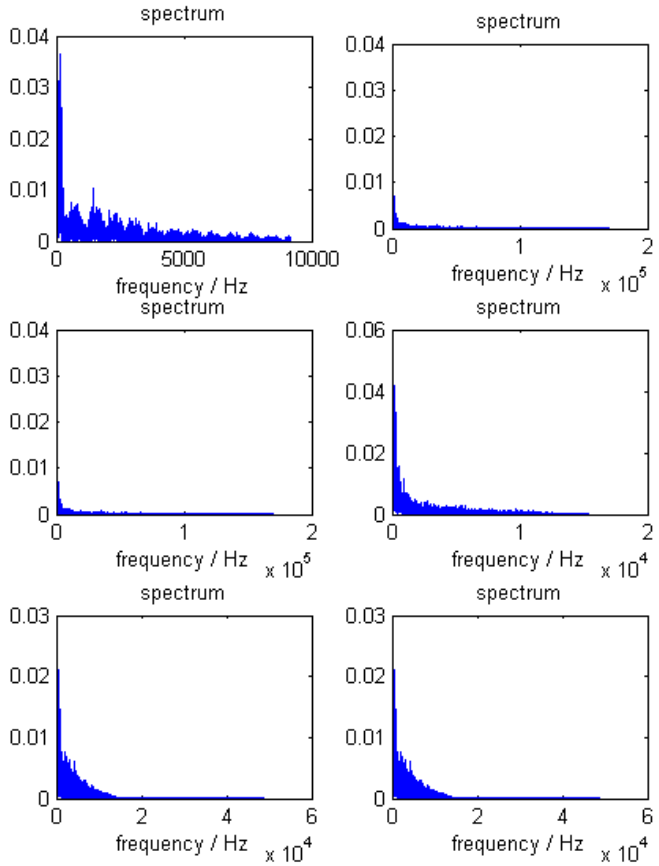
[1] 枪声信号分析与预处理，声学技术，蒋小为，张文等
[2] 枪声定位系统的研究与设计，西安科技大学硕士学位论文，卢慧洋
[12] 基于多组麦克风阵列的枪声定位算法研究，国防科技大学硕士学位论文，佘大鹏

软件架构 Software Architecture
滤波降噪 Filtering & Denoising

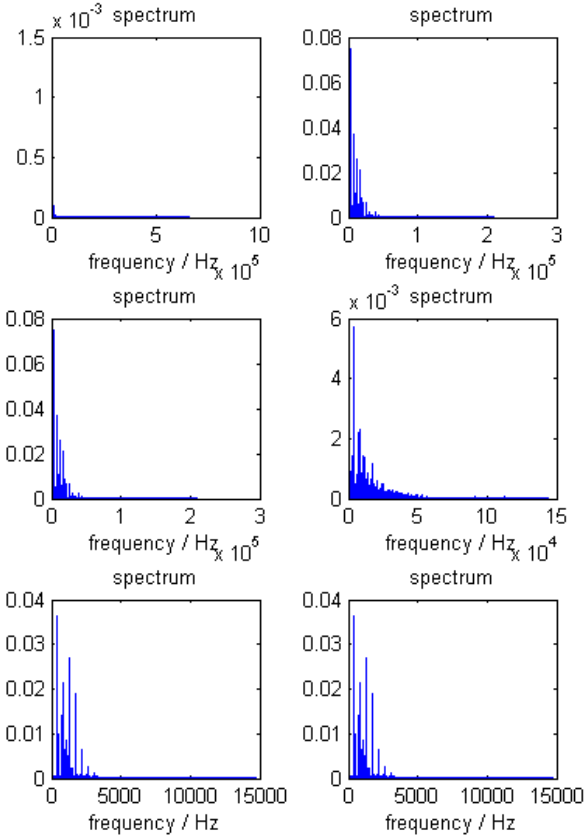
- 根据3类声信号的频谱特征设置截止频率消除高频噪声，同时不会使3类声信号失真
- 白噪声、椒盐噪声：端点检测中会消除其影响
- 对3类声信号进行频域分析：



枪声信号与频谱(仿真)



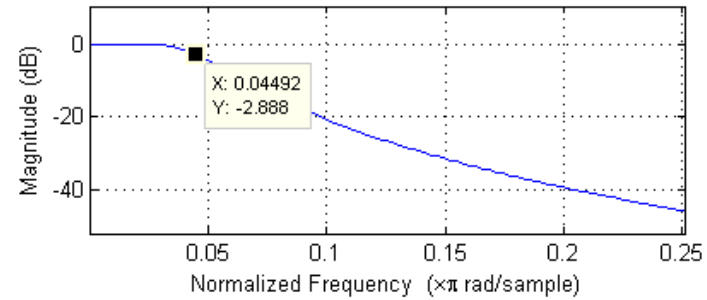
爆炸声信号与频谱(仿真)



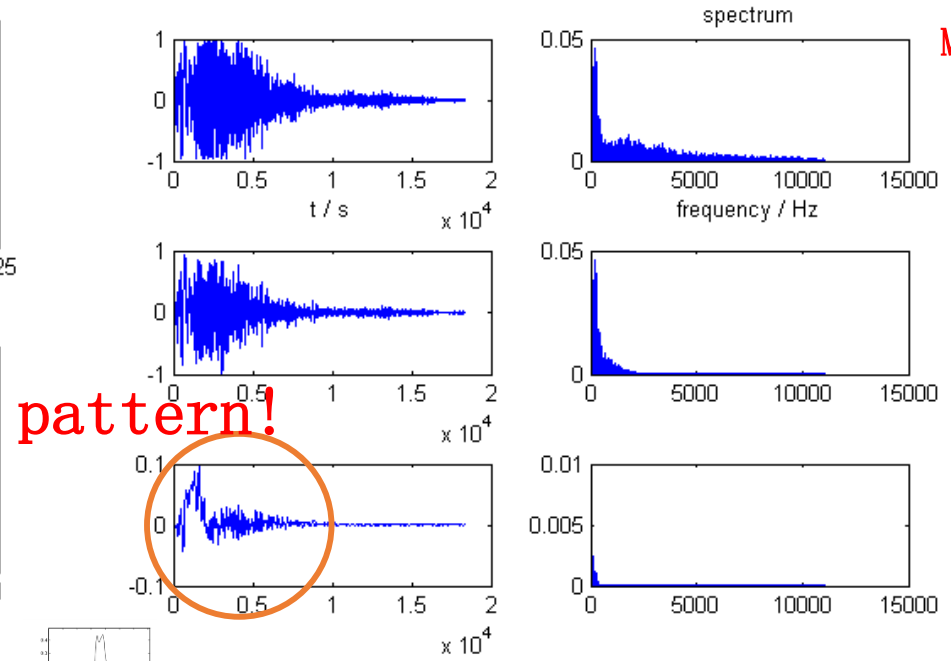
喇叭声信号与频谱(仿真)

软件架构 Software Architecture
滤波降噪 Filtering & Denoising

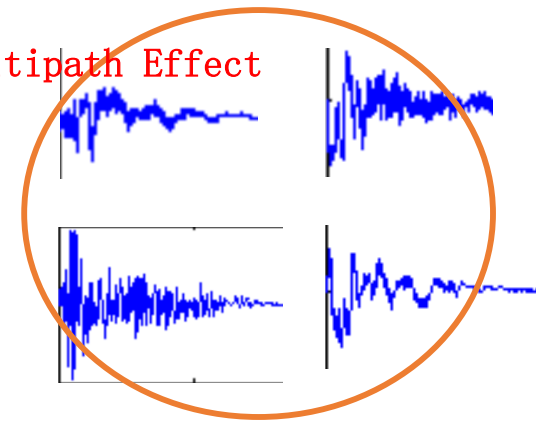
- 滤波降噪方案
  Butterworth Filter实现**低通滤波**($f_{cutoff} = 1\text{kHz}$)
- 考虑使用更好的滤波方案？直接把枪声波形过滤出来后进行**相关分析(correlation)**？
  均值滤波(order≥1k)、**谱减法**[1]等方法的确有可行性，但仿真中出现了各种各样的波形……
  另外，波形模式匹配较难解决低频干扰和多径效应，也无法解决多个目标声信号混叠的问题



Butterworth Filter (仿真)



依次通过低通滤波和均值滤波(仿真)

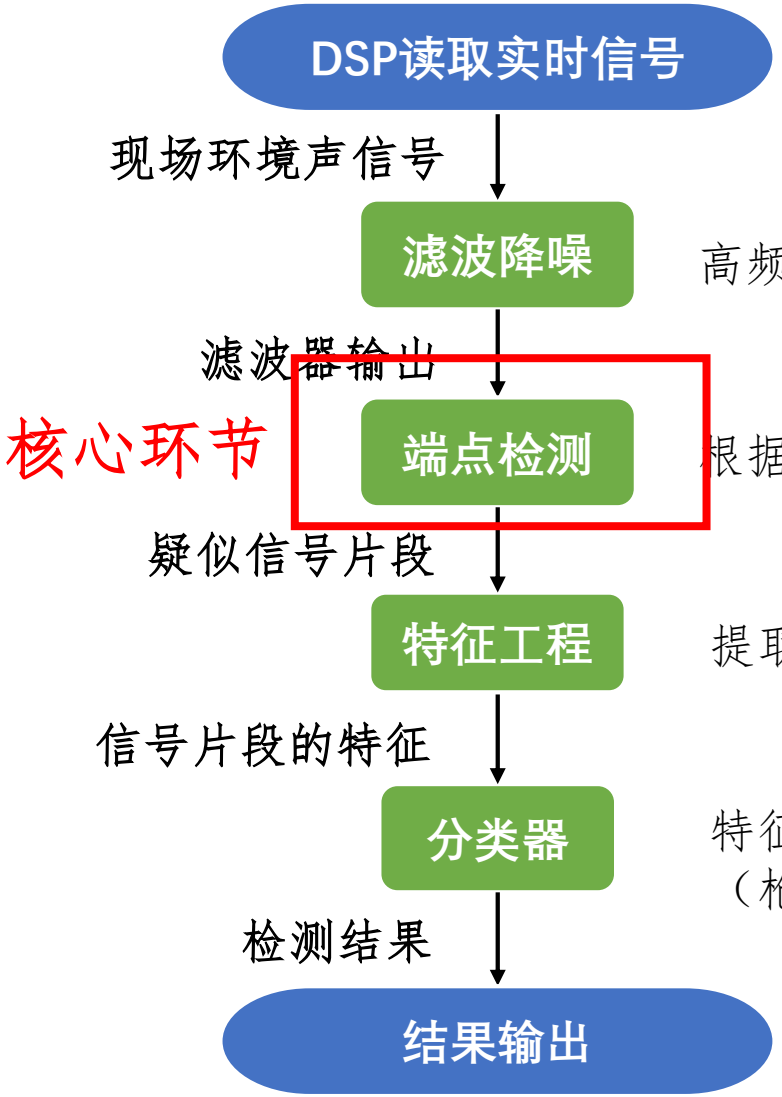


低通滤波和均值滤波后模式仍不明显

[1] 枪声信号分析与预处理，声学技术，蒋小为，张文等

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
端点检测 Endpoint Detection



高频段信号为噪声，用滤波器对信号进行初步清洗

根据疑似信号的特点，检测出疑似信号片段，并从声音序列中抽取出来

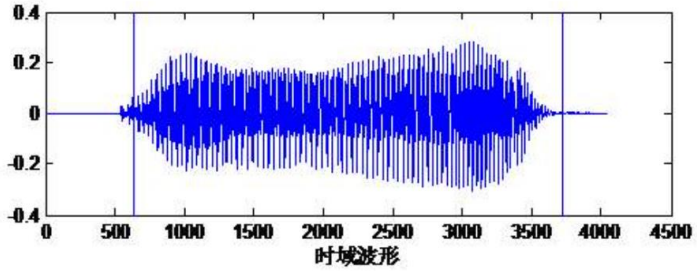
提取每个疑似信号片段的特征

特征作为分类器的输入，用事先训练好的分类器进行分类（枪声/爆炸声/车喇叭），从而实现目标信号的检测

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
端点检测 Endpoint Detection

- **端点检测(Endpoint Detection)**：从一段声信号中准确的找出声信号的起始点和结束点[3]



- 为什么**端点检测**很重要？

端点检测最早出现在语音信号处理的研究里，用于对语音片段进行精确分割，从而为后续的语音识别等语音信号处理做准备。

声信号识别、声学事件检测，跟语音识别有异曲同工的地方，语音识别将语音信号按语音片段进行分割，从而对每个片段分别做识别[8]；声学事件检测同样需要先把连续的声信号分割成一个个事件声片段，再进一步对每个片段进行检测[13]

语音识别：怎么找到人声的开始点和结束点？

声学事件检测：怎么找到声学事件（枪声/爆炸声/喇叭声）的开始点和结束点？

端点检测的准确率会直接关系到分类器的分类准确率（信噪比很低）

[3] 语音信号处理，机械工业出版社，赵力等
[8] 声学事件检测技术的发展历程与研究进展，Journal of Data Acquisition and Processing, 韩纪庆
[13] Events Detection For an Audio-based Surveillance System, IEEE ICME 2005, C. Clavel

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
端点检测 Endpoint Detection

- 常用方法[3]：
  操作最简单：**基于短时能量(short-time energy)、基于短时过零率(short-time ZCR)**
  其他方法：双门限法、自相关法、谱熵法、比例法、对数频谱距离法……
- **基于短时过零率ZCR(short-time Zero Crossing Rate)**[3]

定义语音信号 $x_n(m)$ 的短时过零率 Z_n 为

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

- **基于短时能量(short-time energy)**[3][4]

设第 n 帧语音信号 $x_n(m)$ 的短时能量用 E_n 表示，则其计算公式如下：

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

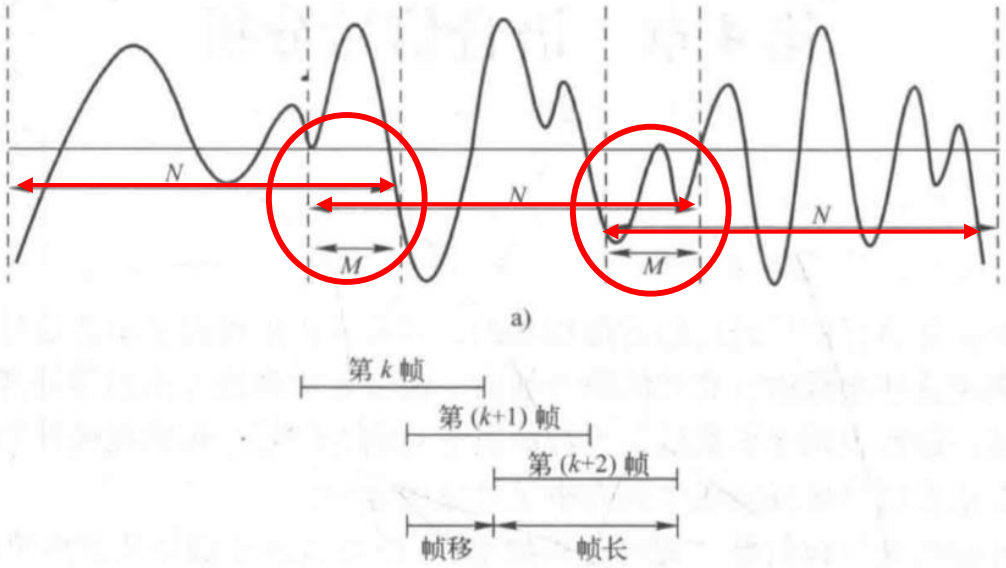
- 综合应用场景（枪声/爆炸/喇叭都是大功率信号）、算法复杂度（可高度并行化）、仿真结果等，采用**基于短时能量的端点检测**

[3] 语音信号处理，机械工业出版社，赵力等
[4] 基于短时能量和小波去噪的枪声信号检测方法，电测与仪表，张克刚等

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
端点检测 Endpoint Detection

- **分帧(frame)**：平稳信号处理方法不能应用于非平稳过程，但如果非平稳信号在一个短时间范围内，其特性基本保持不变，那么可以视作具有**短时平稳性**。**分帧**就是将非平稳信号碎片化为一个个近似平稳的短时信号的操作[13]
- 声学信号处理的许多运算和特征分析都是基于帧的！



设第 n 帧语音信号 $x_n(m)$ 的短时能量用 E_n 表示，则其计算公式如下：

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

- 为了保证帧的连续性，分帧往往会重叠，重叠部分利用**加窗(windowing)**弱化其影响

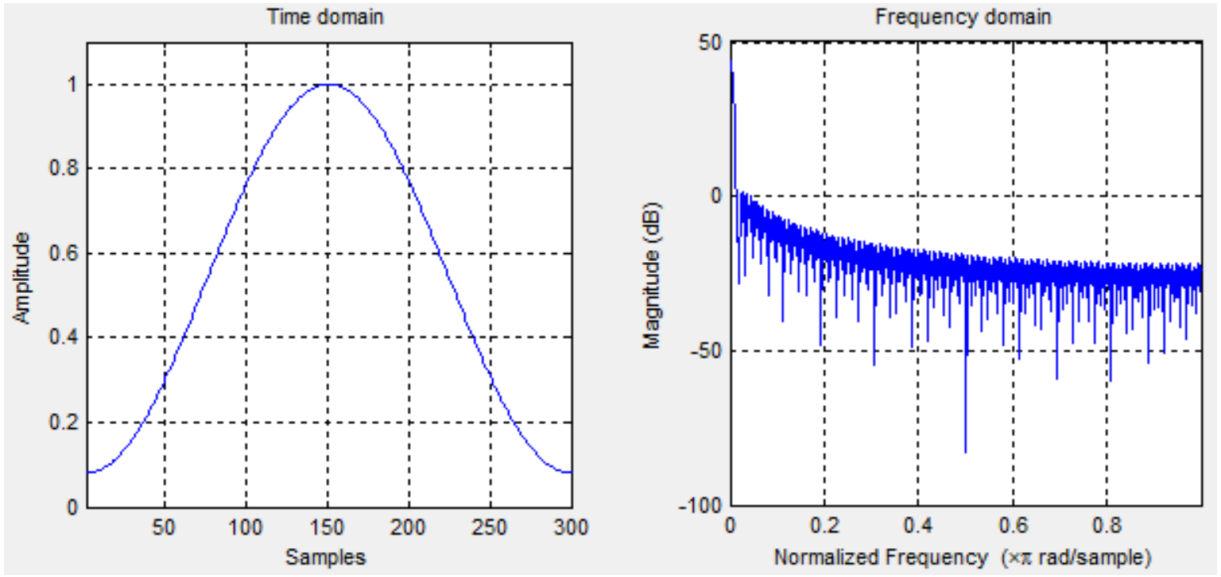
[13] Events Detection For an Audio-based Surveillance System, IEEE ICME 2005, C. Clavel

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

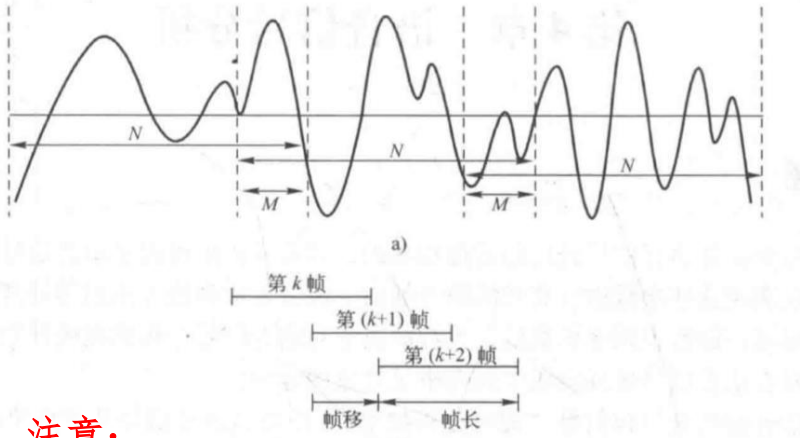
软件架构 Software Architecture
端点检测 Endpoint Detection

- **加窗(windowing)**：常用窗口有矩形窗、Hamming窗、汉宁窗等
- Hamming窗[3]：声学检测、语音处理等研究中非常常用

$$h(n)=\begin{cases}0.54-0.46\cos[2\pi n/(N-1)], & 0\leqslant n\leqslant N-1\\ 0, & n=\text{其他}\end{cases}$$



Hamming窗Matlab仿真



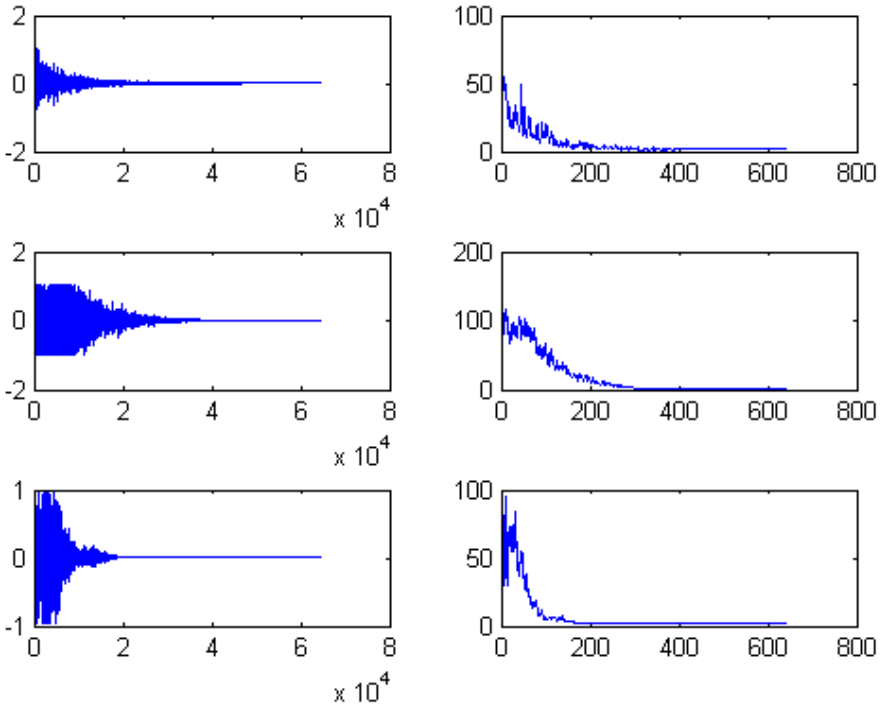
注意：
每一帧的长度相对于整个信号长度非常长非常短，以至信号在帧内近似于平稳信号（而不是像图中剧烈震荡）帧的信息可以认为是该信号的瞬时信息

[3] 语音信号处理，机械工业出版社，赵力等

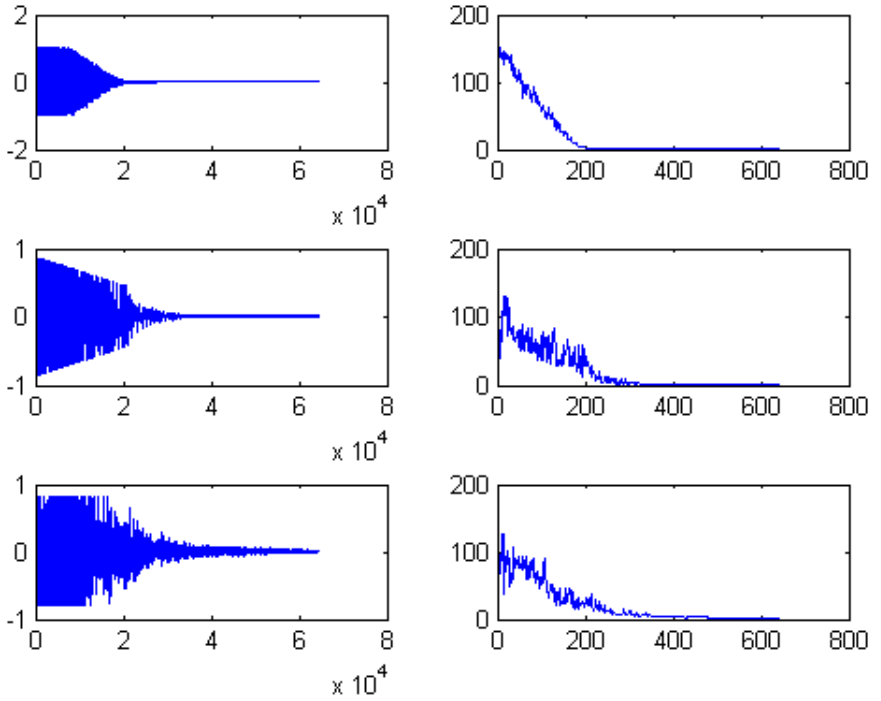
报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
端点检测 Endpoint Detection

- 取帧长 = 300pt，帧移 = 100pt [4]进行分帧、加窗、短时能量计算
  （声信号片段有效长度在30000-60000pt）



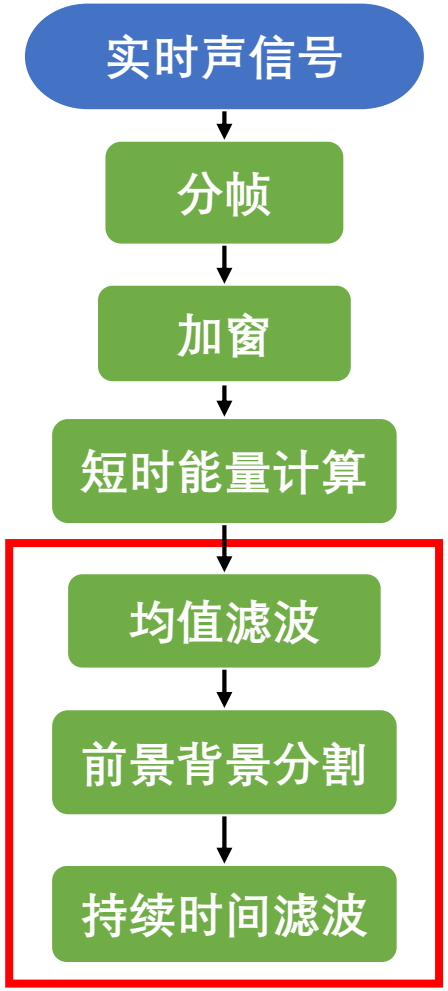
部分加噪的枪声信号(仿真)

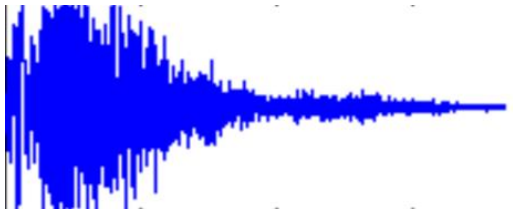


部分加噪的枪声信号(仿真)

[4] 基于短时能量和小波去噪的枪声信号检测方法，电测与仪表，张克刚等

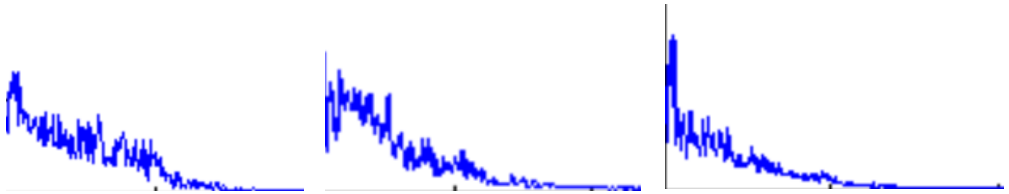
软件架构 Software Architecture
端点检测 Endpoint Detection

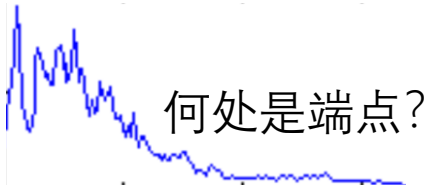




- **均值滤波(mean filtering)**

部分仿真结果中出现一定的高频抖动，考虑使用均值滤波做一个平滑。能够有效防止前景片段明明还没结束，但中间一两个点因抖动掉到阈值以下影响分离





何处是端点?

- 前景 VS 背景

将短时能量作为前景和背景的区分依据，使用**自适应的短时能量阈值**[4]，实现背景片段和可疑片段（前景）的分离

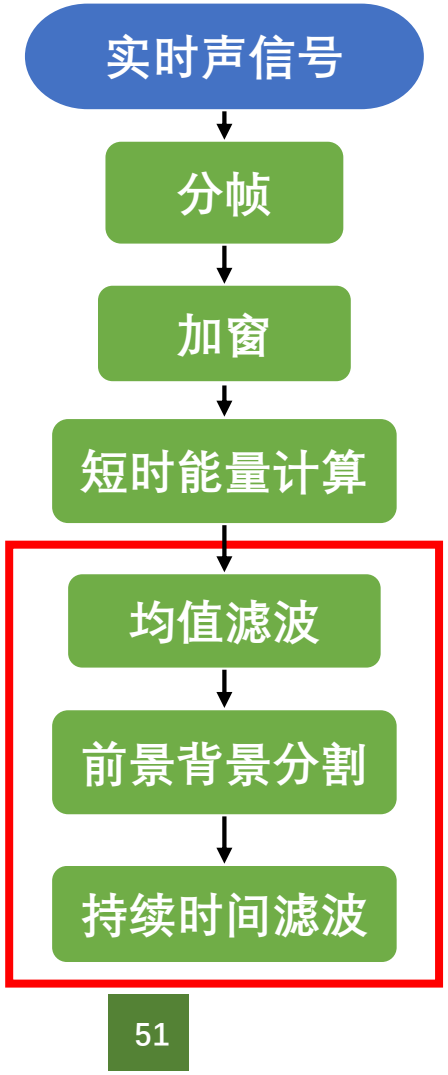
$$THr = \min(En) + 0.2[\max(En) - \min(En)]$$

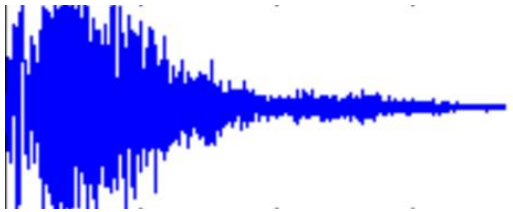
仿真结果发现系数取0.4准确率更高

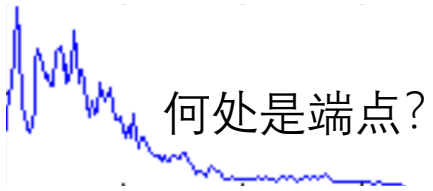
[4] 基于短时能量和小波去噪的枪声信号检测方法，电测与仪表，张克刚等

东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
端点检测 Endpoint Detection

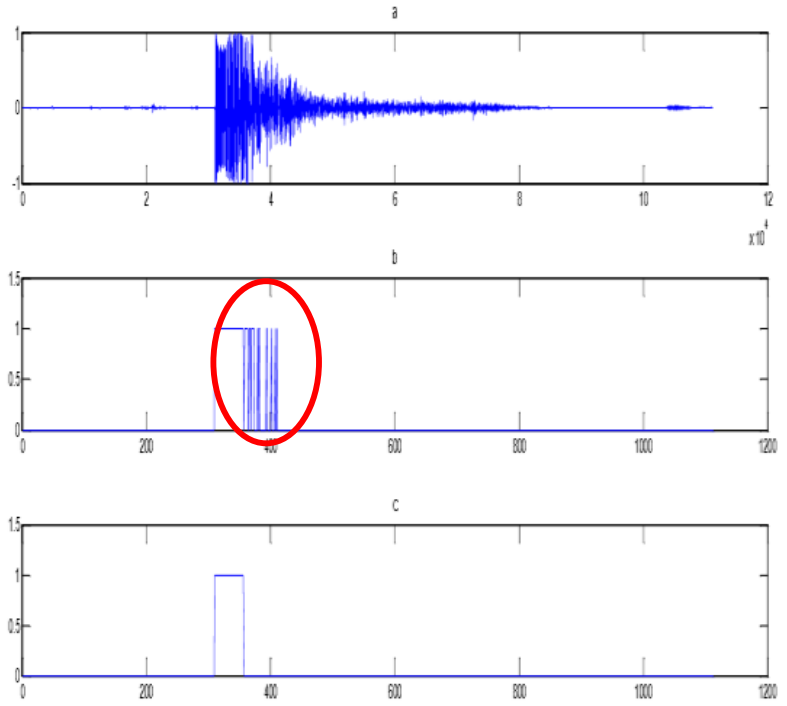






- **持续时间滤波(duration filtering)**

一个突发声信号通过前文的**短时能量自适应阈值分割**，从背景中分离出来后，常常伴随一系列**次要片段**（可以是回响、多径等原因引起）。次要片段高度碎片化，持续时间短，难以提取有效的特征进行检测，用持续时间作为阈值滤去。



[4] 基于短时能量和小波去噪的枪声信号检测方法，电测与仪表，张克刚等

软件架构 Software Architecture
端点检测 Endpoint Detection

实时声信号

分帧

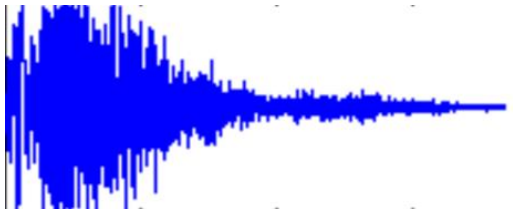
加窗

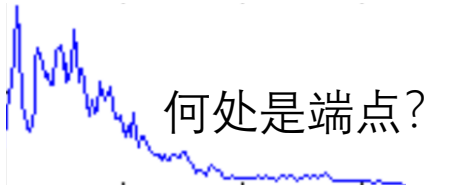
短时能量计算

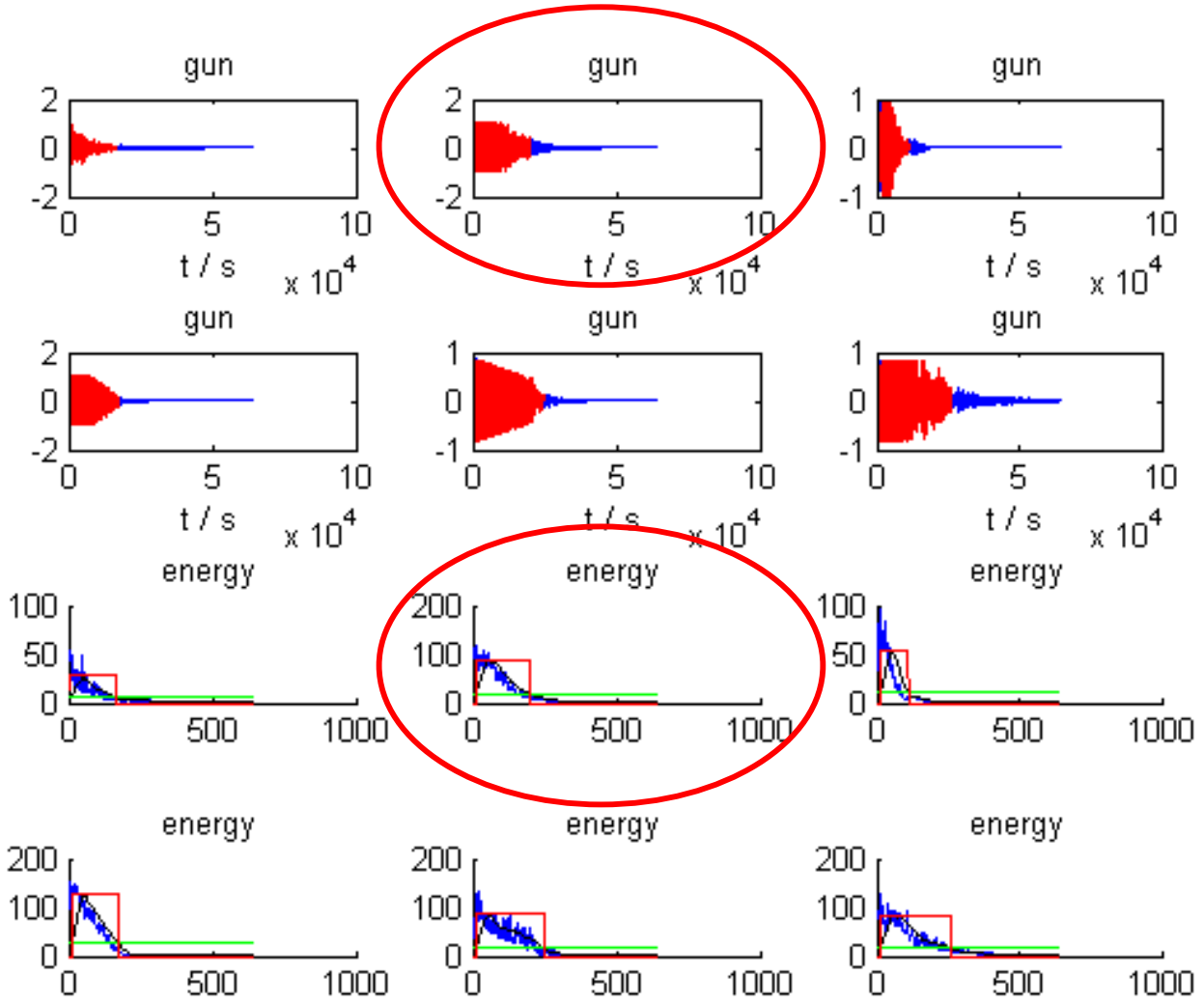
均值滤波

前景背景分割

持续时间滤波

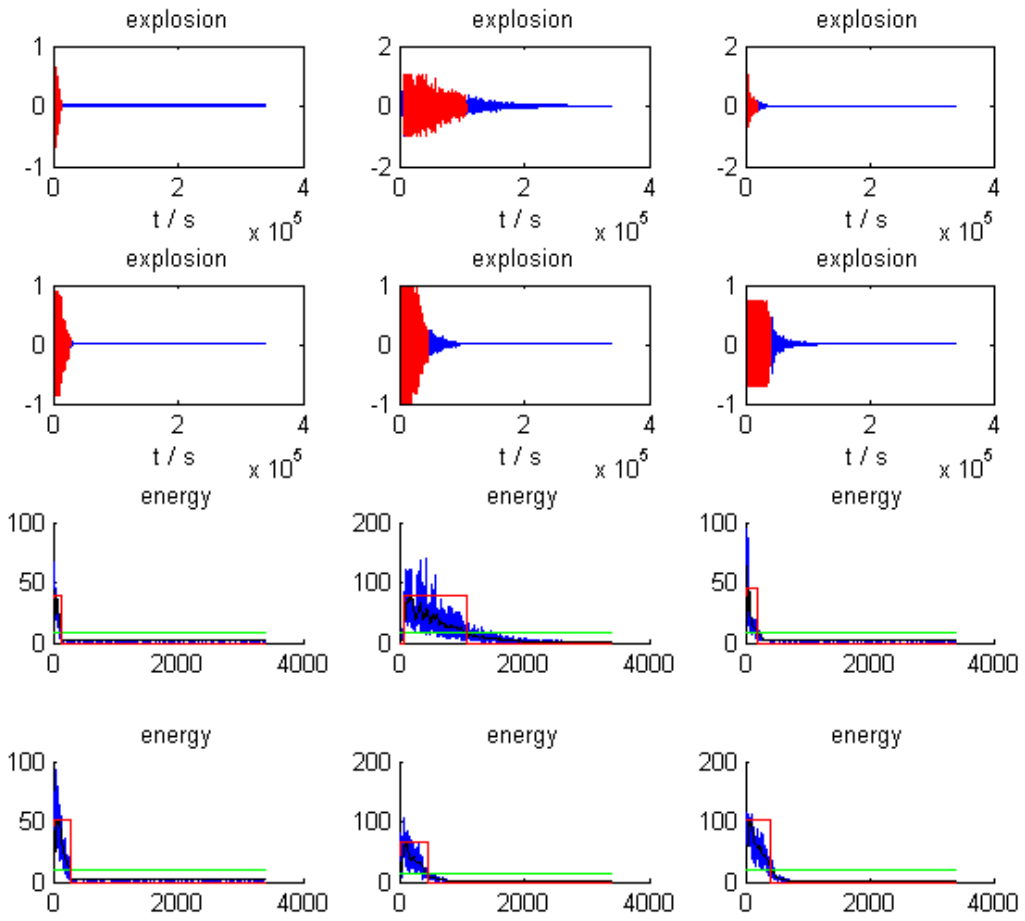




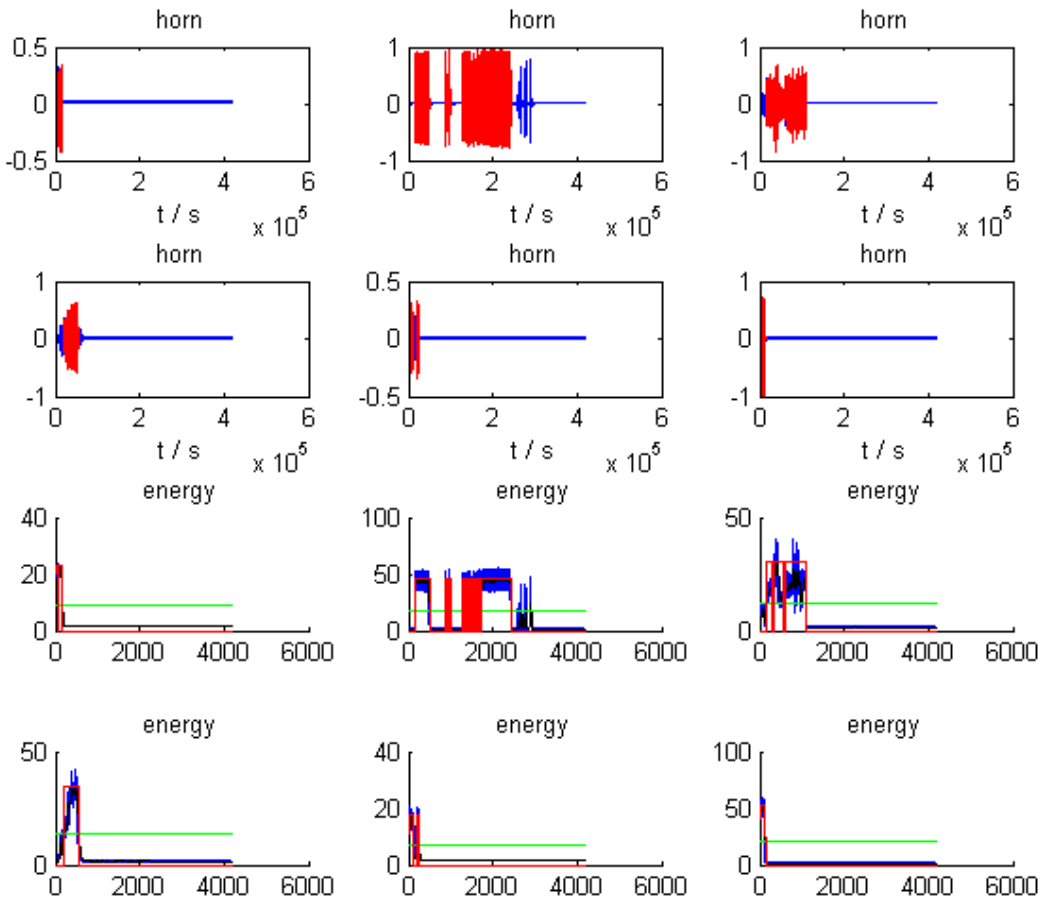


枪声端点检测(仿真)

软件架构 Software Architecture
端点检测 Endpoint Detection

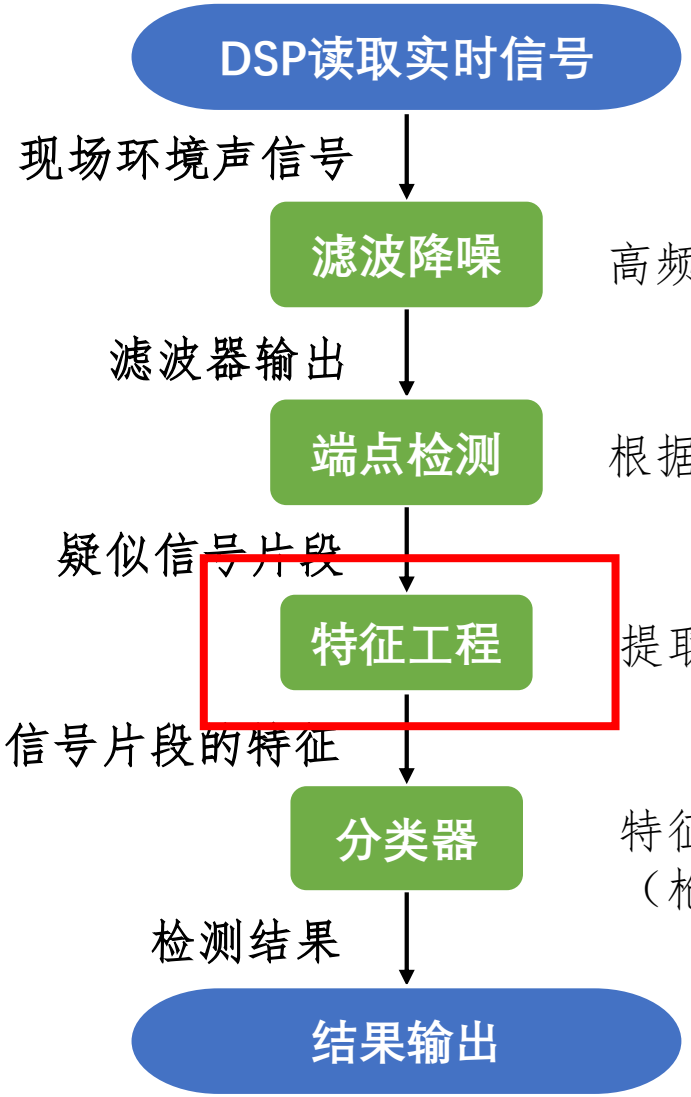


爆炸声端点检测(仿真)



汽车喇叭声端点检测(仿真)

软件架构 Software Architecture
特征工程 Feature Engineering



高频段信号为噪声，用滤波器对信号进行初步清洗

根据疑似信号的特点，检测出疑似信号片段，并从声音序列中抽取出来

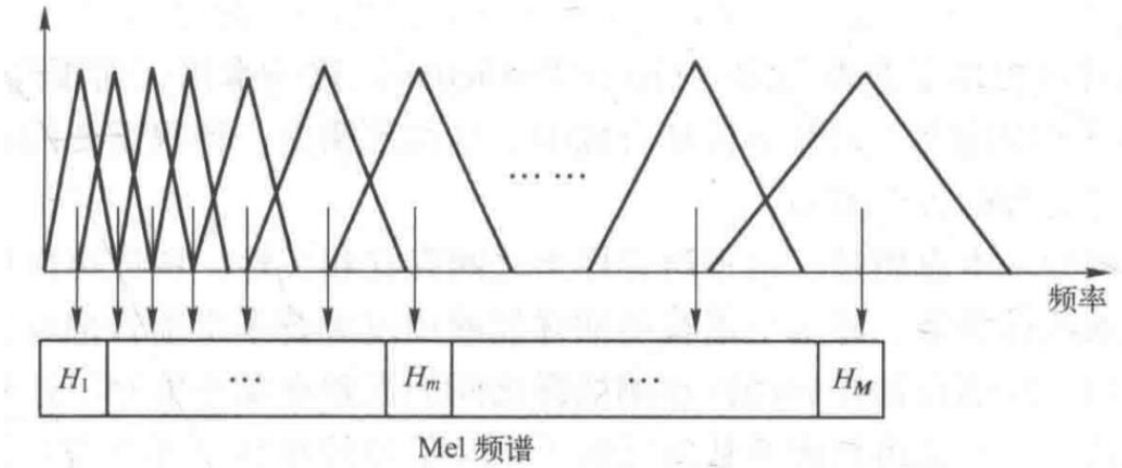
提取每个疑似信号片段的特征

特征作为分类器的输入，用事先训练好的分类器进行分类（枪声/爆炸声/车喇叭），从而实现目标信号的检测

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
特征工程 Feature Engineering

- 为什么需要特征？
  一个孤立、短促的枪声采样点高达6w+个 → 信号长度为6w+（单声道）。端点检测分离出主要片段后呢？仍有2w+
  必须要提取特征作为输入（维数大大减少），分类器的使用才存在可能！
  试图做一个2w+维度输入的分类器不可行、不现实（点之间的距离范数过大，不利于聚类）
- Mel频率倒谱系数MFCC(Mel-Frequency Cepstral Coefficient)[3]
  仿照人耳：设置一个滤波器组，1个输入信号，L个并行滤波器(Mel滤波器) → L个滤波器输出
  用L个输出构造L维向量作为声音片段的特征



Mel 频谱

[3] 语音信号处理，机械工业出版社，赵力等

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture

特征工程 Feature Engineering

疑似信号片段

↓

分帧加窗

↓

Mel频率转换

$$\mathrm{Mel}(f) = 2595 \lg(1 + f/700)$$

↓

滤波器组滤波

$$m(l) = \sum_{k=o(l)}^{h(l)} W_l(k) \mid X_n(k) \mid \qquad l = 1, 2, \cdots, L$$

$$W_l(k) = \begin{cases} \dfrac{k - o(l)}{c(l) - o(l)} & o(l) \leqslant k \leqslant c(l) \\ \dfrac{h(l) - k}{h(l) - c(l)} & c(l) \leqslant k \leqslant h(l) \end{cases}$$

↓

对数处理

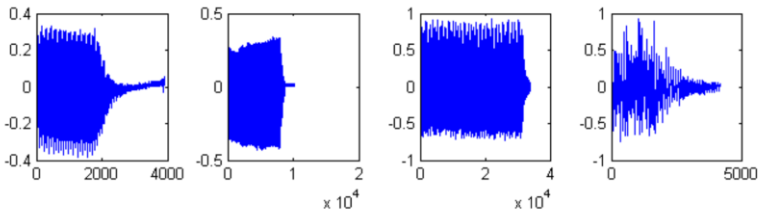
↓

余弦变换

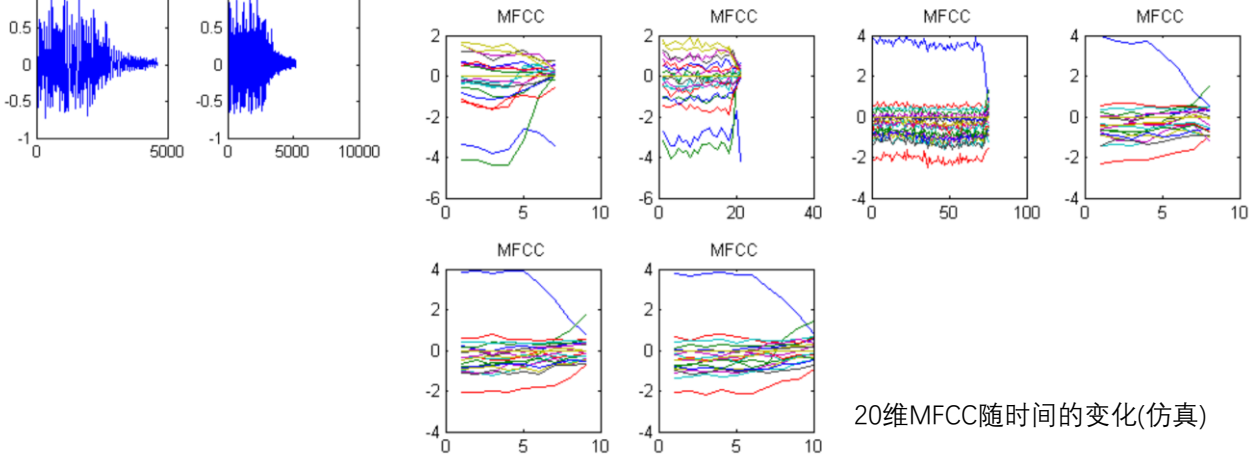
$$c_{\mathrm{mfcc}}(i) = \sqrt{\frac{2}{N}} \sum_{l=1}^{L} \lg m(l) \cos\left\{\left(l - \frac{1}{2}\right)\frac{i\pi}{L}\right\}$$

↓

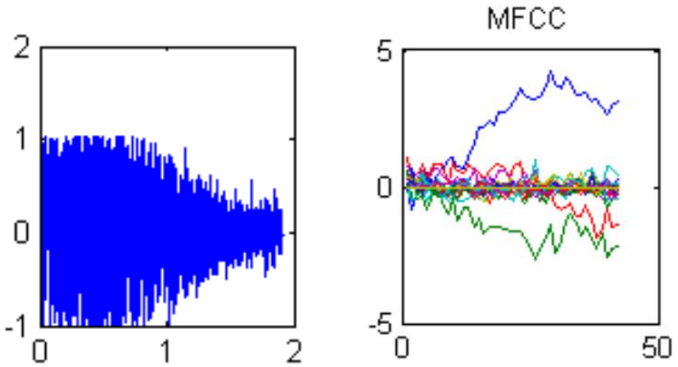
MFCC特征(向量)



部分汽车喇叭声端点检测结果(仿真)



20维MFCC随时间的变化(仿真)

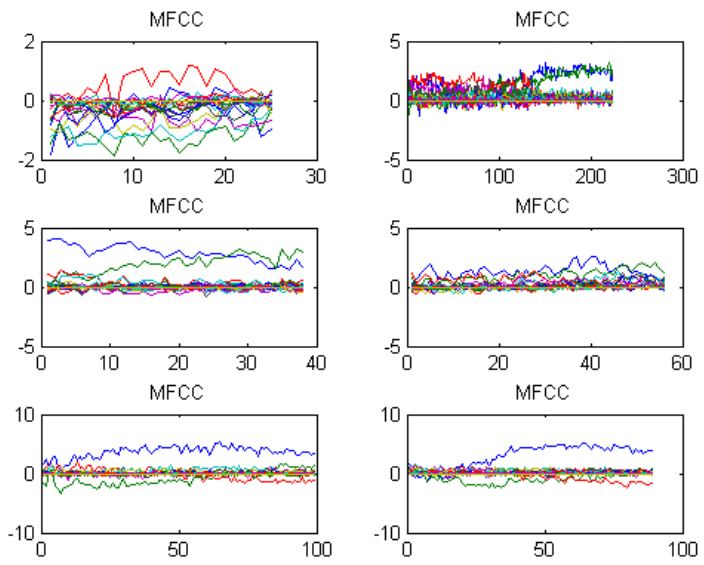


枪声MFCC分析(仿真)

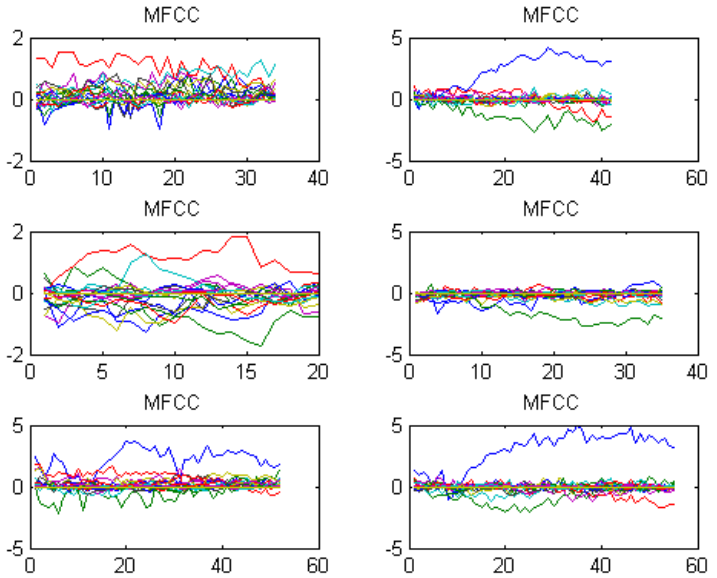
报告人：
招梓枫，林涵

东南大学仪器科学与工程学院
Instrument Science and Engineering

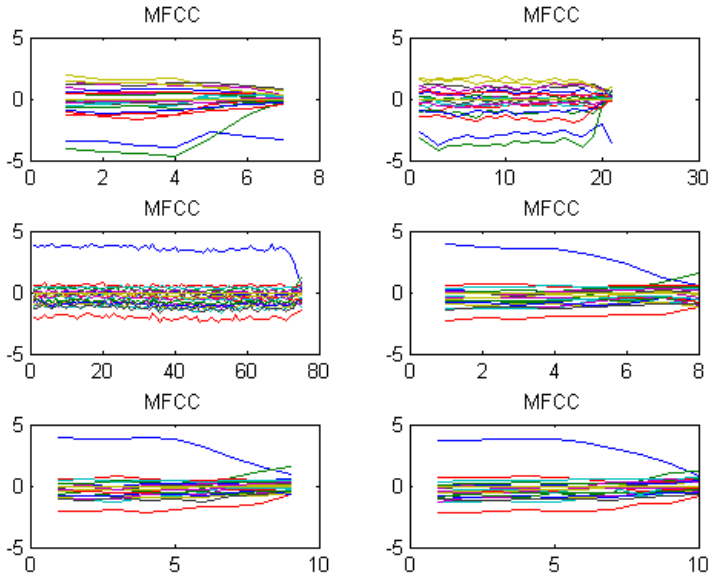
软件架构 Software Architecture
特征工程 Feature Engineering



枪声MFCC分析(仿真)



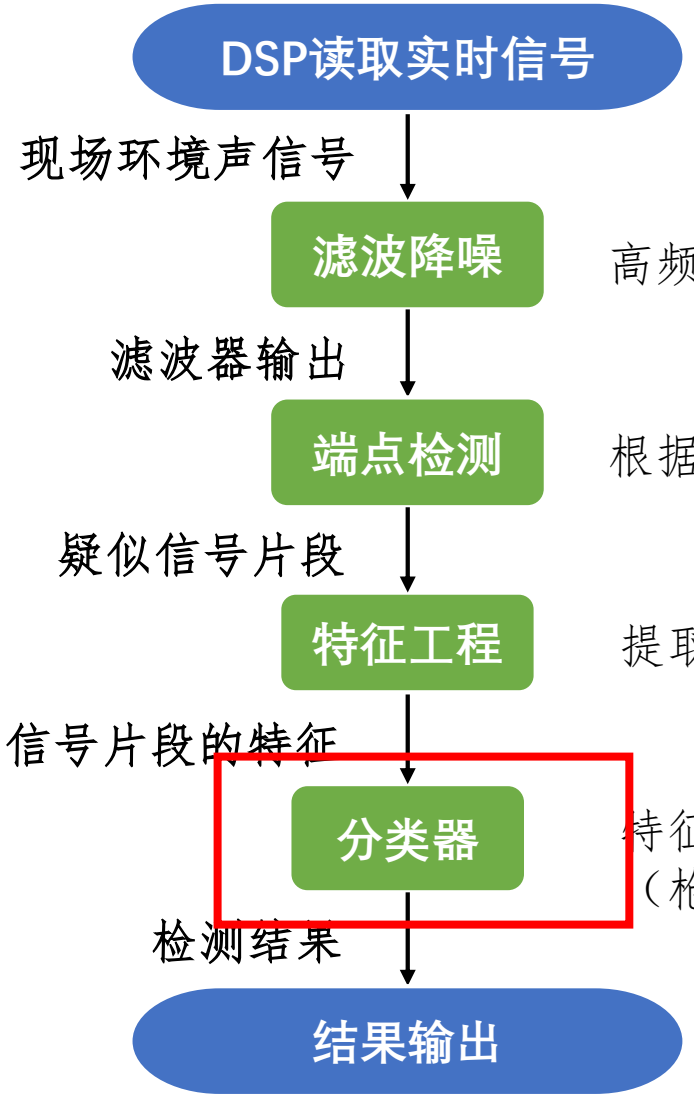
爆炸声MFCC分析(仿真)



喇叭声MFCC分析(仿真)

东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
分类器 Classifier



高频段信号为噪声，用滤波器对信号进行初步清洗

根据疑似信号的特点，检测出疑似信号片段，并从声音序列中抽取出来

提取每个疑似信号片段的特征

特征作为分类器的输入，用事先训练好的分类器进行分类（枪声/爆炸声/车喇叭），从而实现目标信号的检测

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
分类器 Classifier

- 关于**分类器(classifier)**

时下非常非常非常火爆的研究热点，机器学习(Machine Learning)中的一大研究内容，经典的分类器利用概率统计、统计信号处理、贝叶斯估计等理论，在向量空间中，将特征化的输入进行划分，分类器的一些经典模型[5]：

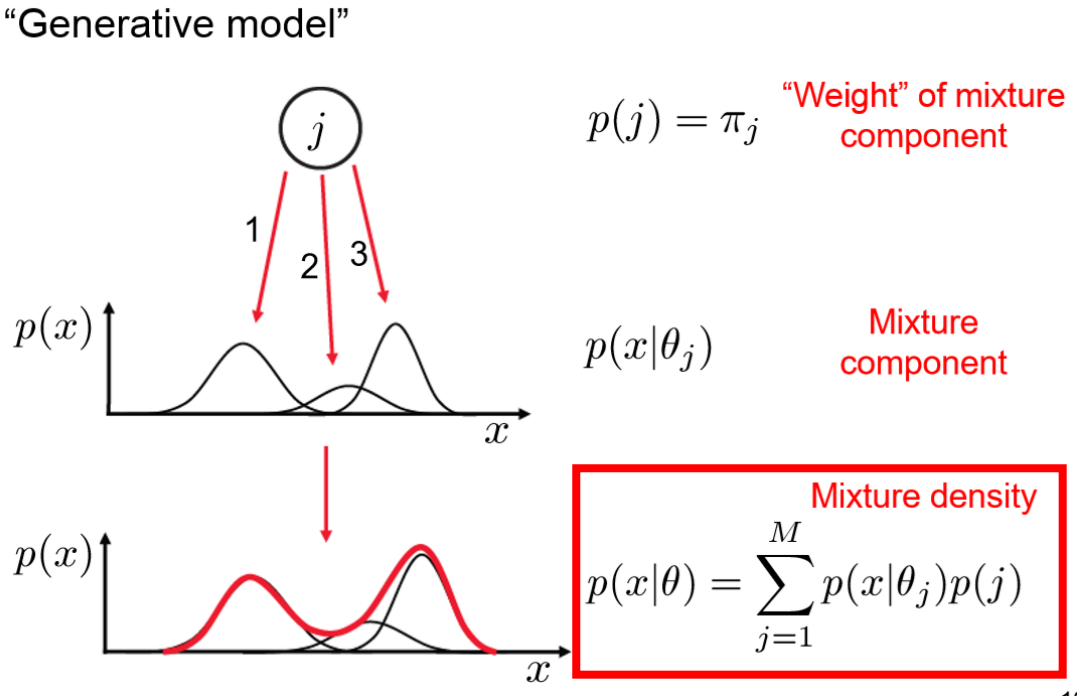
- Bayes决策：需要posterior或者prior & likelihood，需要loss matrix
- 支持向量机SVM(Support Vector Machine)：根据线性可分性分为linear SVM和nonlinear SVM nonlinear SVM中kernel function的选用比较考究[6][7]
- Adaboost(Adaptive Boosting)：sensitive to outliers，手头的样本太少
- 随机森林(Random Forest)：训练复杂，内存消耗大
- GMM + Maximum Likelihood Estimation：本项目中使用
  1. 模型训练好后，易于在检测系统终端部署（嵌入式微机）并完成一站式检测
  2. cluster数目是唯一超参数，且能通过WSS确定。避免了很多模型设计问题（线性非线性、核函数、冗余量）

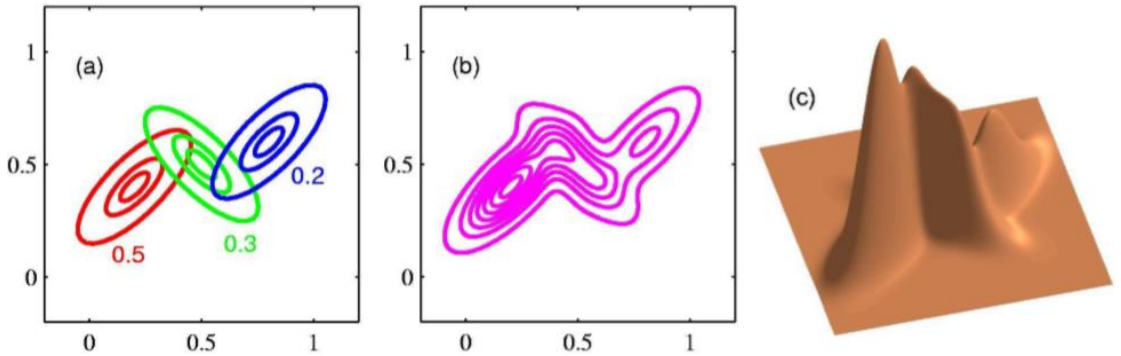
[5] Pattern Recognition and Machine Learning, Springer, Christopher M. Bishop
[6] Learning with Kernels, MIT Press, B. Schoelkopf, A. Smola
[7] A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery
[14] Automatic Sound Detection and Recognition for Noisy Environment, IEEE European Signal Processing Conference, Alain Dufaux

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 Software Architecture
分类器 Classifier

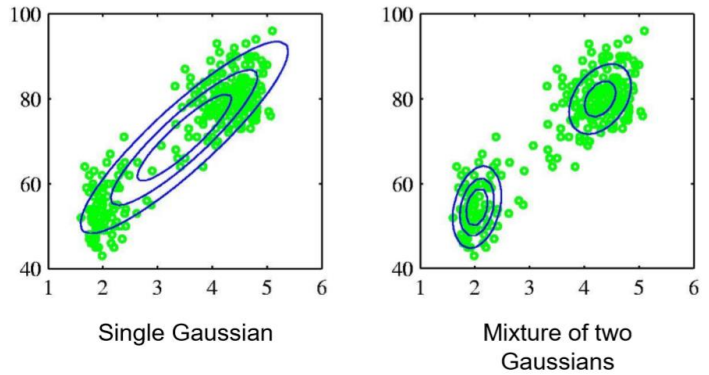
- 混合高斯模型GMM(Gaussian Mixture Model)[5][10]：又叫MoG(Mixture of Gaussian)





A single parametric distribution is often not sufficient

- E.g. for multimodal data



[5] Pattern Recognition and Machine Learning, Springer, Christopher M. Bishop
[9] 公共场所下的枪声检测研究，硕士学位论文，朱强强
[10] Machine Learning (Lecture Notes), RWTH University Aachen, Bastian Leibe, Bernt Schiele

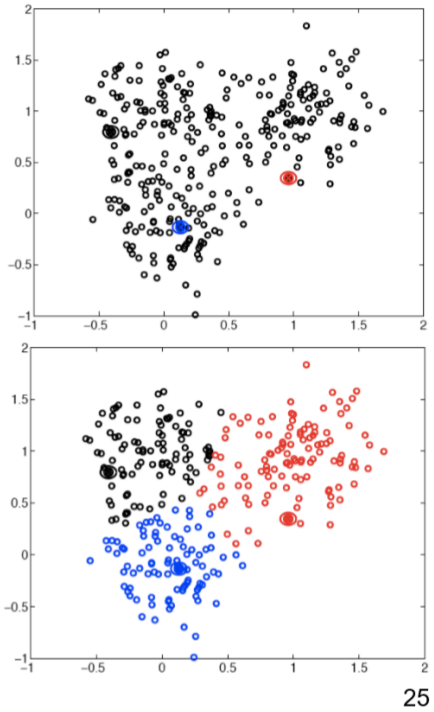
报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

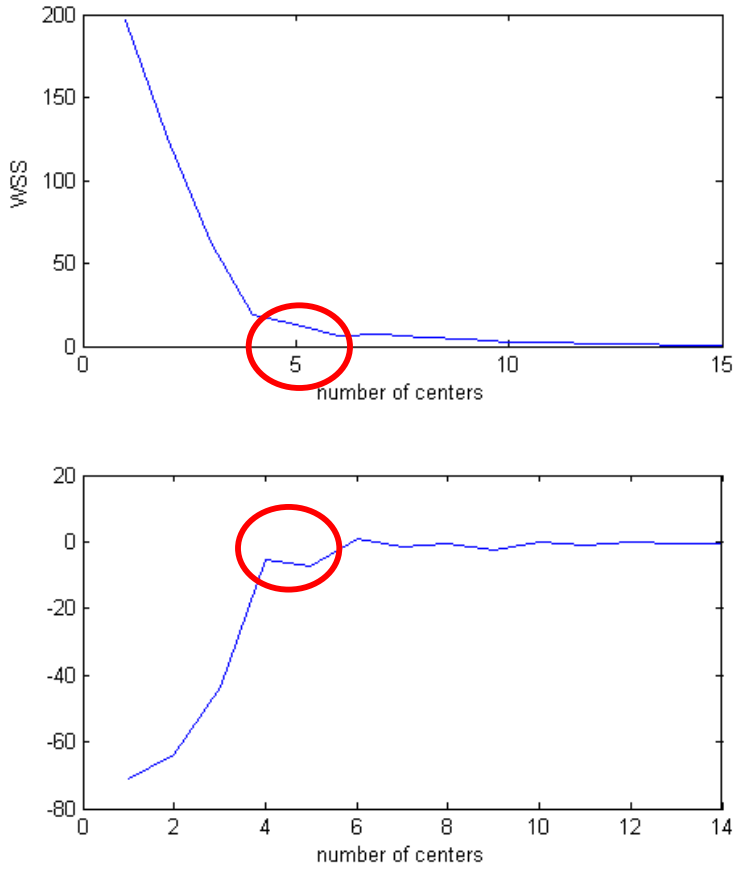
软件架构 Software Architecture
分类器 Classifier

- K-means用于聚类的初始化，同时计算WSS。聚类中心数目在5左右时，WSS随中心的增多不再有明显减小，选定5作为聚类中心数目

K-Means Clustering

- Iterative procedure
  1. Initialization: pick K arbitrary centroids (cluster means)
  2. Assign each sample to the closest centroid.
  3. Adjust the centroids to be the means of the samples assigned to them.
  4. Go to step 2 (until no change)
- Algorithm is guaranteed to converge after finite #iterations.
  - Local optimum
  - Final result depends on initialization.

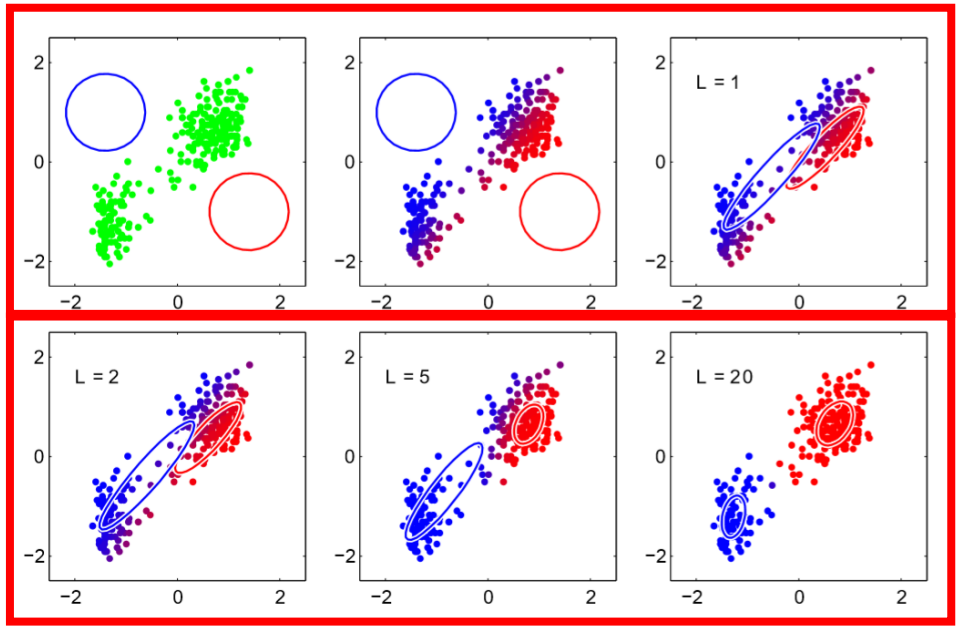




[5] Pattern Recognition and Machine Learning, Springer, Christopher M. Bishop
[10] Machine Learning (Lecture Notes), RWTH University Aachen, Bastian Leibe, Bernt Schiele
[16] A Meta-learning Approach for Recommending the Number of Clusters for Clustering Algorithms, Knowledge Based System, Bruno Almeida Pimentel

软件架构 – 分类器
Software Architecture – classifier

Step 1: Initialization (K-Means)



Step 2
EM Algorithm

Expectation-Maximization (EM) Algorithm

- E-Step: softly assign samples to mixture components

$$\gamma_j(\mathbf{x}_n) \leftarrow \frac{\pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \qquad \forall j = 1, \ldots, K, \quad n = 1, \ldots, N$$

- M-Step: re-estimate the parameters (separately for each mixture component) based on the soft assignments

$$\hat{N}_j \leftarrow \sum_{n=1}^{N} \gamma_j(\mathbf{x}_n) = \text{soft number of samples labeled } j$$

$$\hat{\pi}_j^{\text{new}} \leftarrow \frac{\hat{N}_j}{N}$$

$$\hat{\boldsymbol{\mu}}_j^{\text{new}} \leftarrow \frac{1}{\hat{N}_j} \sum_{n=1}^{N} \gamma_j(\mathbf{x}_n)\mathbf{x}_n$$

$$\hat{\boldsymbol{\Sigma}}_j^{\text{new}} \leftarrow \frac{1}{\hat{N}_j} \sum_{n=1}^{N} \gamma_j(\mathbf{x}_n)(\mathbf{x}_n - \hat{\boldsymbol{\mu}}_j^{\text{new}})(\mathbf{x}_n - \hat{\boldsymbol{\mu}}_j^{\text{new}})^{\mathrm{T}}$$

38

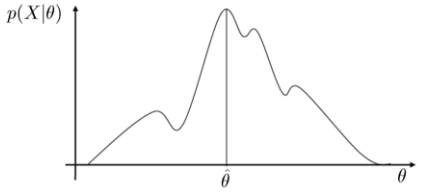
[5] Pattern Recognition and Machine Learning, Springer, Christopher M. Bishop
[10] Machine Learning (Lecture Notes), RWTH University Aachen, Bastian Leibe, Bernt Schiele

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构 – 分类器
Software Architecture – classifier

- 极大似然估计ML(Maximum Likelihood Estimation)[5][9][10][11]:对3种声学事件（枪声/爆炸/汽车喇叭）分别训练GMM，得到三个GMM模型。将待检测结果的MFCC特征**x**分别输入3个GMM，得到3个概率密度，概率密度最大者即认为是该类别

We want to obtain $\hat{\theta}$ such that $L(\hat{\theta})$ is maximized.



$$p(x|\theta) = \sum_{j=1}^{M} p(x|\theta_j)p(j)$$

测试集	枪声似然	爆炸似然	喇叭似然	分类
gun	24.1%	0.1%	0%	gun
gun	25.1%	0.3%	0%	gun
gun	97.3%	0%	0%	gun
gun	75.7%	0%	0%	gun
gun	27.5%	0.2%	0%	gun
gun	22.4%	0%	0%	gun

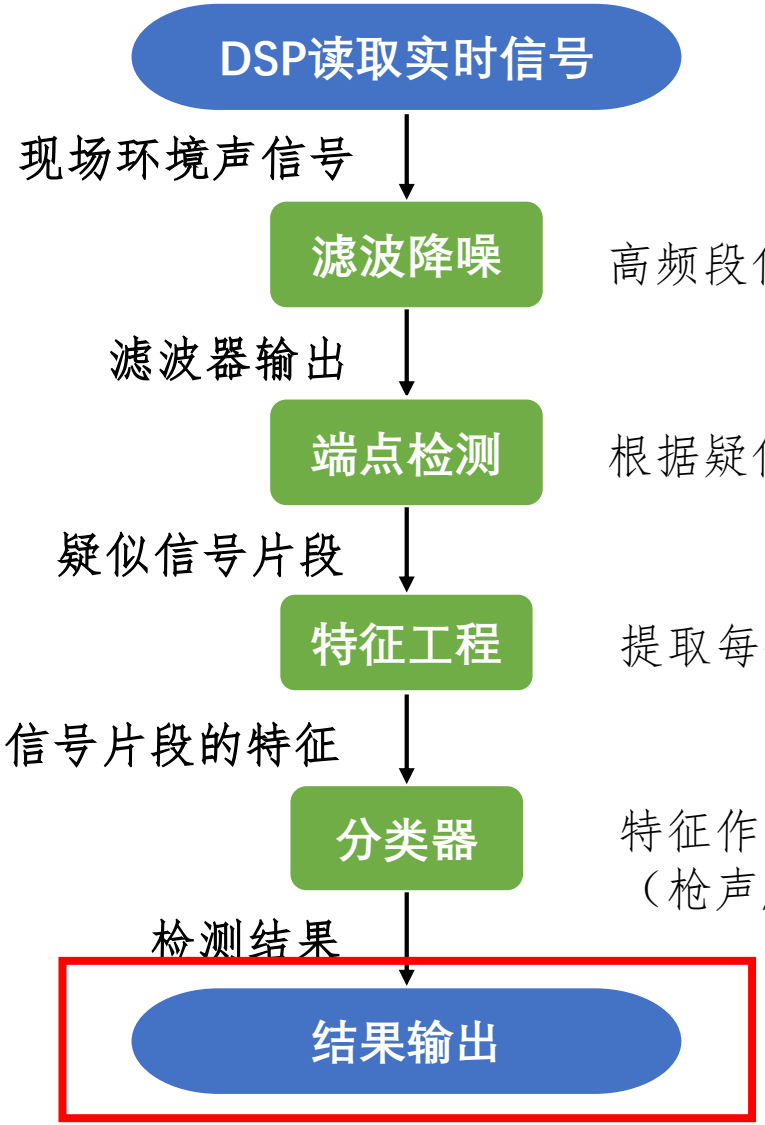
测试集	枪声似然	爆炸似然	喇叭似然	分类
explosion	0.4%	54.6%	0%	explosion
explosion	0%	52.1%	0%	explosion
explosion	0%	93.4%	0%	explosion
explosion	2.6%	24.3%	0%	explosion
explosion	1%	11.9%	0%	explosion
explosion	1.8%	77.8%	0%	explosion

测试集	枪声似然	爆炸似然	喇叭似然	分类
horn	0%	0%	14.3%	horn
horn	0%	0%	23.3%	horn
horn	0%	0%	17.6%	horn
horn	0%	0%	14.4%	horn
horn	0%	0%	13.8%	horn
horn	0%	0%	47.1%	horn

[5] Pattern Recognition and Machine Learning, Springer, Christopher M. Bishop
[9] 公共场所下的枪声检测研究，硕士学位论文，朱强强
[10] Machine Learning (Lecture Notes), RWTH University Aachen, Bastian Leibe, Bernt Schiele
[11] 统计信号处理讲义，东南大学，蒋忠进，孟桥等

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

软件架构
Software Architecture



高频段信号为噪声，用滤波器对信号进行初步清洗

根据疑似信号的特点，检测出疑似信号片段，并从声音序列中抽取出来

提取每个疑似信号片段的特征

特征作为分类器的输入，用事先训练好的分类器进行分类（枪声/爆炸声/车喇叭），从而实现目标信号的检测

报告人：
招梓枫，林涵
东南大学仪器科学与工程学院
Instrument Science and Engineering

展望
Future Work

- 扩展研究对象，支持更多的突发事件声学检测和分类（例如火灾、倒塌、人群恐慌），从而进一步提高检测系统的抗干扰性能和检测范围
- 扩大数据集，做进一步的数据增强(data augmentation)，提高前端信号处理模块性能和分类器的准确率
- 增加声源定位(source positioning)模块，实现突发事件检测后的粗定位，从而为监控摄像头动态对准、相应的安保行动提供信息
- 探索基于低功耗传感器网络节点的枪声检测[15]与定位系统

[15] Evaluation of Gunshot Detection Algorithms, IEEE Transactions on Circuits and Systems, Alfonso Chacón-Rodríguez

感谢聆听

招梓枫　林涵

东南大学仪器科学与工程学院
Instrument Science and Engineering