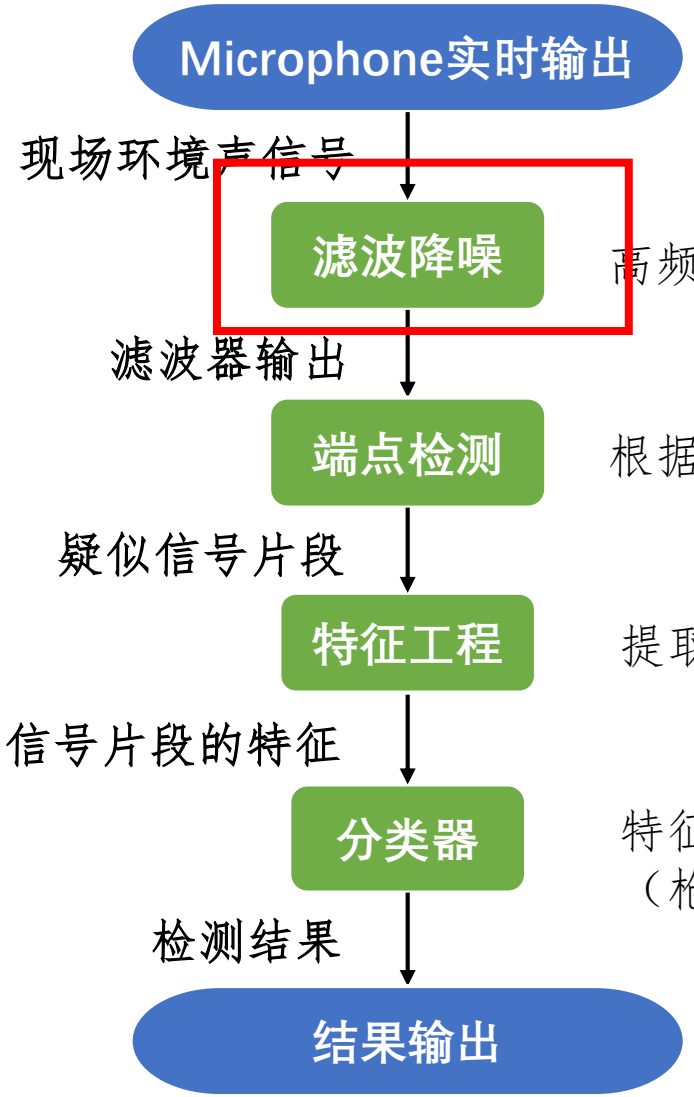
软件架构
Software Architecture



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

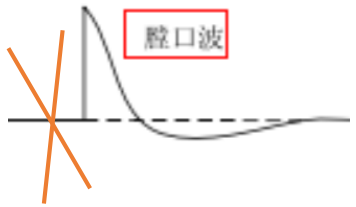
软件架构 – 滤波降噪
Software Architecture - filtering & denoising

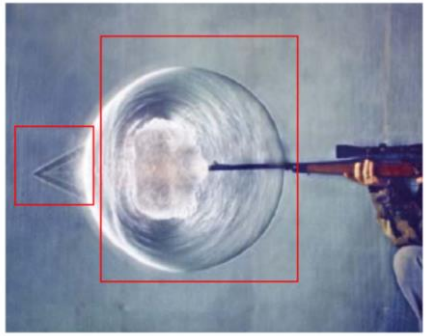
- 为什么要滤波降噪？

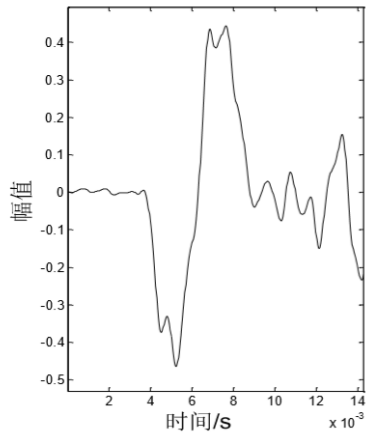
先来听一段典型的枪声信号

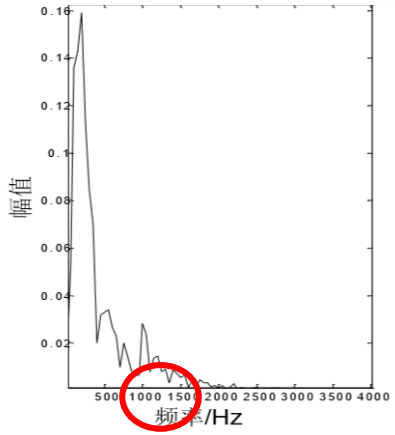
仅考虑**膛口波(muzzle blast)**，典型的枪声信号是一个**负压-正压**的过程

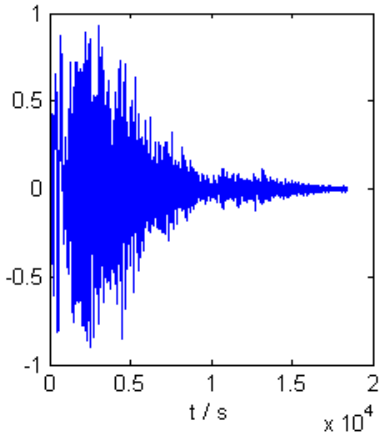
理论波形的的频率集中在**低频**【1】【2】，若要在检测的基础上做进精确定位可以综合**膛口波**与**马赫波(shock wave)**做分析【12】

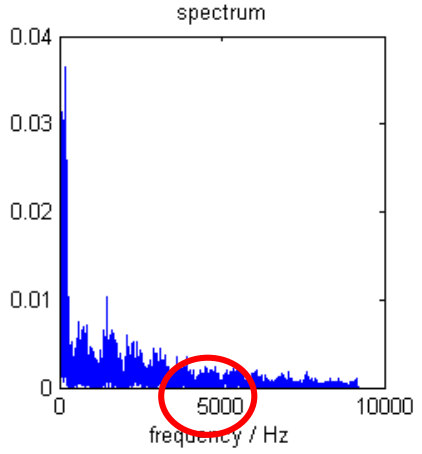












【1】枪声信号分析与预处理，声学技术，蒋小为，张文等

【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 – 滤波降噪
Software Architecture - filtering & denoising

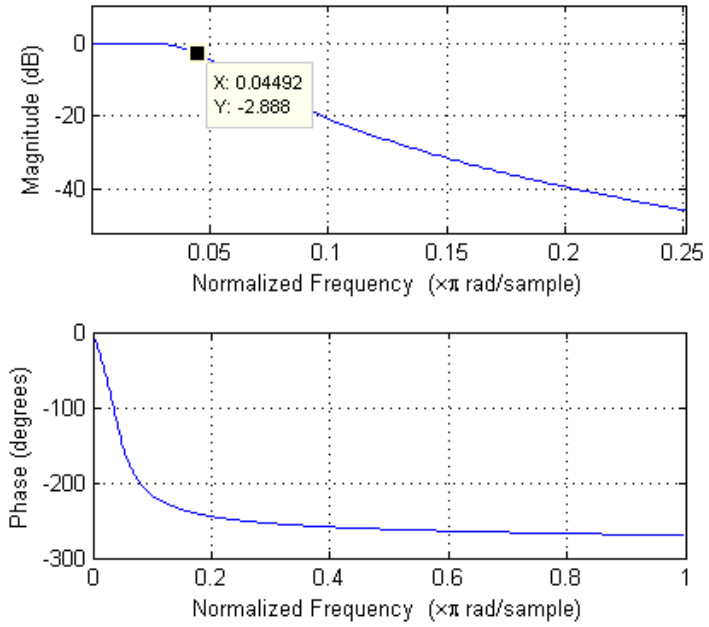
- 滤波降噪方案

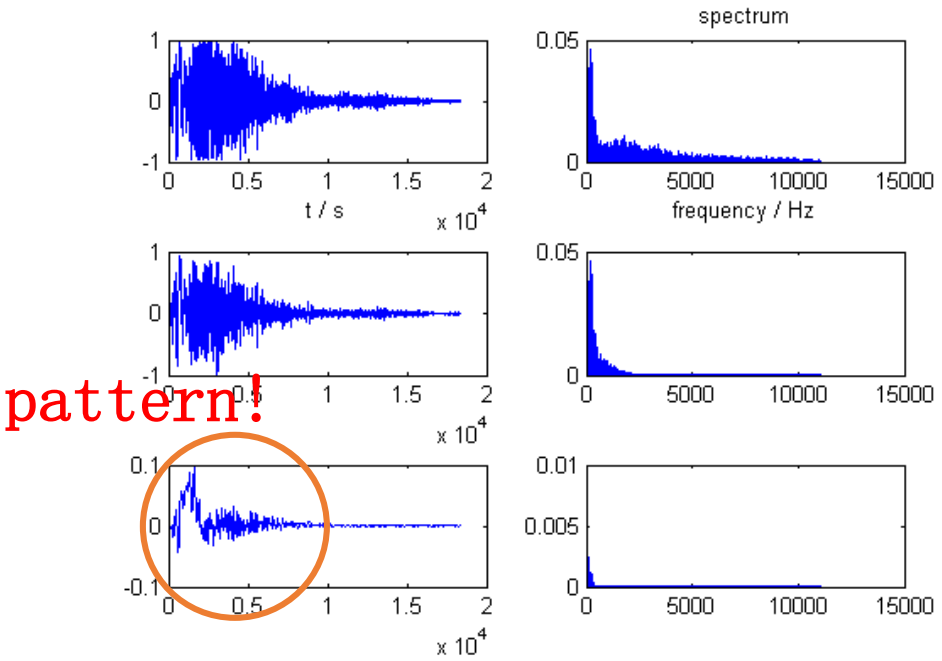
Butterworth Filter实现**低通滤波**(cutoff frequency = 1kHz)

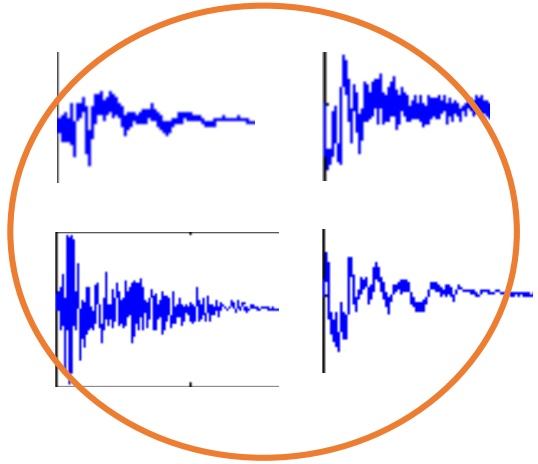
- 考虑使用更好的滤波方案？直接把枪声波形过滤出来后进行**相关分析(correlation)**？

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：这势必无法解决其他低频信号的干扰，比如汽车喇叭！







【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

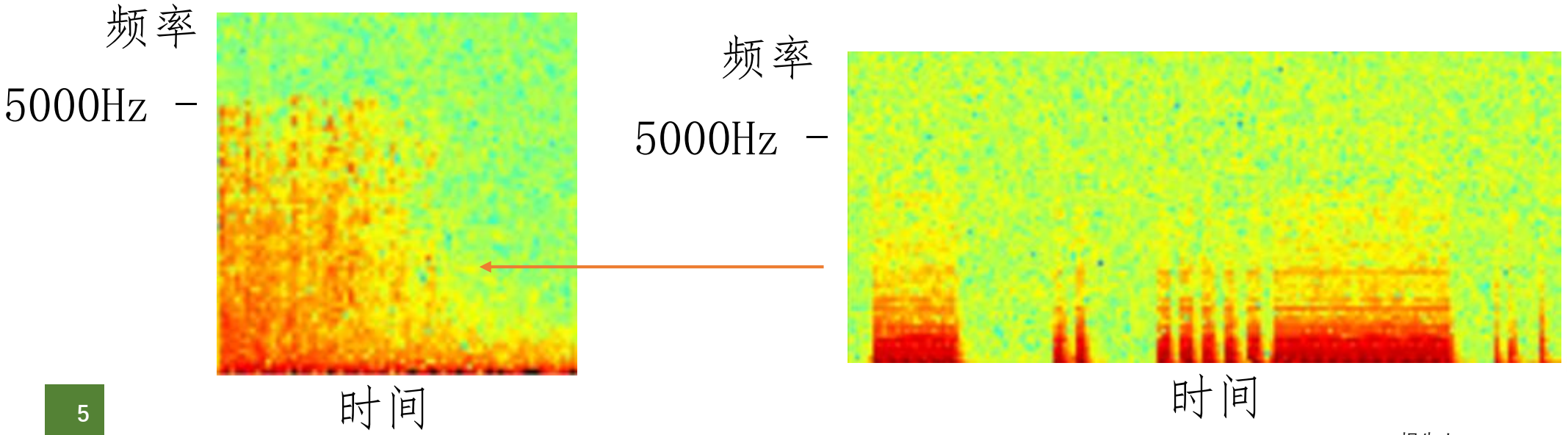
软件架构 – 滤波降噪
Software Architecture - filtering & denoising

- 滤波降噪方案

  Butterworth Filter实现**低通滤波**(cutoff frequency = 1kHz)
- 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：（均值滤波）势必无法解决其他低频信号的干扰，比如汽车喇叭！



【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

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 – 端点检测
Software Architecture – endpoint detection

- 端点检测(Endpoint Detection)：从一端语音信号中准确的找出语音信号的起始点和结束点【3】
- 为什么要**端点检测**？

端点检测最早出现在语音信号处理的研究里，用于对话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，或者说**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；声学事件检测同样需要先把可疑的声学信号片段分割出来，然后再进一步对每个可疑片段进行检测【8】

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？

【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

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 – 端点检测
Software Architecture – 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(short-time ZCR)**

定义语音信号 $x_n(m)$ 的短时过零率 Z_n 为

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

- **基于短时能量(short-time energy)【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：

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

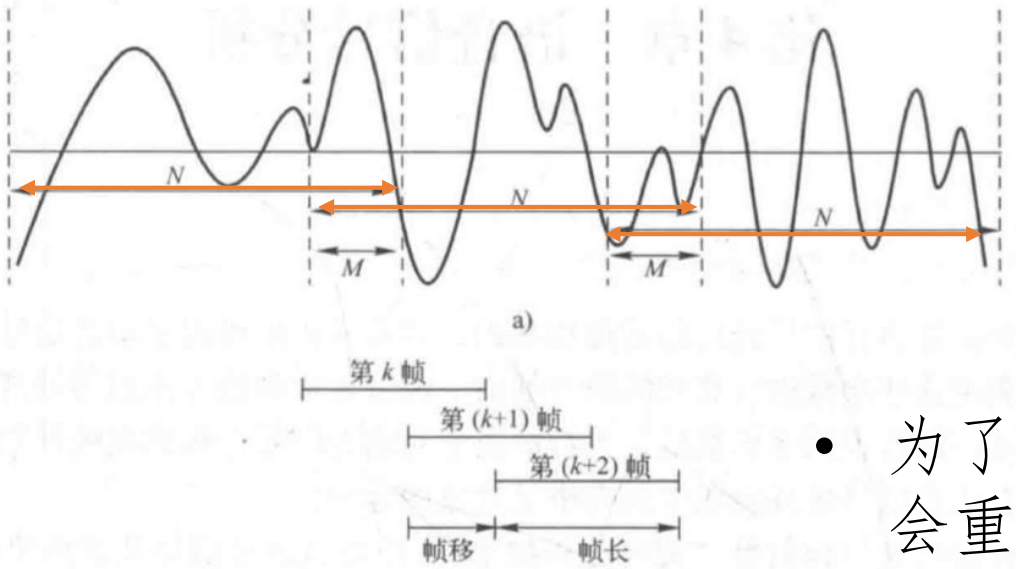
- 综合应用场景（枪声/爆炸/喇叭都是大功率信号）、算法复杂度（可高度并行化）、仿真结果等，采用**基于短时能量的端点检测**

【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 – 端点检测
Software Architecture – 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。
- 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：

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

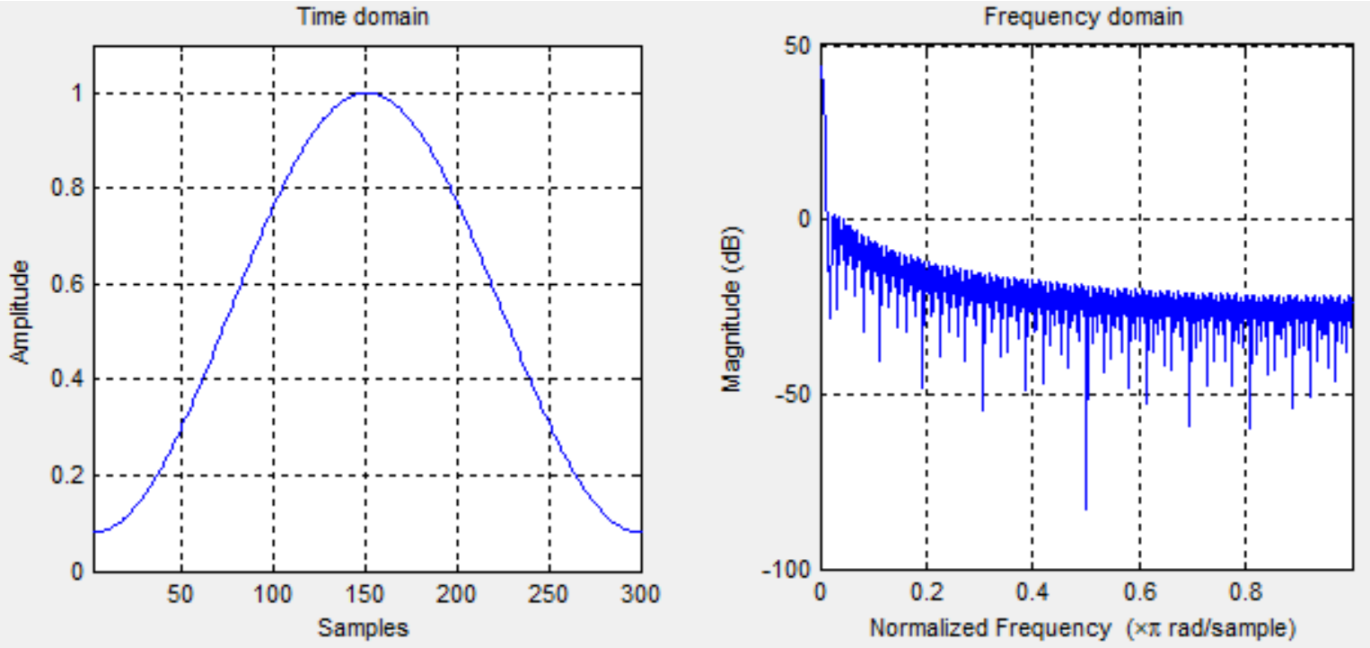
- 为了保证帧的连续性，分帧往往会重叠，重叠部分利用**加窗(windowing)**弱化其影响

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 – 端点检测
Software Architecture – 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

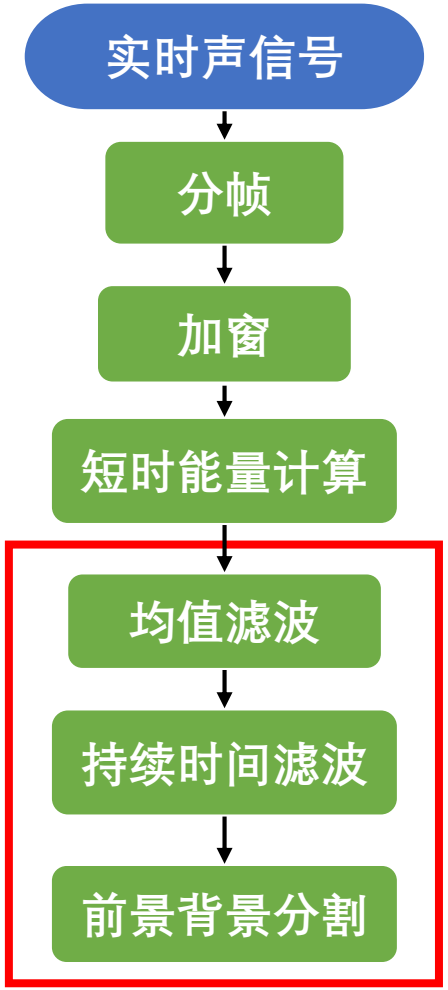
$$h(n)=\begin{cases}0.54-0.46\cos[2\pi n/(N-1)], & 0\leqslant n\leqslant N-1\\ 0, & n=\text{其他}\end{cases}$$

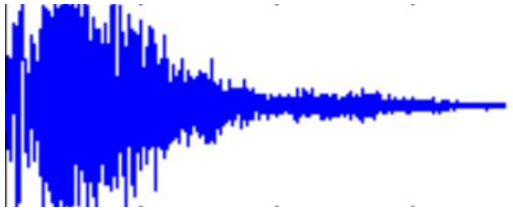


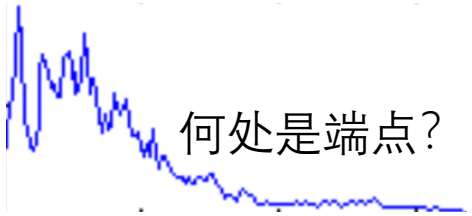
【3】语音信号处理，机械工业出版社，赵力等

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 – 端点检测
Software Architecture – endpoint detection







- 前景 VS 背景

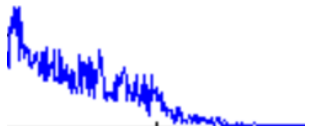
将短时能量作为前景和背景的区分依据，使用**自适应的短时能量阈值【4】**，实现背景片段和可疑片段（前景）的分离

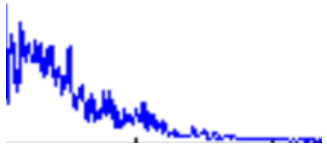
$$THr = \min(En) + 0.2[\max(En) - \min(En)]$$

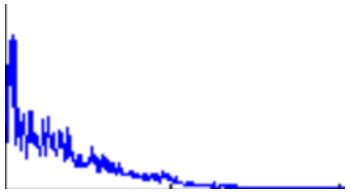
仿真结果发现系数取0.4准确率更高

- 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

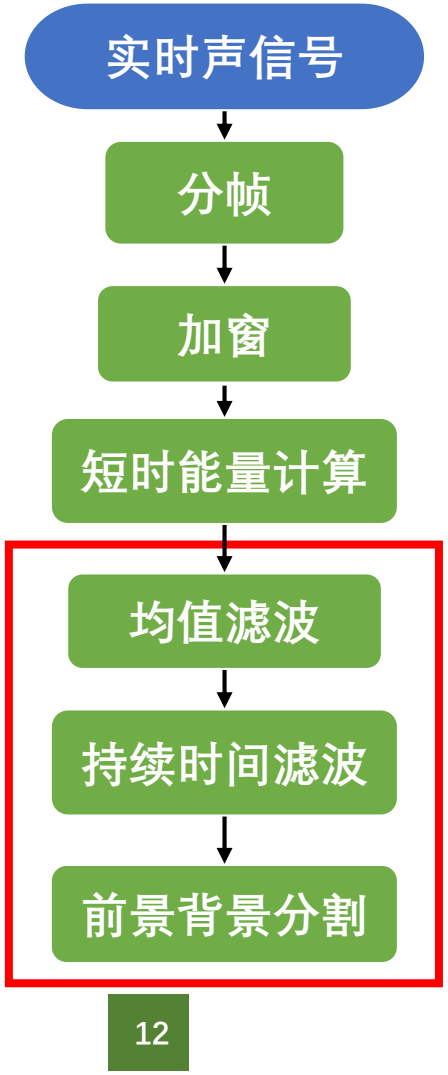


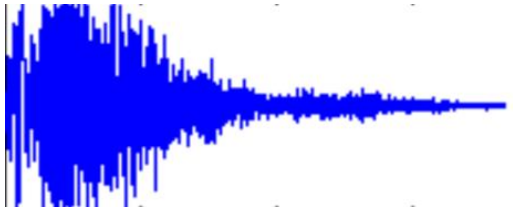


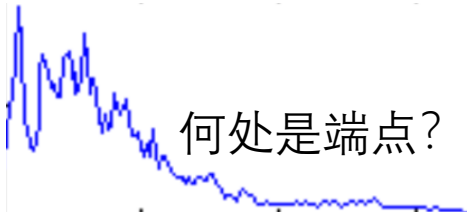


【4】基于短时能量和小波去噪的枪声信号检测方法，电测与仪表，张克刚等

软件架构 – 端点检测
Software Architecture – endpoint detection

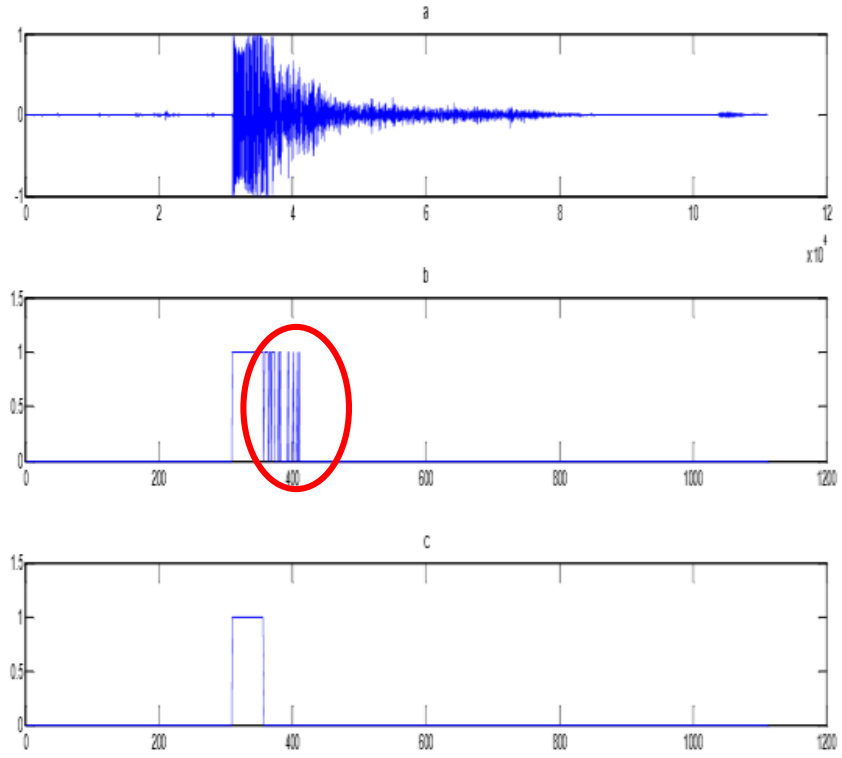






- **持续时间滤波(duration filtering)**

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

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

软件架构 – 端点检测
Software Architecture – endpoint detection

实时声信号

分帧

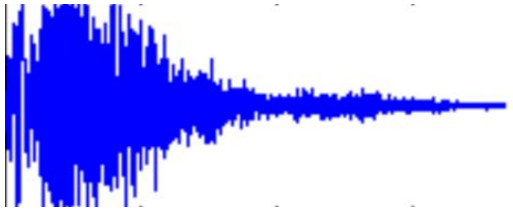
加窗

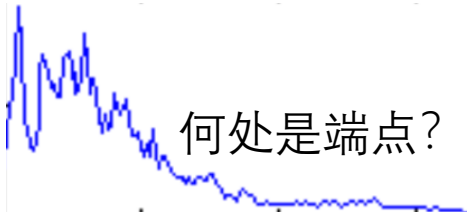
短时能量计算

均值滤波

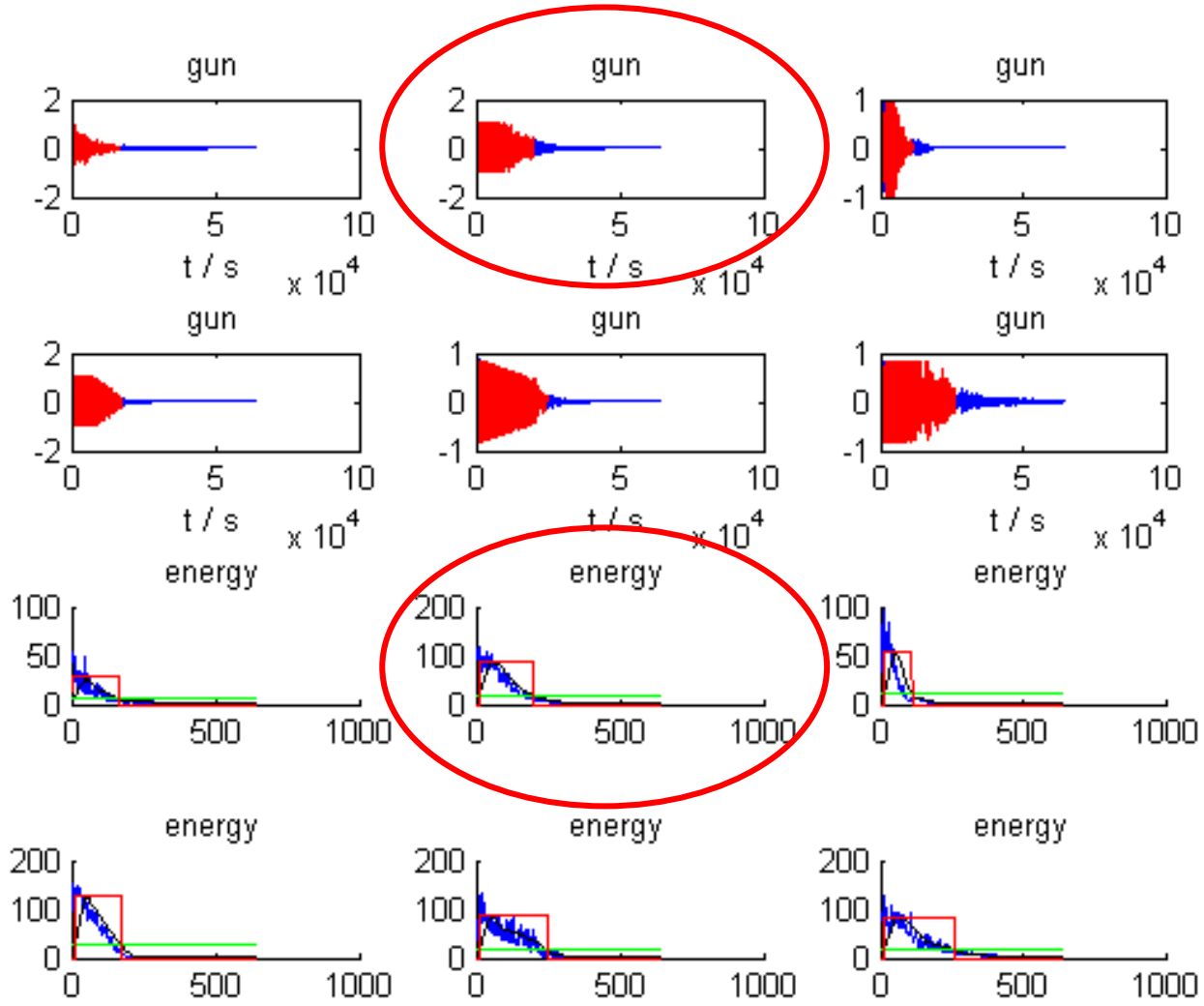
持续时间滤波

前景背景分割



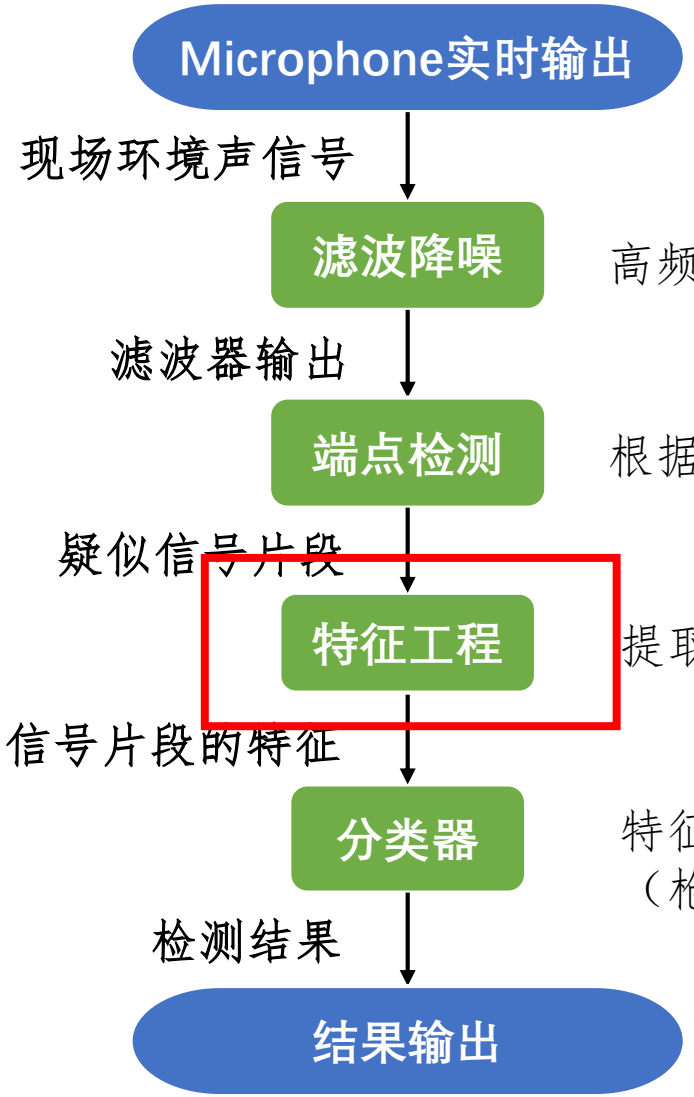


何处是端点？



【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



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 – 特征工程
Software Architecture – 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+

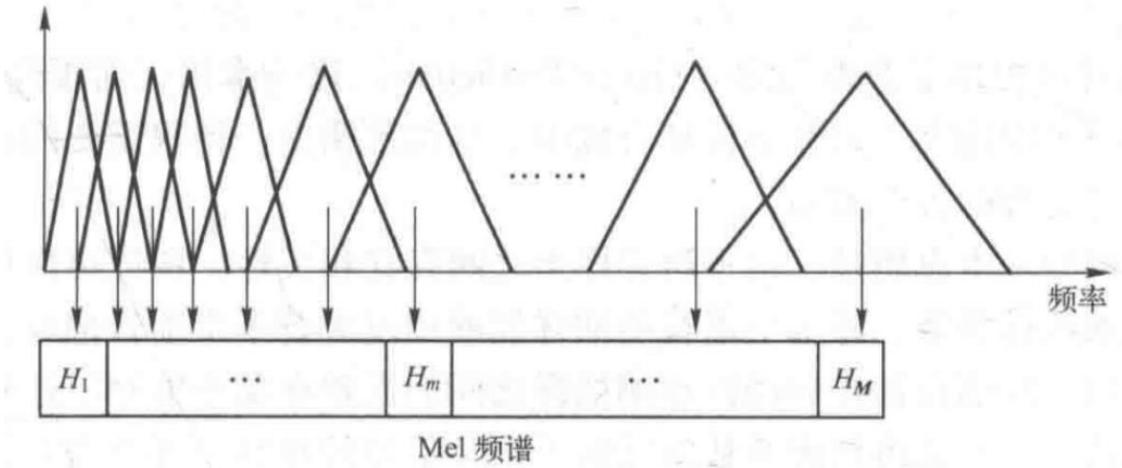
  必须要提取特征作为输入（维数大大减少），分类器的使用才存在可能！

  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

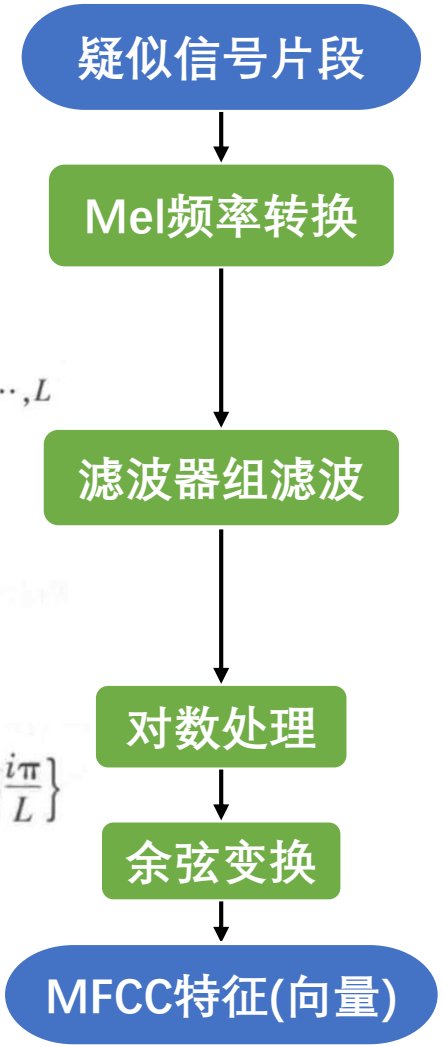


Mel 频谱

【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 – 特征工程
Software Architecture – feature engineering

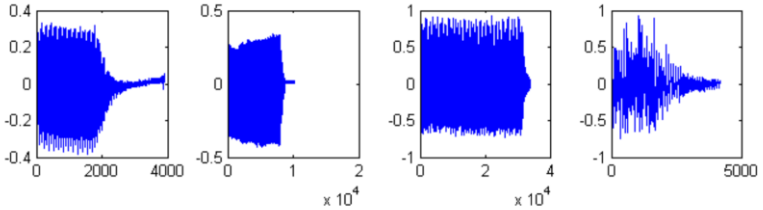


$$\mathrm{Mel}(f) = 2595 \lg(1 + f/700)$$

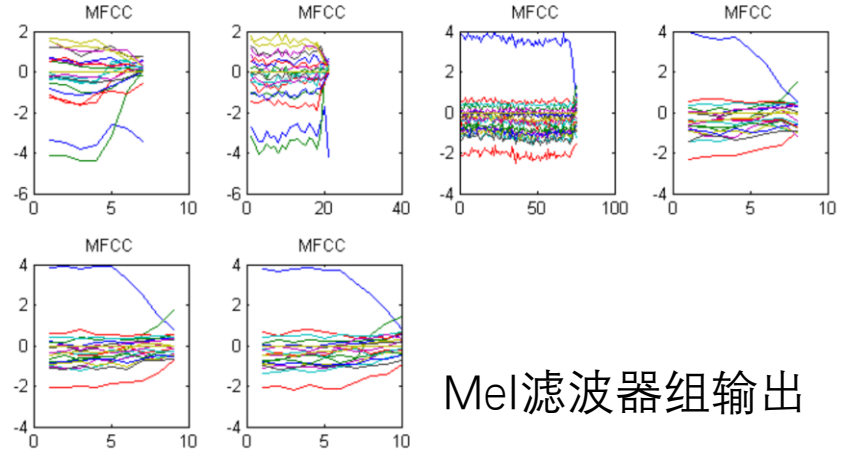
$$m(l) = \sum_{k=o(l)}^{h(l)} W_l(k) \mid X_n(k) \mid \qquad l = 1, 2, \cdots, L$$

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

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

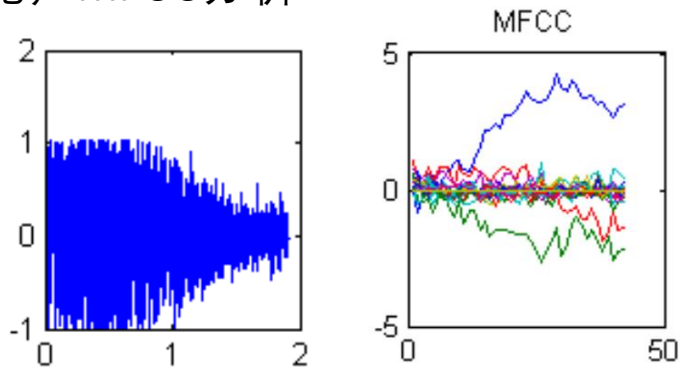


汽车喇叭



Mel滤波器组输出

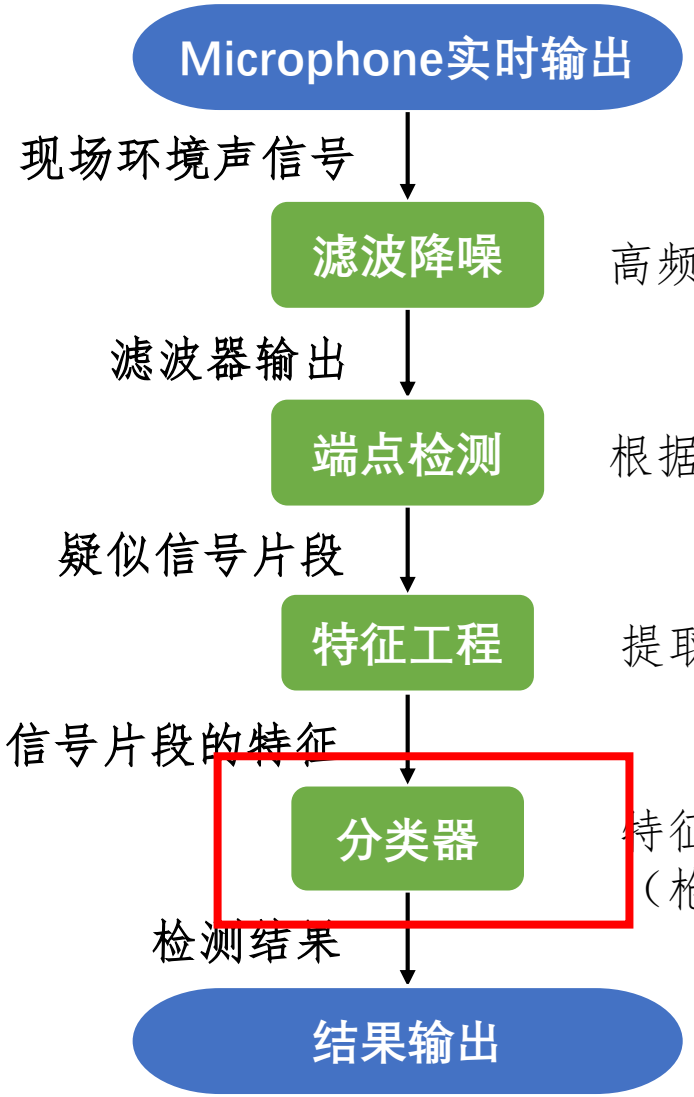
枪声MFCC分析



【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



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 – 分类器
Software Architecture – 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】：

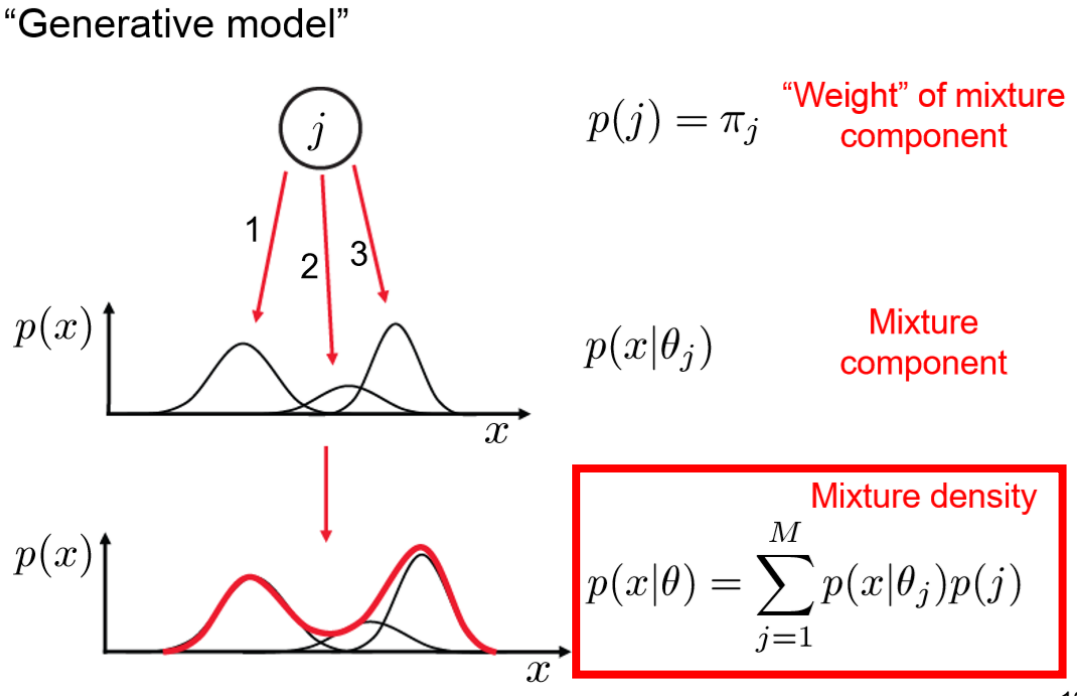
- Bayes决策：需要posterior或者prior & likelihood，需要loss matrix
- 支持向量机SVM(Support Vector Machine)：根据线性可分性分为linear SVM和nonlinear SVM nonlinear SVM中kernel function的选用比较考究，有一整套理论【6】【7】。~~水太深了……~~
- Adaboost(Adaptive Boosting)：sensitive to outliers，母前手头的样本太少
- 随机森林(Random Forest)：训练有点复杂
- GMM + Maximum Likelihood Estimation：本项目中使用

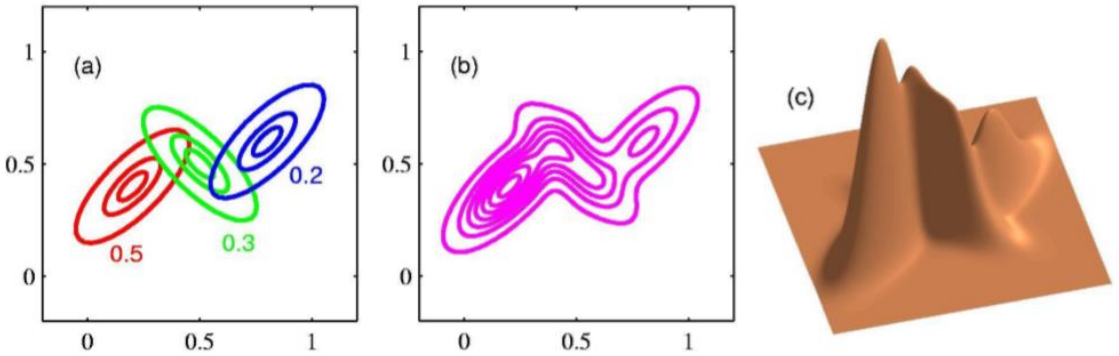
【5】Pattern Recognition and Machine Learning, Springer, Christopher M. Bishop
【6】Learning with Kernels, MIT Press, B. Schoelkopf, A. Smola
【7】A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery

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

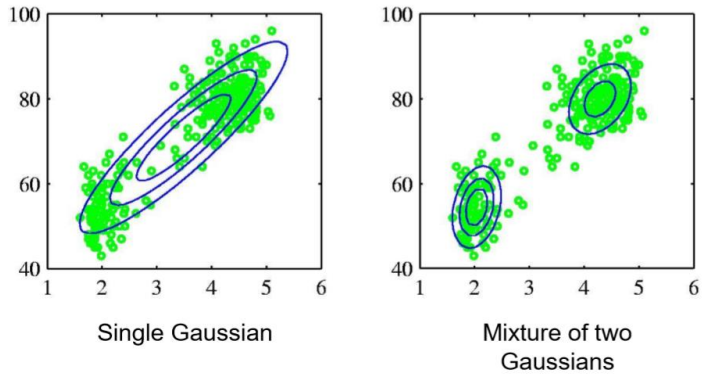
- 混合高斯模型GMM(Gaussian Mixture Model)【5】【10】：又叫MoG(Mixture of Gaussian)





A single parametric distribution is often not sufficient

- E.g. for multimodal data

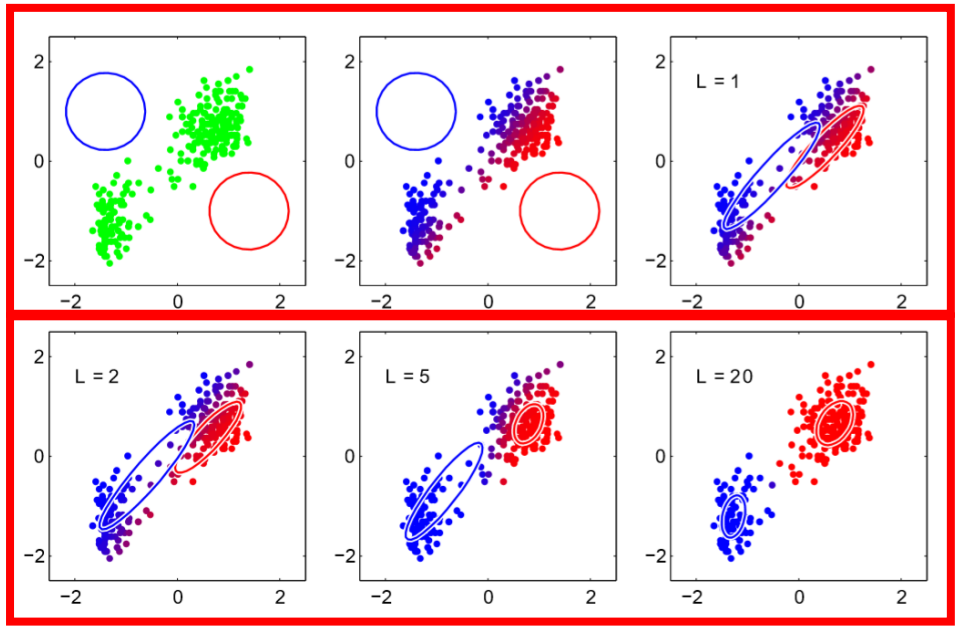


【5】Pattern Recognition and Machine Learning, Springer, Christopher M. Bishop
【9】公共场所下的枪声检测研究，硕士学位论文，朱强强
【10】Machine Learning (Lecture Notes), RWTH University Aachen, Bastian Leibe, Bernt Schiele

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 – 分类器
Software Architecture – classifier

Step 1: Initialization (eg. K-Means)



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

Step 2
EM Algorithm

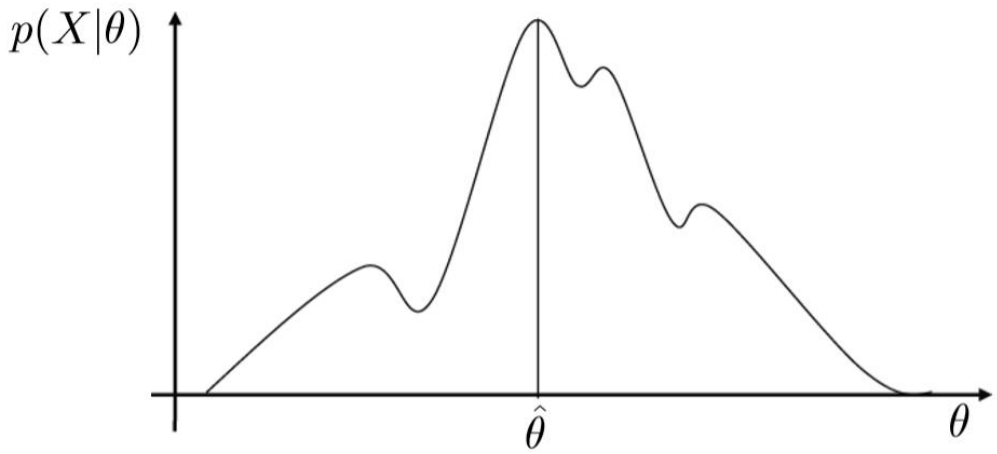
【5】Pattern Recognition and Machine Learning, Springer, Christopher M. Bishop
【10】Machine Learning (Lecture Notes), RWTH University Aachen, Bastian Leibe, Bernt Schiele

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

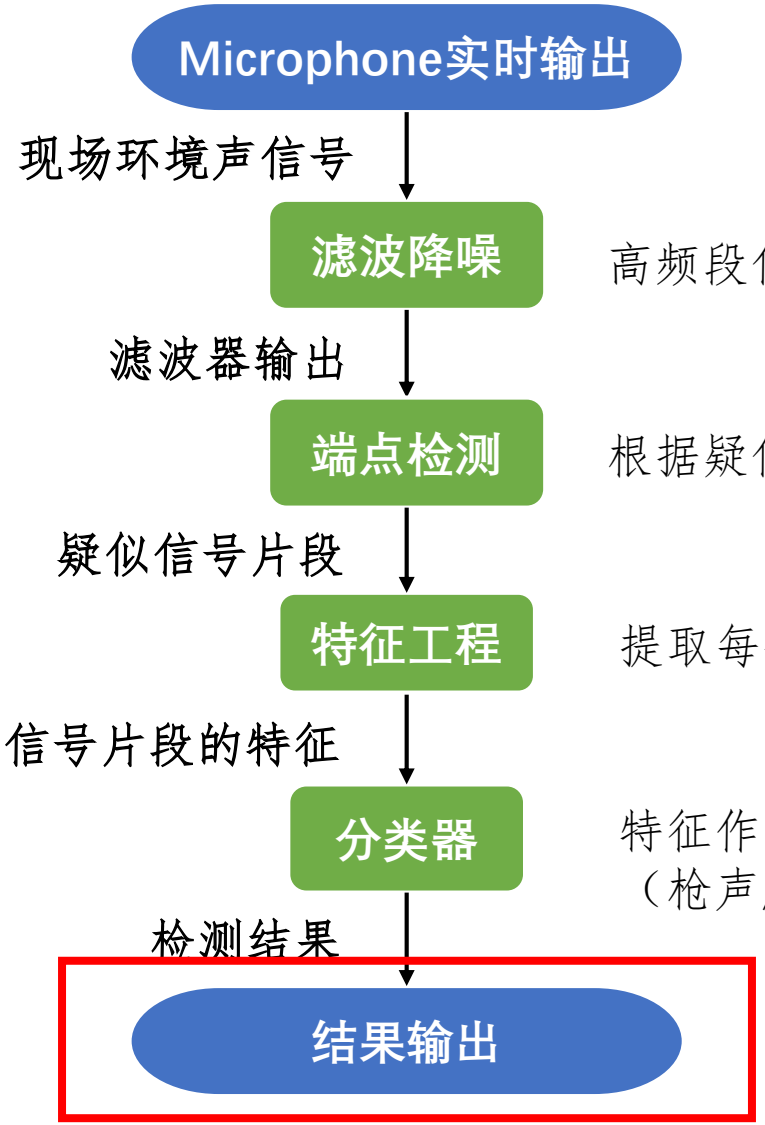
We want to obtain $\hat{\theta}$ such that $L(\hat{\theta})$ is maximized.



【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

参考文献
Reference

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