

# 咳嗽自动定位与检测系统

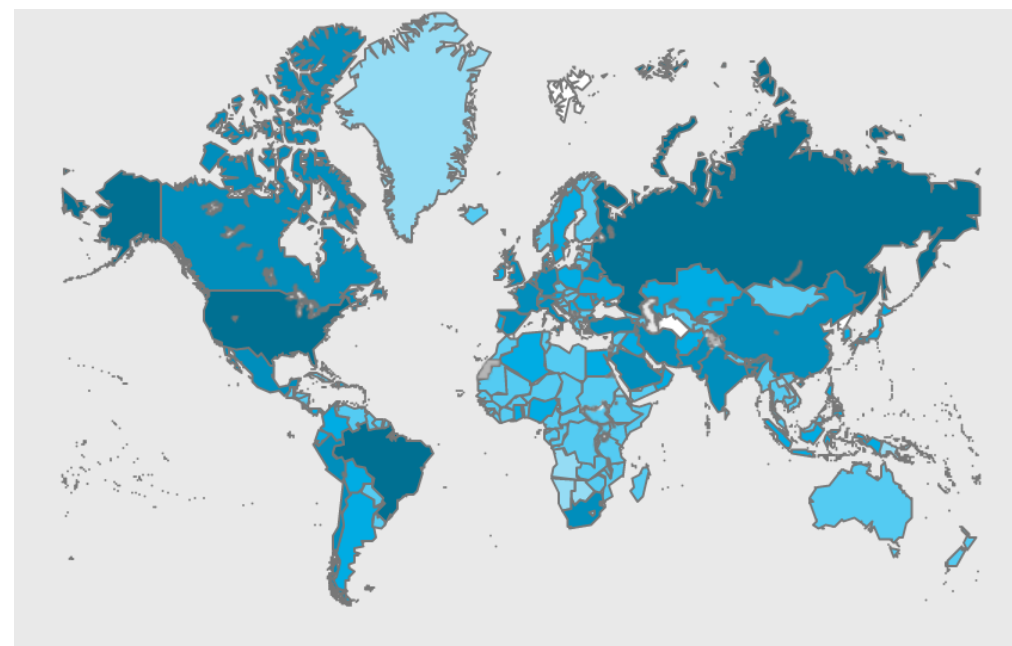
姓名：单心茹      指导教师：JIE LIU教授

# 目录

- 课题背景
- 相关工作
- 声源定位算法与仿真
- 波束形成算法与仿真
- 基于SVM进行咳嗽检测
- 实时检测系统
- 后续工作

# 课题背景

- 2020年3月11日，WHO宣布COVID-19构成“全球大流行”。
- 主要通过飞沫，人与人的密切接触造成感染。
- **典型症状**包括发热，干咳，乏力，咳痰，气短，咽痛。
- 结合AI，麦克风阵列信号处理，设计咳嗽自动定位与检测系统，**定位咳嗽声源，检测咳嗽事件的发生，**实时监测公共场所的潜在病例。



 Worldwide

Confirmed  
**7.04M**  
+109K

Recovered  
-

Deaths  
**404K**  
+3,539

# 相关工作

- **咳嗽检测仪**: HACC, LCM(HMM+KWS, 灵敏度91%, 特异性99%, 用于临床实验)。常用于医疗康复, 为医生提供辅助诊断。
- **咳嗽检测模型**: 早期(HMM+KWS), 目前(ML, CNN, RNN)。
- **数据集**: 医院提供, 网络音频。
- **一般处理流程**: 声音采集, 静音检测, 特征提取及选择, 分类器训练, 模型评估及测试。

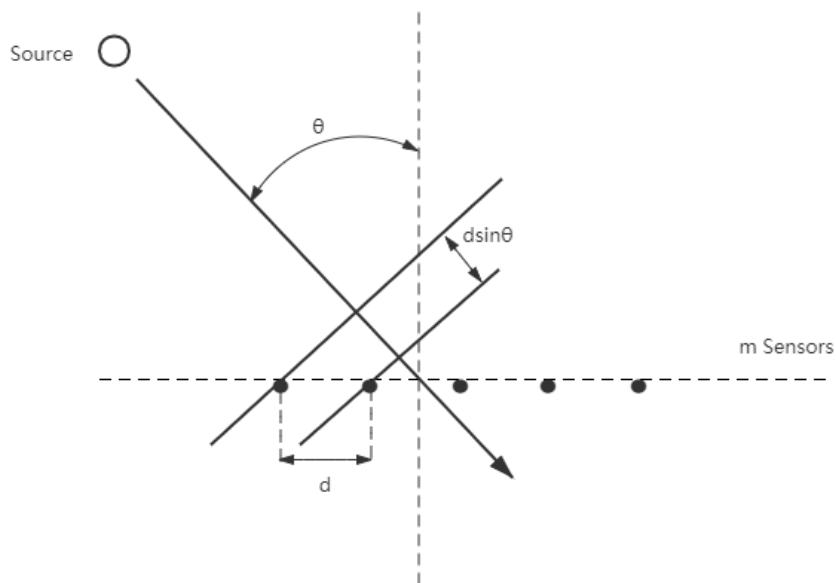


HACC咳嗽检测仪



LCM咳嗽检测仪

# 声源定位算法



麦克风阵列模型

- 麦克风阵列通道输出信号:

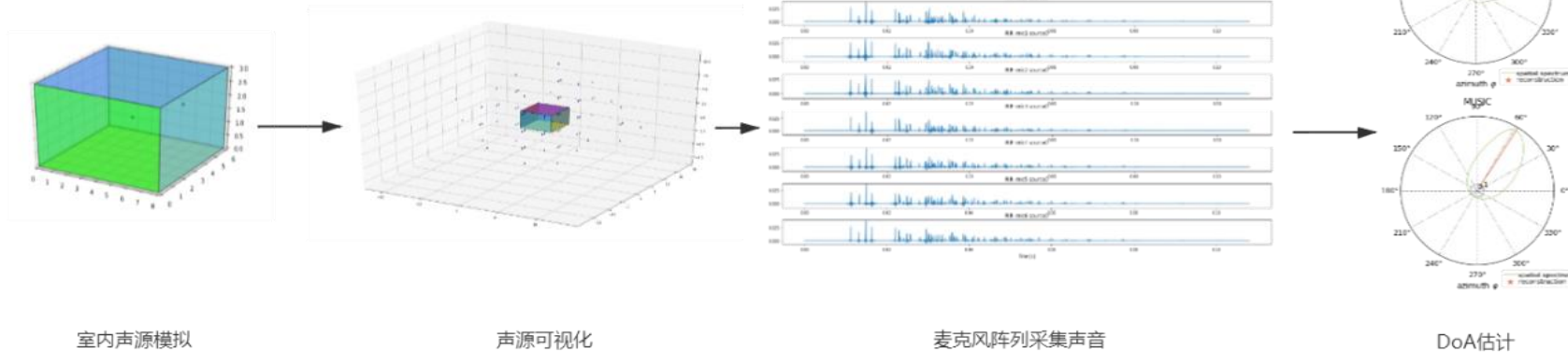
$$\begin{aligned}y_m(k) &= x_m(k) + v_m(k) \\ &= x(k - \tau_m) + v_m(k)\end{aligned}$$

- 估计通道之间**相对时延**:

$$\tau_m = (m - 1) \frac{d \sin \theta}{c}$$

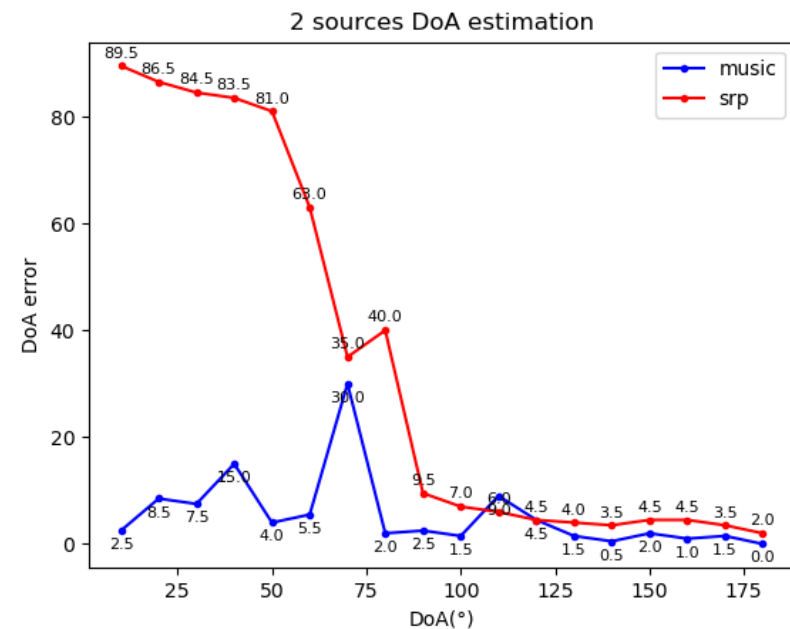
- **MUSIC算法**: 将阵列输出的协方差矩阵进行特征分解, 根据特征值的大小顺序将特征空间划分为信号子空间和噪声子空间, 然后利用两个子空间的正交性来估计DoA。
- **SRP-PHAT算法**: 计算两两阵元的GCC之和, 作为SRP值, 并进行PHAT加权。
- **DoA估计可分为两步**: TDOA估计->网格化峰值搜索估计DoA

# DoA估计仿真

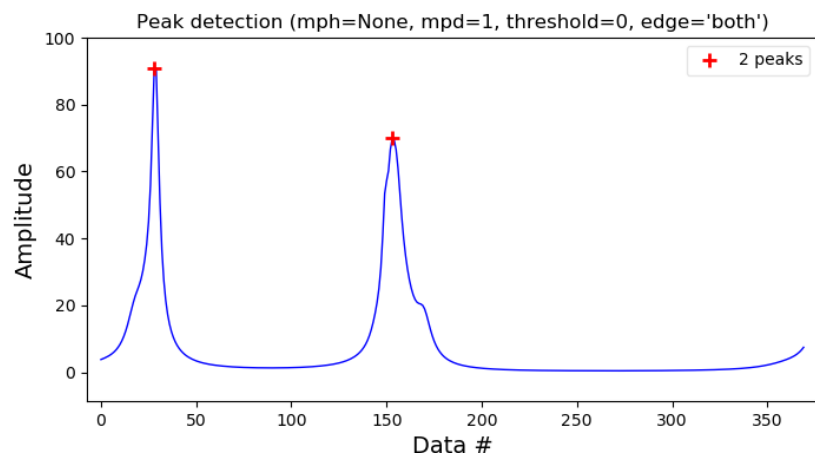


- **多声源二维仿真：** 房间(10m×10m)，最大反射次数为2，墙面吸收率为0.1，SNR=30dB。  
固定一声源位置(0°, 3m)。  
变换另一声源角度(DoA ∈ [10°, 180°], 3m)。

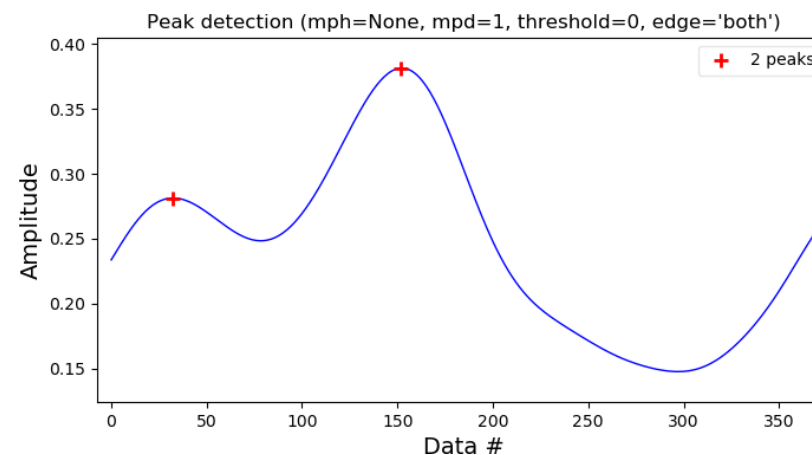
**SRP-PHAT算法在两声源DoA相近时，误差较大，而MUSIC算法表现良好。**



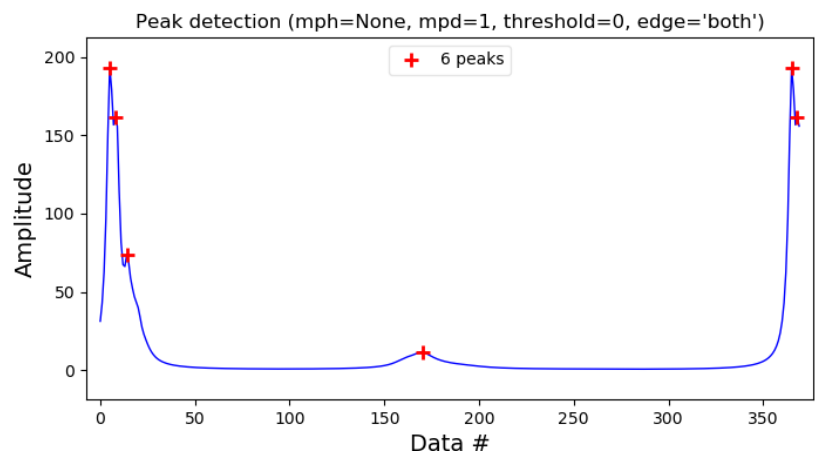
# DoA估计仿真



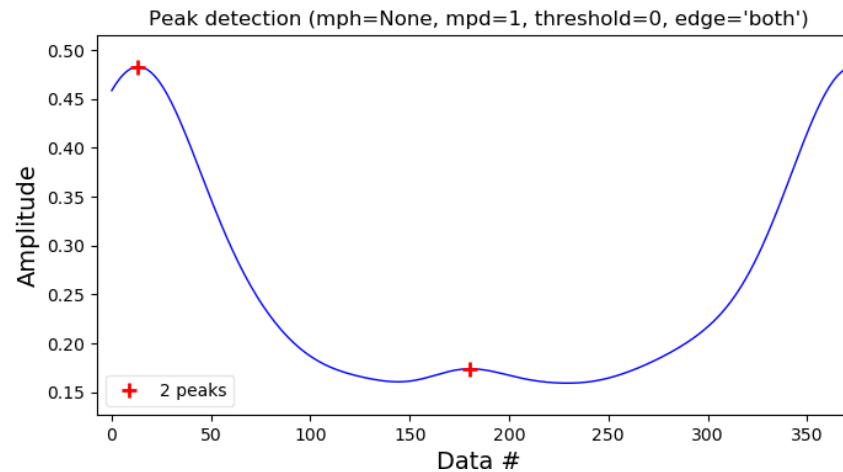
a) MUSIC峰值(两声源位置分离)



b) SRP峰值(两声源位置分离)



c) MUSIC峰值(两声源位置相近)



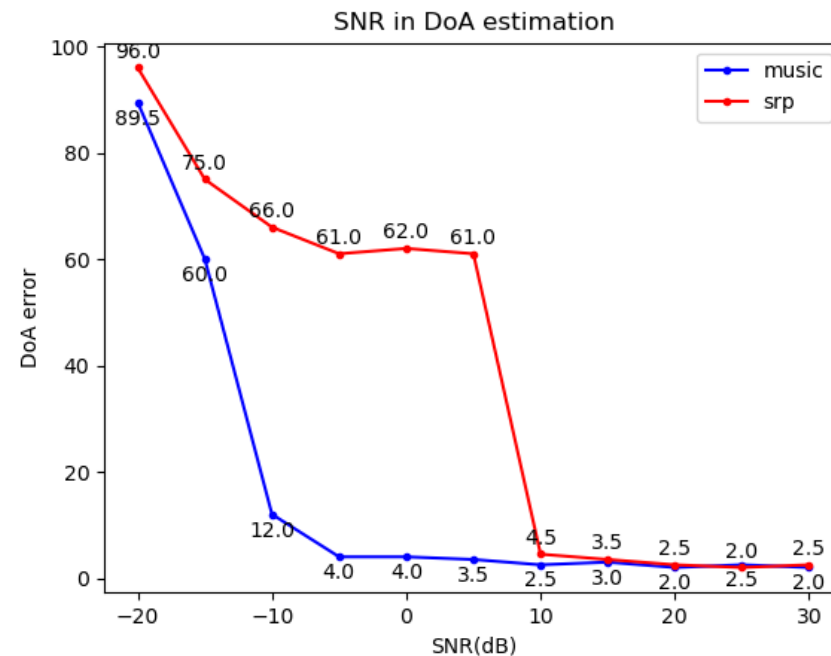
d) SRP峰值(两声源位置相近)

# DoA估计仿真

- 探究SNR对DoA估计的影响，两声源位置较分离，如右图。

## 结论：

- 低信噪比环境严重影响DoA估计，提高信噪比有利于声源位置估计。
- 在声源位置相近，要求多声源估计分辨率高时，MUSIC算法峰值明显，而 SRP-PHAT峰值较平滑。



SNR对DoA估计的影响



# 波束形成算法

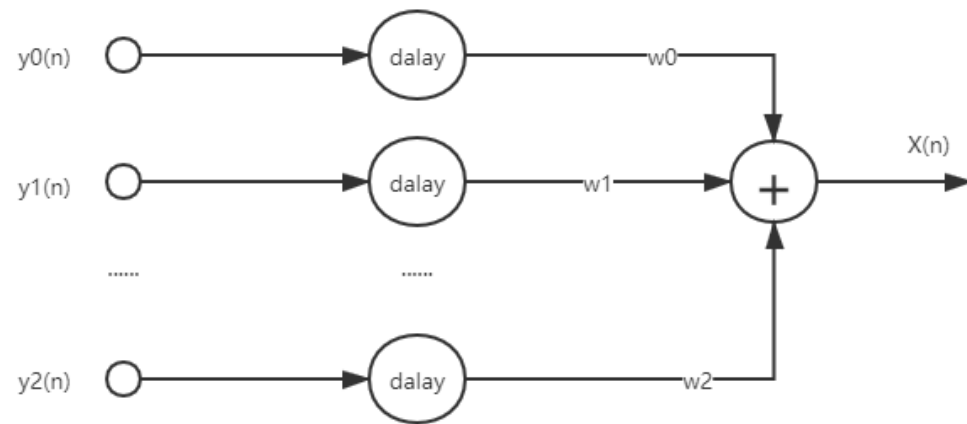
主要作用：增强期望信号，抑制干扰，将阵列接收聚集到一个方向，形成“波束”。

- 固定波束形成，可以对某一方向的声源进行抑制，抑制的强度不变。
- 自适应波束形成，抑制性能可以对环境信号的变化做出调整，鲁棒性更好

- 设计滤波器 $h(\omega)$ ：增强感兴趣的信号，抑制干扰信号

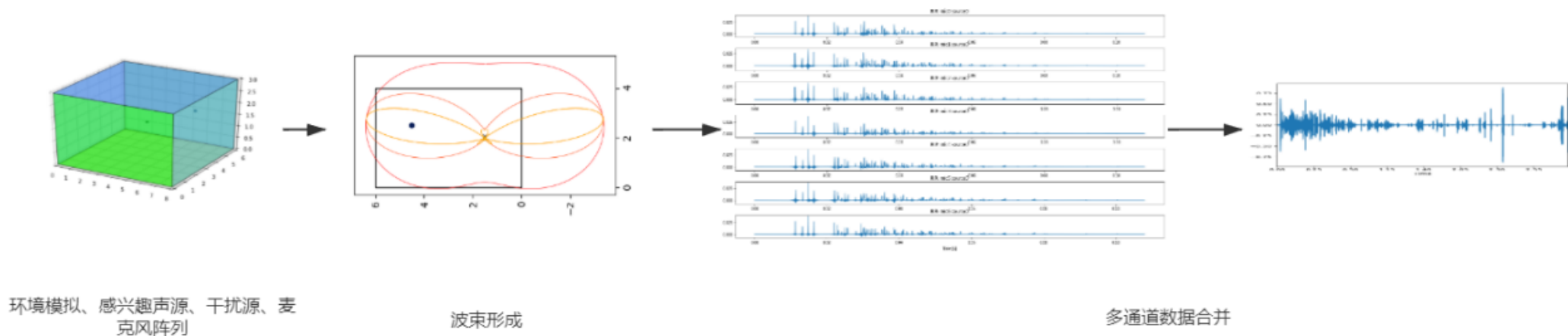
$$y(\omega) = x(\omega) + v(\omega)$$

$$\begin{aligned} Z(\omega) &= h^H(\omega)y(\omega) \\ &= h^H(\omega)x(\omega) + h^H(\omega)v(\omega) \end{aligned}$$



延时-求和滤波器

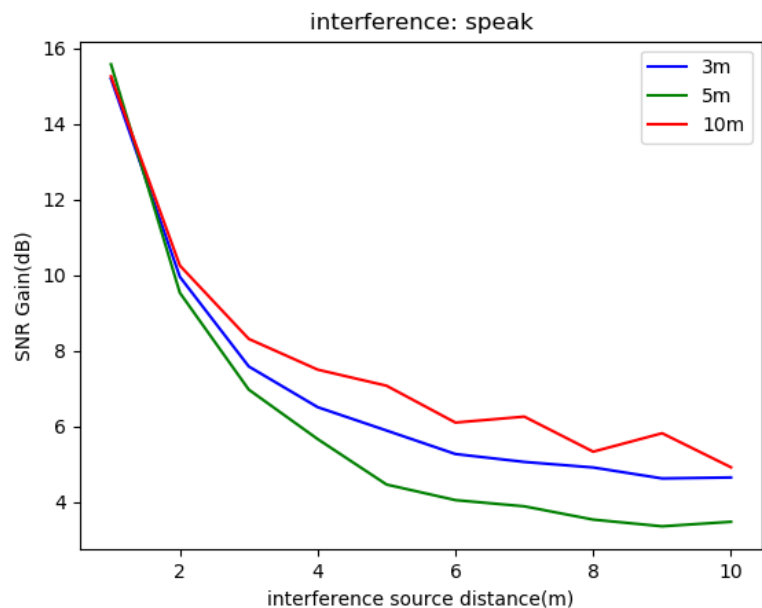
# 波束形成仿真



- **二维MVDR仿真，探究干扰源距离对SNR增益的影响：**感兴趣声源（咳嗽）DoA =  $0^\circ$ ，相对麦克风阵列距离为3m/5m/10m；干扰声源（说话）DoA从 $30^\circ$ 至 $180^\circ$ ，分辨率为 $30^\circ$ ，相对阵列距离从1m到10m，分辨率1m。

## 结论：

- 当干扰源相对麦克风阵列位置越近，则信噪比增益较高，越远信噪比增益越低，在麦克风阵列基本拾音范围（5m内）信噪比增益大致范围为[6, 15]，可以明显提高信噪比，便于咳嗽识别。



# SVM检测咳嗽

- 数据集:

部分Google AudioSet咳嗽音频集与YouTube咳嗽音频，共采用82例咳嗽样本，84例非咳嗽样本（包括说话声和音乐等教室常见声音类型）。

- 特征提取:

短期特征->中期特征->长期特征

短期特征：帧长50ms，帧移25ms，提取34个音频特征（过零率，能量熵，MFCC等）。

中期特征：计算短期特征的均值和标准差。

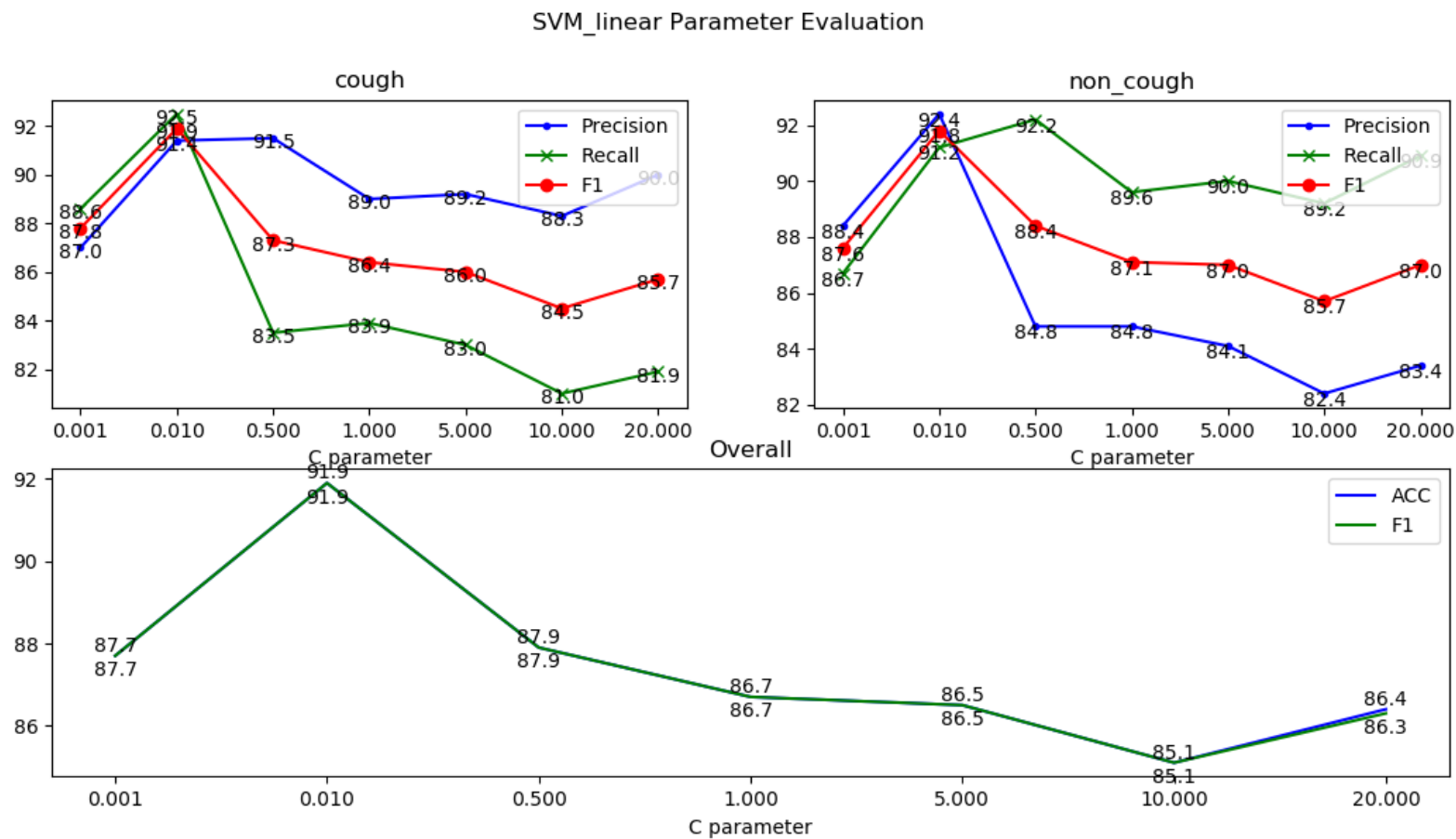
长期特征：计算中期特征的均值和标准差。

- SVM模型:

对比SVM\_linear模型与SVM\_RBF模型，测试不同惩罚系数C，进行10折交叉验证。

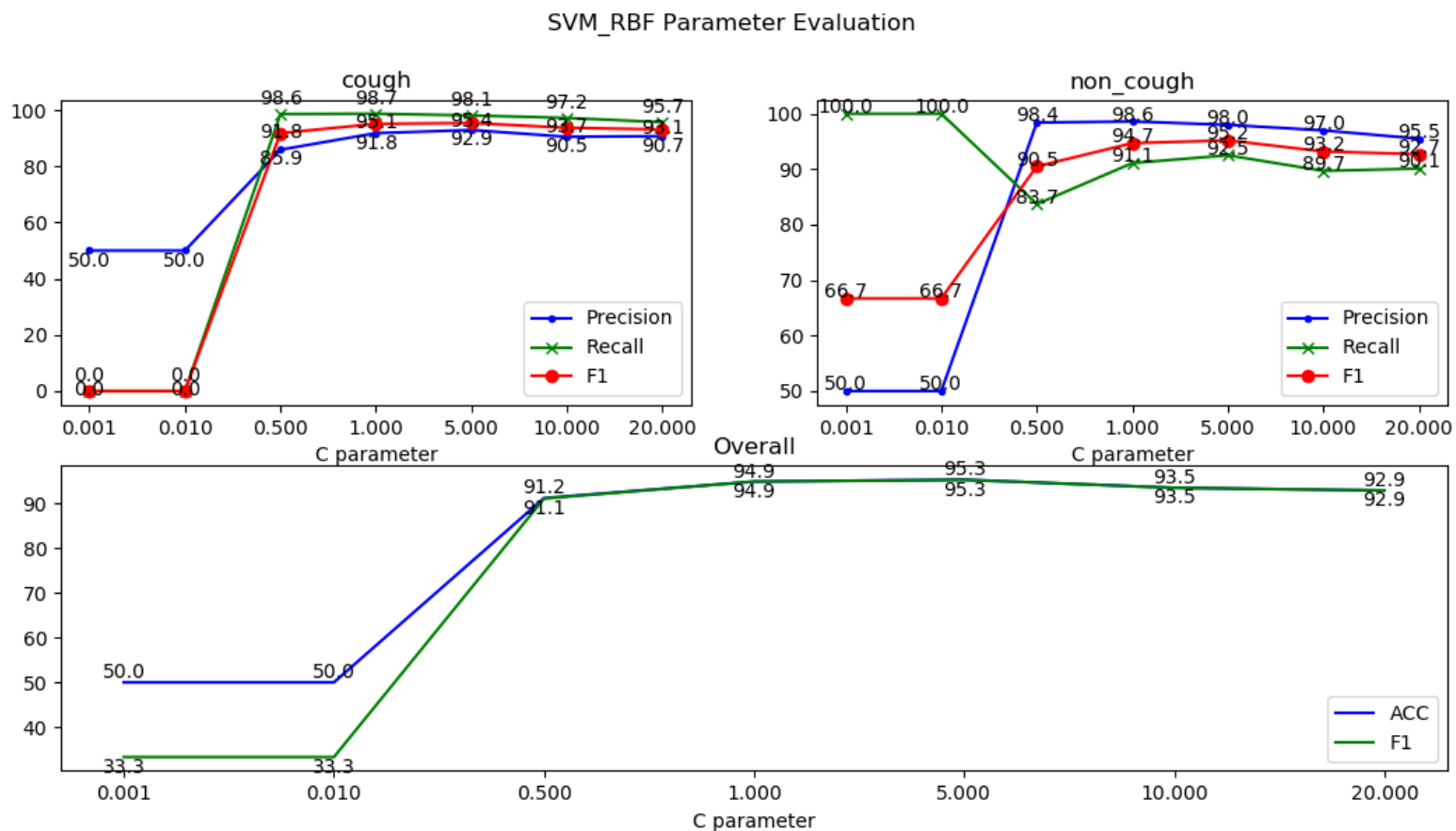
# SVM检测咳嗽

- SVM\_linear模型，最优参数C为0.01，F1指数为91.9%



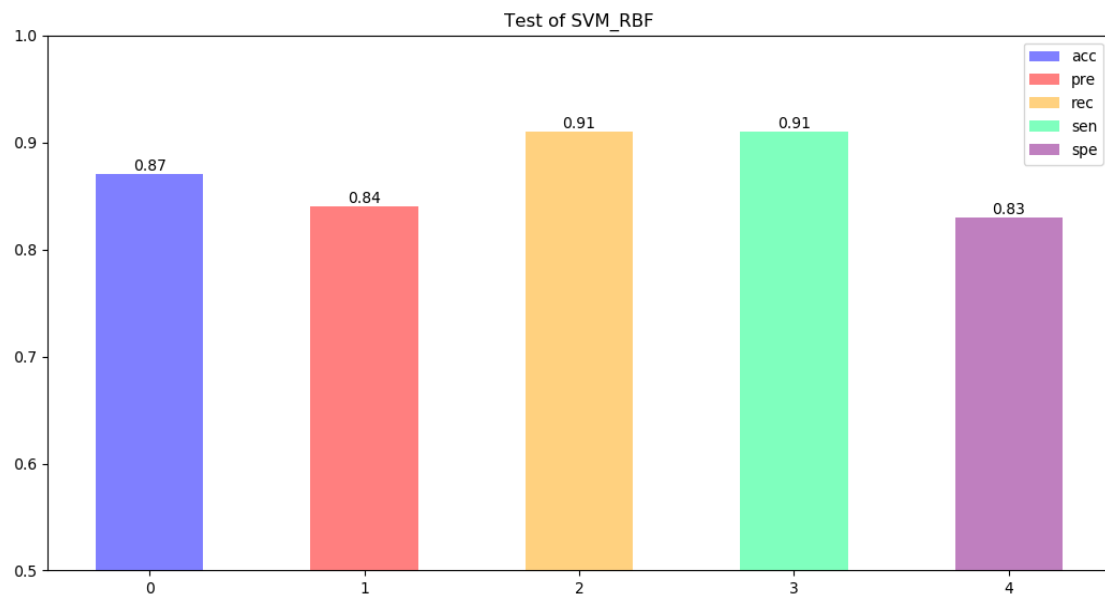
# SVM检测咳嗽

- SVM\_RBF模型, 最优参数C为1.0, F1指数为95.3%

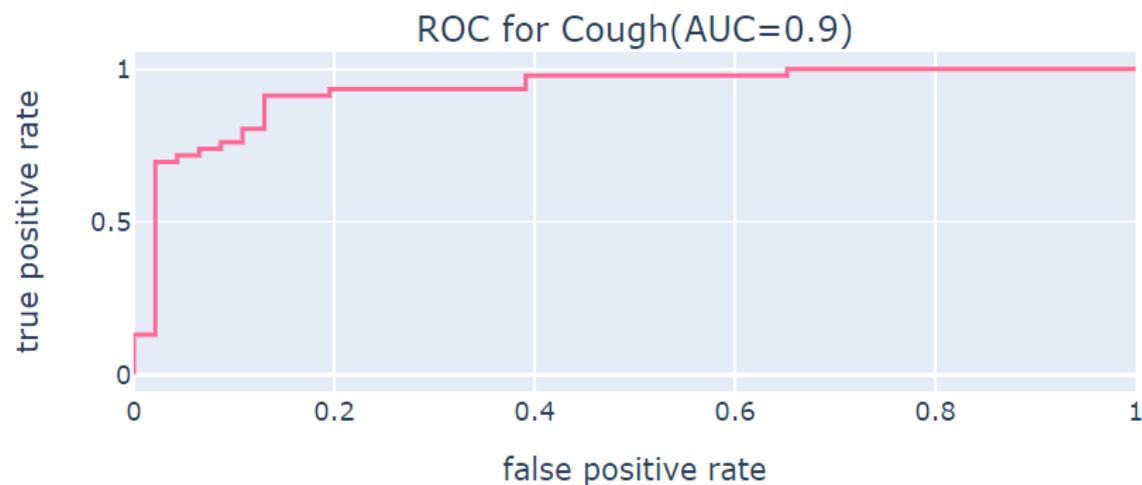


# SVM检测咳嗽

- **测试集：**基于AudioSet的咳嗽，说话，音乐等音频数据集，包含咳嗽样本46例，非咳嗽样本46例。
- 准确率为87%,精确度为84%，召回率为91%，灵敏度为91%，特异性为83%，AUC值0.9。



测试集评价指标

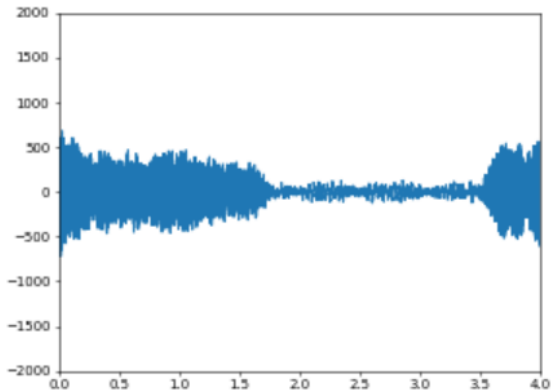


测试集ROC曲线

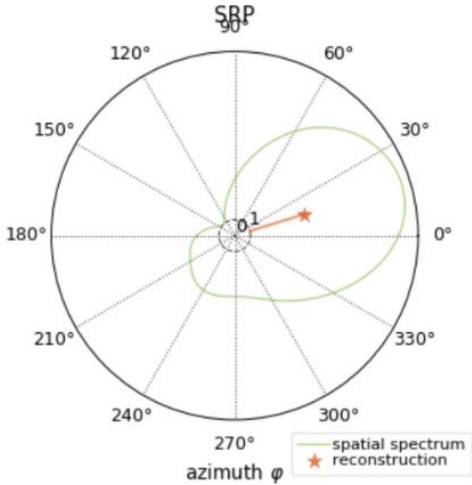
# 实时检测系统



系统流程



Html+js实时可视化波形和DoA



```
[2020-06-03 21:28:11,980][7364][INFO] ---- SRP-PHAT estimate azimuth:[23.]degrees
[2020-06-03 21:28:12,176][7364][INFO] ---- nonCough
[2020-06-03 21:28:12,176][7364][INFO] ---- Total run time: 15.39s |Mean process time: 1.05s
[2020-06-03 21:28:13,024][7364][INFO] ---- #####
[2020-06-03 21:28:14,128][7364][INFO] ---- SRP-PHAT estimate azimuth:[19.]degrees
[2020-06-03 21:28:14,287][7364][INFO] ---- nonCough
[2020-06-03 21:28:14,287][7364][INFO] ---- Total run time: 17.51s |Mean process time: 1.07s
[2020-06-03 21:28:15,044][7364][INFO] ---- #####
[2020-06-03 21:28:15,996][7364][INFO] ---- SRP-PHAT estimate azimuth:[12.]degrees
[2020-06-03 21:28:16,151][7364][INFO] ---- nonCough
[2020-06-03 21:28:16,151][7364][INFO] ---- Total run time: 19.37s |Mean process time: 1.08s
[2020-06-03 21:28:17,062][7364][INFO] ---- #####
[2020-06-03 21:28:17,624][7364][INFO] ---- silent
[2020-06-03 21:28:17,624][7364][INFO] ---- Total run time: 20.84s |Mean process time: 1.02s
[2020-06-03 21:28:19,083][7364][INFO] ---- #####
[2020-06-03 21:28:19,670][7364][INFO] ---- silent
[2020-06-03 21:28:19,670][7364][INFO] ---- Total run time: 22.89s |Mean process time: 0.99s
```

日志记录详细信息

# 运行时间测试

- 系统配置参数:

CPU	RAM	OS	Mic Array
Intel Core i5-6200U (4 CPUs)	8G	x64, win10	Azure Kinect

- 评价指标 (代码块总运行时间与代码块运行次数的比值) :  $P_{mean} = \frac{\sum_i^N time(i)}{N}$

系统总运行 时间	测试1 $P_{mean}(s)$	测试2 $P_{mean}(s)$	测试3 $P_{mean}(s)$	均值 $\mu$	标准差 $\sigma$
1min	0.84	1.14	1.15	1.04	0.0207
10min	0.96	1.45	0.83	1.08	0.0712
30min	0.91	1.42	0.88	1.07	0.0614
60min	0.87	1.34	0.91	1.04	0.0453



# 实时检测系统

- 对比静音与非静音片段，多声源运行时间：

静音片段 $P_{mean}(s)$	单声源非静音片段 $P_{mean}(s)$	两声源非静音片段 $P_{mean}(s)$	三声源非静音片段 $P_{mean}(s)$
0.63	1.17	1.33	1.48

结论：

- 单声源检测算法平均运行时间为1.05s，其中静音片段检测算法平均运行时间为0.63s。
- 静音检测减少算法运行时间，降低功耗。
- 随着声源数量增加，算法后续步骤耗时增加。

# 后续工作

本文实现了咳嗽自动定位与检测的基线系统，存在很大优化空间，后续工作可从以下几个方面进行：

- 自动估计活跃声源数量。
- 优化SRP-PHAT算法，解决在声源位置相近时，峰值平滑的问题。
- 根据DoA估计算法复杂度，量化其对实时系统性能的影响，优化算法性能。
- 实验多声源情况下，量化通过波束形成提高SNR对咳嗽检测的影响。
- 部署到嵌入式终端设备，进行实际测试。

**THANKS ;)**