**咳嗽自动定位与检测系统**

**单心茹**

院 （系）：计算机科学与技术学院 专 业：物联网工程

学　　号：1160100626 指导教师：JIE LIU

**2020年6月**



**毕业设计（论文）**

题 目 咳嗽自动定位与检测系统

专 业 物联网工程

学　　 号 1160100626

学 生 单心茹

指 导 教 师 JIE LIU

答 辩 日 期 2020.6.12

# 摘 要

2020年，新型冠状病毒肺炎（COVID-19）感染爆发，在全球形成“大流行”。主要通过飞沫与密切接触造成呼吸系统感染，典型症状包括发热、咳嗽、肌肉酸痛、疲劳等。目前，中国疫情防控处于中后期阶段，做好观察、隔离和追踪，逐步复工复学阶段。由于新冠肺炎可能仍会持续一段时间，防控工作不能放松，监控公共场所的潜在病例非常有必要。

本文以新冠肺炎的典型症状——咳嗽作为切入点，研发咳嗽自动定位与检测系统，结合人工智能技术，麦克风阵列信号处理技术，在教室，实验室等学校公共场所，实现自动检测咳嗽并定位的功能，可监控潜在病例，作为一种自动辅助诊断，解决复工复学情况下人员聚集空间中不易排查的问题，提供除发烧症状以外的早期新冠症状监测，减少监测过程中大量人力物力消耗，为呼吸系统疾病诊断与复工复学监控提供帮助。

系统集成音频实时处理，声源定位，波束形成，咳嗽检测等功能。在声源定位模块，对比MUSIC算法与SRP-PHAT算法，并通过仿真实验验证算法性能。系统通过自适应波束形成算法提高信噪比，利于咳嗽检测，在仿真实验部分验证了MVDR算法可有效提高信噪比。基于AudioSet与YouTube音频训练SVM\_RBF模型，进行咳嗽检测，10折交叉验证中F1指数为95.3%，基于测试集灵敏度为91%，特异性为83%，AUC值达0.9，分类性能良好。

系统积极响应疫情科研攻关的号召，实现咳嗽自动定位与检测的功能，为新冠肺炎时期的抗疫工作提供新思路。

**关键词：**音频信号处理；声源定位；波束形成；机器学习；实时系统；

# Abstract

In 2020, the outbreak of COVID-19 infection has caused a global pandemic. It is spreads mainly from person to person, through respiratory droplets produced when an infected person coughs or sneezes. The typical symptoms include fever, cough, muscle soreness and fatigue. At the time when this thesis is written, the epidemic is basically under control in China. However, when the country gradually resumes work and school, or in the next flu season, the danger of a resurgence is still possible. If COVID-19, or similar respiratory infectious diseases, spread, it is imperative to monitor potential cases in public places.

This thesis takes coughing, a typical symptom of COVID-19, as the entry point to develop an automatic cough location and detection system. The system combines AI and microphone array signal processing technologies to monitor potential cases in classrooms, laboratories and other public places. This can be used as an automatic auxiliary diagnosis device to solve the problem of checking out gathering people. Besides, it can also provide early monitoring of COVID-19 symptoms in addition to fever, reducing a large amount of manpower and material consumption in the monitoring process, thus provide help for the diagnosis of respiratory diseases and the resumption of work and school monitoring.

The system integrates audio signal real-time processing, direction of arrival estimation(DoA), beamforming, cough detection, etc. In the DoA estimation module, two DoA algorithms are compared, and the performance of the algorithms is verified by simulation experiments. Adaptive beamforming algorithm aimed at improving the signal-to-noise ratio (SNR) of audio, which is beneficial to cough detection. The simulation experiment verifies that MVDR algorithm can effectively improve the SNR. In the cough detection phase, we trained a SVM\_RBF model based on Google AudioSet and YouTube audio dataset. In 10-fold cross validation, our method yields a F1 index 95.3%. Based on the test set, the sensitivity is 91%; the specificity is 83%; and the AUC is up to 0.9, indicating good classification performance.

The system directly responds to the call for scientific research on the pandemic, realizes the function of capability of automatic cough location and detection, and provides new ideas to the counter COVID-19 like epidemics.

**Keywords:** Audio signal processing, DoA estimation, Beamforming, Machine learning, Real-time system

目 录

[摘 要 I](#_Toc42936825)

[Abstract II](#_Toc42936826)

[第一章 绪 论 1](#_Toc42936827)

[1.1 课题背景及研究的目的和意义 1](#_Toc42936828)

[1.2 国内外研究现状分析 2](#_Toc42936829)

[1.3本文主要研究内容 6](#_Toc42936830)

[第二章 声源定位与波束形成算法理论基础 7](#_Toc42936831)

[2.1麦克风阵列信号处理理论 7](#_Toc42936832)

[2.2声源位置估计(DoA)算法 8](#_Toc42936833)

[2.3波束形成(beamforming)算法 10](#_Toc42936834)

[2.4声源定位与波束形成仿真 12](#_Toc42936835)

[2.5 本章小结 18](#_Toc42936836)

[第三章 基于SVM的咳嗽检测方法 19](#_Toc42936837)

[3.1数据来源 19](#_Toc42936838)

[3.2 音频采集及特征提取 20](#_Toc42936839)

[3.3 基于SVM二分类的咳嗽检测 21](#_Toc42936840)

[3.4 模型验证与评价 23](#_Toc42936841)

[3.5本章小结 28](#_Toc42936842)

[第四章 咳嗽自动定位与检测系统集成 29](#_Toc42936843)

[4.1 实时系统流程 29](#_Toc42936844)

[4.2 实时监测系统集成 30](#_Toc42936845)

[4.3本章小结 32](#_Toc42936846)

[结论 33](#_Toc42936847)

[参考文献 34](#_Toc42936848)

[哈尔滨工业大学本科毕业设计（论文）原创性声明 37](#_Toc42936849)

[致 谢 38](#_Toc42936850)

# 第一章 绪 论

1.1 课题背景及研究的目的和意义

1.1.1 课题研究背景

2019年12月起，新型冠状病毒肺炎（Corona Virus Disease 2019，COVID-19）爆发，蔓延全球，截止2020年5月22日，全球已确诊511万例，已恢复195万例，死亡人数达33.3万。根据世界卫生组织（WHO）的资料，目前尚无针对新冠病毒的疫苗或特定的抗病毒治疗方法。2020年1月30日，WHO宣布新冠肺炎为“国际关注的突发公共卫生事件”（PHEIC）；2月12日，WHO将新型冠状病毒引起的疾病命名为（COVID-19）；3月11日，WHO宣布其构成“全球大流行”，感染人数远超SARS。

病原学研究证明，新冠状病毒和SARS病毒同属一个病毒科，根据目前的数据基础，COVID-19最初可能寄生于蝙蝠，通过穿山甲或者其他野生动物传染人类[1]。从流行病学的角度，COVID-19潜伏期范围为0~24天，主要通过被感染者咳嗽，打喷嚏时产生的飞沫传播，人与人的密切接触同样造成感染。COVID-19的临床症状是非特异性的，常见症状包括发烧、咳嗽、肌肉酸痛或疲劳，发烧前几天可能出现腹泻、恶心，少数患者出现头痛或者咳血，甚至是无症状。在较严重病例中，感染可导致肺炎、严重急性呼吸综合征、肾衰竭，甚至死亡。多次聚集性传播事件证明，新冠肺炎时期，隔离防治能有效的防止病毒传播。中国在抗疫期间，采取了坚决果断的措施对新冠肺炎病例和密切接触者进行隔离观察，多次检测和轨迹追踪，实践证明这一措施是积极有效的。

中国国内新冠肺炎疫情从2019年12月首次发现，到2020年2月达到疫情高峰，3月疫情得到控制，4月出现境外输入病例。由于新型冠状病毒潜伏期较长（1-14天），且病毒在传染过程中可能出现突变等，隔离，观察，追踪都是抗疫中的关键。全球范围内，新冠肺炎的大流行导致2020年日本东京奥运会延期，直接敬意损失约为60亿元；国际经济遭受重大影响，美股四次熔断，油价出现史无前例的负值；全球失业率创新高，毕业生就业也称得上“史上最难”。

目前中国国内抗疫工作主要集中于：检测并控制境外输入病例，对密切接触者的检测与观察，发现无症状感染者，疫苗与治疗方法研究。同时，各行各业逐步复工复学。

1.1.2 课题研究意义

全球疫情当下，防疫战不仅是医疗领域的任务，打的也是科技战。当前疫情防控到了一个新的关键时期，需要结合国家疫情防控的趋势与特点，尽快研发出用得上，有效果的科研成果，助力抗疫站。疫情攻关科研诞生了许多优秀的科研工作，例如通过大数据分析，知识图谱等人工智能技术，进行接触者排查，预测人流等，让疫情防控更加科学、精细、智能、可追溯。“无人配送车”、“一体化人体安检测温系统”、“自动消毒喷洒机器人”等实现抗疫期间的无人化。深圳北科瑞声公司研发的智能语音系统，已部署火神山、雷神山医院，帮助医护人员解决了穿戴厚重、手动操作医疗管理系统不便的难题。声扬科技采用声纹识别、语音识别等技术，助力防疫部门快速、准确、批量地完成受访对象个人信息跟踪搜集，率先实现健康状况及情绪状况的实时识别。

根据世卫专家组考察报告：新冠肺炎的症状是非特异性的，典型症状包括发热（87.9%），干咳（67.7%），乏力（38.1%）咳痰（33.4%），气短（18.6%），咽痛（13.9%）等[2]。可见咳嗽是新冠肺炎的典型症状之一，飞沫也是传播的主要途径。课题通过设计咳嗽自动定位检测系统，实时监测教室，实验室，办公室等公共场所的声音，自动定位咳嗽声源，检测咳嗽事件的发生，统计咳嗽频率等，监测潜在病例，为复工复学提供保障。系统可作为一种自动辅助诊断，解决复工复学情况下人员聚集空间中不易排查的问题，提供除发烧症状以外的早期新冠症状监测，减少监测过程中大量人力物力消耗，为呼吸系统疾病诊断与复工复学监控提供帮助。

咳嗽自动检测系统常应用于医疗康复，养老和辅助诊断。例如咳嗽检测仪用于病人康复时期，可自动计数咳嗽，计算频率等，为医生提供辅助诊断。新冠肺炎疫情下，感染可能会持续一段时间，甚至成为呼吸系统常见流行病。抗疫不能放松，更需要提高警惕，监控公共场所的潜在病例很有必要。本文以新冠肺炎的典型症状——咳嗽作为切入点，研发咳嗽自动定位与检测系统，积极响应疫情科研攻关号召，结合人工智能技术，麦克风阵列信号处理技术，研发咳嗽自动定位与检测系统，为新冠肺炎时期的抗疫工作提供新思路。

1.2 国内外研究现状分析

1.2.1 国内咳嗽检测研究现状

国内关于咳嗽自动检测的相关工作较少，论文[3]提取音频MFCC特征，通过SVM模型识别咳嗽事件，并比较隐马尔可夫(HMM)和动态时间规划(DTW)的检测结果和运行时间。实验样本来自医院病房，SVM和HMM较DTW的识别效果更佳，SVM运行效率高于HMM。论文[4]提取MFCC特征，并在预处理阶段进行小波去噪，为咳嗽和非咳嗽分别构建两个GMM模型，然后对GMM模型进行线性组合，得到最终的GMM概率模型，实验证明对模型进行线性组合提高了识别率。论文[5]中提出了一种简单的基于Mel倒谱系数的咳嗽识别方法，在病房中录音，对Mel倒谱系数本身的分布情况进行研究，采用三角滤波器组，设定阈值进行识别，快速简单，但阈值的选取较受限于诸多参数的影响。论文[6]中结合了HMM和ANN，分析了线性预测系数、Mel频谱倒谱系数（MFCC）和线性预测倒谱系数（LPCC），发现MFCC基于人耳特性，在咳嗽识别上更优。实验中利用HMM较好的时序建模能力和ANN强大的分类能力，研究了基于混合模型对咳嗽进行学习和识别的算法。

1.2.2 国外咳嗽检测研究现状

咳嗽自动检测的国外研究大多集中在两类：咳嗽监测设备和咳嗽自动检测算法。检测设备可用于观察医院病人康复情况，统计并检测咳嗽频率，咳嗽频率是判断康复的一个客观标准，为医生提供辅助判断依据。早期监测对咳嗽进行人工计数，但费时费力，自动检测咳嗽事件及频率是趋势，但目前还没有完全准确咳嗽自动检测系统。医学文献表明，需要一个客观可靠的工具衡量咳嗽的严重程度[7]。

研究[7]中将典型的咳嗽分为三个阶段，1）爆发性阶段：由于声门突然打开而引起的爆发；2）中间阶段：咳嗽声减弱；3）浊音阶段：由声带闭合引起。但咳嗽是因病而异也是因人而异的。咳嗽评估指标可以简单分为：咳嗽频率。咳嗽强度，咳嗽类型和声学特征。目前，咳嗽频率是评估中最有价值的指标，相关研究工作开展较多，技术较成熟。咳嗽检测也是一个活跃的研究领域，一些研究人员已经提出了从音频记录中识别咳嗽声音的方法。这些方法可以分为三大类：1）自动咳嗽检测和分割；2）对已经检测到的咳嗽进行自动分类；3）根据咳嗽的声音和类型诊断疾病。

早期的关于咳嗽检测的研究工作大多基于HMM，当下更多工作引入了SVM模型，尤其基于神经网络例如ANN，CNN的研究取得了不错的效果。

HACC（Hull Automatic Cough Counter, UK）通过麦克风24小时采集声音，构建ANN检测咳嗽事件，该系统可以自动标记咳嗽但仍需要人工计数[8]。LCM咳嗽检测仪（Leicester Cough Monitor，UK）通过自由场麦克风录音，系统采用基于HMM的KWS（keyword spotting）模型检测咳嗽事件的发生，然后，人工从这些咳嗽声音片段中选择一些来建立一个统计模型来适应当前的记录。最后，利用所设计的模型对整个录音过程进行处理，系统灵敏度91%，特异性99%，已经用于临床实验[9]。Crooks等人使用了由HACC和LCM组成的混合系统，测量COPD加重期恢复期咳嗽频率，总体灵敏度为57.9%，特异性为98.2%[10]。 Drugman等人设计了一个使用ANNs的声学系统，测试了10名健康受试者在不同情况下的自主咳嗽，敏感性和特异性约为95%[11]。在文献[12]中，Larson等人提出了一种使用低成本移动麦克风的精确、隐私保护咳嗽监测仪。他们使用主成分分析(PCA)和随机森林分类器对咳嗽声进行重建和分类，平均True-Positive为92%，False-Positive为0.5%。咳嗽检测设备，例如LCM咳嗽检测仪，特殊于专项任务，论文的工作[13]通过手机麦克风实时监测咳嗽，通过手机APP收集音频数据，训练数据通过SoundSnap网站下载了5个咳嗽样本，41个YouTube视频，构建CNN网络，识别灵敏度为92%，特异性99%。Tracey等人开发了一种用于咳嗽检测的算法，以监测结核病患者的恢复情况。他们从音频信号中提取MFCC特征，结合ANN和SVM分类器的组合实现算法，整体灵敏度为81%[14]。Swarnkar等人使用光谱特征，如共振峰频率、峰度和B-score结合MFCC特征进行咳嗽检测。输入到神经网络中，对于一个仅由3个受试者的342次咳嗽组成的测试数据集，敏感度为93%，特异性为94%[15]。Amrulloh等人，使用ANN分类开发了一种咳嗽检测器，该检测器使用非接触式记录系统，用于儿科病房，使用14名受试者的1400多个咳嗽声音，分别获得93%和98%的敏感性和特异性值[16]。Liu等人使用了GMCC特征，支持向量机分类器对4名受试者咳嗽903次，敏感度和特异度分别为91%和95%[17]。 Renard等人提出的百日咳自动检测算法[18]，数据集包括38例YouTube样本，20例正例（百日咳），18例其他类型咳嗽样本，基于百日咳的特点，构建逻辑回归模型，首先检测咳嗽事件，再检测是否为百日咳， 并针对百日咳中whooping的特点进行检测。该算法能够从所有音频记录中成功诊断出百日咳，而不存在任何错误诊断。它还可以自动检测个别咳嗽的声音，准确率为92%，PPV为97%。

在咳嗽检测算法的研究中，文献[19,20]提出咳嗽检测的四个挑战为：1）环境声去噪；2）与人说话，大笑，喷嚏的声音区分开；3）咳嗽个体差异；4）干咳或湿咳的分类:这是一个重要的医学指标。通过神经网络，SVM，贝叶斯分类器等检测咳嗽事件。一般音频处理的过程为：声音采集，静音检测，特征提取及特征选择，训练分类器 ，评估模型并测试。咳嗽信号的频率通常转化为感知频率，这样可以更好地模拟听觉处理文献。卷积神经网络（CNN）的高效识别能力使其受到广泛关注，机器学习中的分类器也用于检测咳嗽事件的发生，如支持向量机(SVM)、朴素贝叶斯分类器(Bayesian)、隐马尔可夫模型(HMM)、动态时间规正(DTW)等。MFCCs+SVM的应用更广泛，神经网络更具有准确识别的潜力[21,22]。

1.2.3 国外声音事件检测与定位研究现状

声音事件检测（Sound Event Detection, SED），是指在音频中检测说话，狗叫，汽车鸣笛等事件，并标注出事件的起止时间。声音事件检测与定位（Sound Event Detection, SELD）是指在音频中检测事件的同时，估计事件的空间位置。近年来，诞生了将声音事件检测与定位集成到一个神经网络模型中完成的研究工作，DCASE2019中的任务三发布了SELD的挑战，通过麦克风阵列和真实室内环境中录音，发布公开数据集。数据集中包括11个声音事件类别，每类20个样本。在5个不同的房间中，通过32通道球形麦克风阵列录音。距离阵列1米时，方位角（azimuth）分辨率10°，仰角(elevation)从-40°到40°每10°，共36×9=324 DoAs。距离阵列2米，方向角每10°，仰角从-20°到20°每10°，共36×5=180 DoAs。其中最多同时存在两个声音事件重叠。

比赛中发布的基线系统[23]中提出，基于DNN的方法被用来克服参数估计方法的一些缺陷，同时对混响和低信噪比的场景具有鲁棒性。并通过卷积循环神经网络（CRNN）同时预测声音事件与事件DoA，其中通过CNN过滤器提取特征，而RNN则对时序特征更敏感。论文中提出的SELDnet作为DCASE2019 Task3的基线系统，声音事件检测错误率0.35， F指数80%， DoA误差30.8°，帧召回率84%。比赛中基于基线系统的改进工作很多，Slawomir等人提出一种将任务分解的系统[23]，将SELD任务分解为：1）估计活跃声源数量；2）单声源时，估计DoA；3）两个声源活跃时，已知一声源的DoA，估计另一声源DoA；4）对声音事件分类。为每一个任务修改并训练单独的SELDnet，系统错误率0.08，F指数94.7%，DoA误差3.7，帧召回率96.8%。Jingyang Zhang等人提出了一个完整的流程，其中包括数据集扩充，网络预测和后处理阶段[24]。系统中通过SpecAugment扩充数据集，在特征提取阶段中指出，梅尔频谱系数可以提供与人类听觉相似的特征，利于SED任务，而DoA所需信息可以从幅值相位等特征得到。网络预测部分，分布对SED和DoA任务训练类CRNN网络。在后处理部分，论文提出了一种基于先验知识的正则化方法(PKR)，它计算出每个事件段的定位预测的平均值，并用这个平均值代替该事件的预测。最终系统错误率0.06，F指数96.7%，DoA误差22.4，帧召回率94.1%。

SELD可以是辅助侦听系统，场景信息可视化系统，沉浸式交互式媒体以及用于基于场景的服务部署的空间机器认知中的重要模块。一个直接的实际应用是识别并跟踪感兴趣的声源的机器人。DCASE近年的挑战任务表明，SED和DoA估计领域中，逐渐通过神经网络整合为一体，刚处于研究领域阶段，数据集较少，可达成效果有限。该领域未来必然会诞生更多优秀的有价值的研究工作，潜力无穷。

1.3本文主要研究内容

由于目前正处于新冠肺炎疫情期间，通过医院采集样本数据非常困难，本文通过Google AudioSet公开数据集以及YouTube音频作为数据样本。本文设计咳嗽自动定位与检测系统，实时监测室内咳嗽情况，可部署于教室，实验室，办公室，医院病房等公共场所。

**本文主要研究：**

1. 麦克风阵列实时捕获声源数据：采用七通道环形阵列麦克风部署在室内公共场所。
2. 静音检测：声源定位算法，咳嗽检测都有一定消耗，从低功耗的角度，首先检测是否为静音，确保后续步骤作用于非静音片段。
3. 声源定位与波束形成：深入理解两种声源位置估计算法（MUSIC算法，SRP-PHAT算法），并通过仿真测试不同算法的性能，适用条件。项目的背景是室内多声源的，则通过波束形成算法定向拾音，提高信噪比，为后续检测做准备。
4. 提取音频特征：采用三阶段音频特征提取的方法，提取34个短期特征，合成中期特征和长期特征。
5. 咳嗽检测模型：构建SVM二分类模型，将音频分类为咳嗽与非咳嗽。
6. 提取咳嗽事件音频：可在后续对咳嗽进行分类（例如干咳，湿咳等）。

本文将在第二章详述麦克风阵列信号处理，不同声源定位（DoA）算法研究，波束形成（Beamforming）算法研究，并进行仿真实验。在第三章详述本文使用数据集，音频特征提取，基于SVM模型进行咳嗽检测，并进行模型评价。第四章将描述集成的咳嗽自动定位与检测系统。

# 第二章 声源定位与波束形成算法理论基础

2.1麦克风阵列信号处理理论

与传统的信号传感器相比，传感器阵列具有较高的信号增益和较强的抗干扰能力。因此，传感器阵列在信号接收和参数估计方面具有更好的分辨率和性能。阵列信号的应用之一就是对位置的估计。

2.1.1 麦克风阵列模型

麦克风阵列形状大致可分为：线性阵列，圆形阵列，三维阵列等。复杂阵列原理都由线性阵列模型扩展而得，考虑由m个传感器构成的均匀线性阵列，如图2-1：

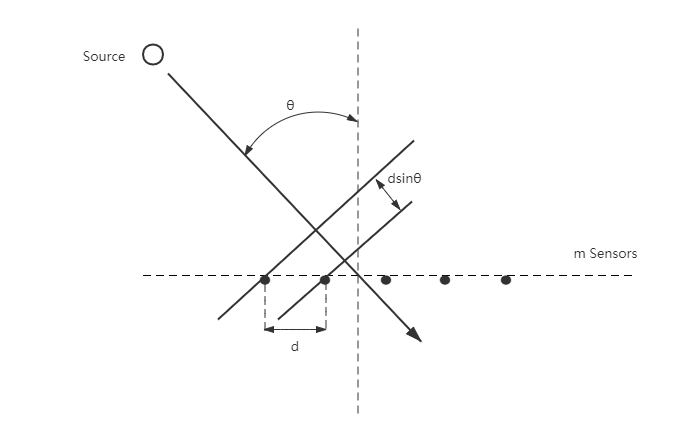


图2-1线性麦克风阵列模型

设d为传感器间距，为声源信号与阵列法线的逆时针夹角，c声音信号的传播速度，为以第一传感器为参考和平面波假设，可得第k个传感器与第一个传感器接收信号的时间差：

则阵列传播向量（方向向量）为：

通过公式(2-1)和公式(2-2)我们可以得到：

线性阵列的范围为[-90°, 90°]，如果声源位置关于阵列对称，则无法区分。每个麦克风接收到的数据与第一个麦克风之间存在延迟，其中为干扰信号：

设为波长，频率为，可得：

并定义：

根据公式(2-5)和(2-6)代入(2-3)，可得Vandermonde向量，它类似于由正弦信号：

声源位置估计和波束形成的理论都基于麦克风阵列模型，例如可以通过计算麦克风之间的时间差估计声源位置，设计滤波器提取感兴趣的声源，同时抑制噪声。

2.2声源位置估计(DoA)算法

2.2.1 DoA算法简介

阵列信号处理的一个基本问题是空间信号的到达方向（Direction of Arrival）估计问题，是雷达、声呐等许多领域的重要任务之一。经典的DoA估计方法括：基于波束形成（beamforming）的谱估计算法，Capon算法 （又称最小方差法），基于空间谱估计的MUSIC算法（MUltiple SIgnal Classiﬁcation）。本文主要研究基于波束形成的DoA估计算法：SRP-PHAT（Steered-Response Power Phase Transform）和基于子空间方法的MUSIC算法。

2.2.2 基于特征分解的子空间算法—MUSIC

MUSIC算法的提出在DoA估计理论发展史上是一次质的飞跃。其核心思想为：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。是一种超分辨率的方法，后续有不少MUSIC优化算法，例如Barabell提出的Root-MUSIC算法，减小了计算量，陆续出现改进算法，子空间迭代的快速算法及去相关空间的平滑技术等。

MUSIC算法原理：

1. 阵列接收数据模型，假设阵列通道数量为，共个远场信号源，功率相同且互不相干，s为个输入信号源，是个通道的输出信号，为个通道的噪声向量*，*为的Vandermonde矩阵：

2. 阵列输出的协方差矩阵为：

其中是的信号自相关矩阵，I为的单位矩阵，是噪声方差。为Hermite矩阵，其特征向量具有正交性，对其进行特征值分解，可得到共个正实特征值，其中D个较大的特征值对应信号，M-D个较小的特征值对应噪声，那么相应的特征向量也就相应构成信号空间和噪声空间。则根据特征值构造信号空间矩阵，的各列与信号源方向对应，噪声矩阵。

3. 则可定义空间谱函数：

根据变化计算谱函数，寻找峰值来估计声源方向（DoA）

2.2.3 基于波束形成(beamforming)的方法—SRP-PHAT

SRP-PHAT（Steered-Response Power Phase Transform**）**是一种结合波束形成和TDOA的DoA估计算法，鲁棒性强，在混响、多声源环境中表现良好。对空间不同方向的声音进行增强，得到声音信号最强的方向即为声源方向。

SRP可由滤波求和波束形成器（filter and sum beamformer）引入，在频域FS的输出功率表示为：

在音频处理中通常在短时频域进行计算，故表达式中添加。其中表示自适应滤波器，每个麦克风通道有差异，表示声源信号的频域形式，表示时延。

那么：

表示t时刻的SRP值，在FS固定时，SRP值只与波达方向有关，将公式(2-12)带入公式(2-13)，可得：

可以发现SRP是所有麦克风两两组合的GCC（广义互相关）之和：

PHAT加权相当于白化滤波，可以锐化GCC，定义为：

通道滤波器与加权函数的关系为：

2.3波束形成(beamforming)算法

2.3.1 波束形成算法简介

波束形成时阵列信号处理的标志任务之一，主要思想是：通过将各阵元输出进行加权求和，增强期望信号，抑制干扰，将阵列的接收方向增益聚集到一个方向，形成“波束”。波束形成算法大致可分为三类：固定波束形成，自适应波束形成和后置滤波算法。

固定波束形成，适用于稳定不变的噪声干扰环境，可以对某一方向的声源进行抑制，抑制的强度不变。当环境中的噪声信号发生变化时，固定波束形成的算法并不能跟随着做出调整，鲁棒性差，但其运算复杂度低，算法的可实现性强。自适应波束形成的方法，可以利用信号的输出来自适应地调整滤波的权重系数，其抑制性能可以对环境信号的变化做出调整，鲁棒性更好。但是会残留一部分噪声，而且对某些噪声的抑制能力不强，所以就有了把单通道的方法和麦克风阵列相结合的方法，即后置滤波算法。

2.3.2 延时-求和（Delay and Sum）

延时-求和是最早最经典的固定波束形成算法，如图2-2，首先**计算不同麦克风通道之间的时延，进行时延补偿，将信号对齐，然后根据不同权重约束叠加信号取均值，最后输出一路信号。**在Flanagan最初的DS算法中，每路信号权重相同，因此抑制噪声能力较差。

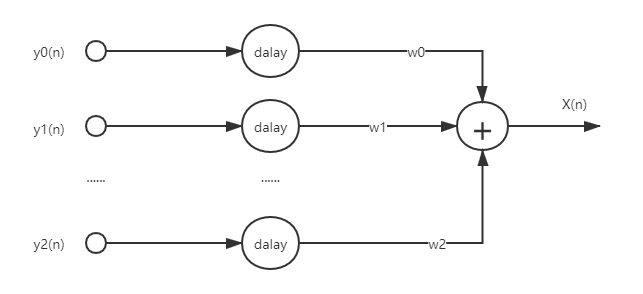


图2-2 延时-求和滤波器

2.3.3 MVDR

MVDR（Minimum Variance Distortionless Response）是非常广泛的波束形成方法，它通过最小化残留噪声的能量，同时约束期望方向的信号不失真来实现。在频域上阵列信号模型为：

其中定义为声源到达麦克风的信号，为衰减因子，相对时延为，s为声源信号，表示噪声信号。则：

并令：

其中被称为方向向量，那么公式(2-20)可以写为：

波束形成的核心在于设计滤波器，对频域上的每个点，完成到的映射，即：

**滤波器的功能很明显，增强的信号（desired speech），抑制信号（这部分被称为残留噪声，residual noise）。**

再由功率谱密度（PSD）的定义，对于复数域的向量：

假设噪声和原始信号之间不相关，有：

MVDR的求解问题可以表示为：

约束条件可以保证MVDR滤波器再满足的方向或者条件下，增强的信号不是真，根据拉格朗日乘子法得到：

2.4声源定位与波束形成仿真

本文基于开源库pyRoomacoustics进行室内仿真，人为放置声源，设置信噪比等参数，模拟七麦克风阵列采集房间脉冲响应（RIR），测试DoA算法及波束形成算法。

2.4.1 DoA仿真

基于库pyRoomacoustics的DoA估计仿真流程如图2-3，首先1）自定义室内环境（房间大小，墙面反射，墙面吸收等）；2）在房间内添加声源和麦克风阵列，可人工设置声源和麦克风阵列位置，相对角度等参数；3）模拟真实环境，通过麦克风阵列采集房间脉冲响应（RIR）；4）对信号进行分帧短时傅里叶变换，然后进行DoA估计（本文实验MUSIC算法和SRP-PHAT算法）：

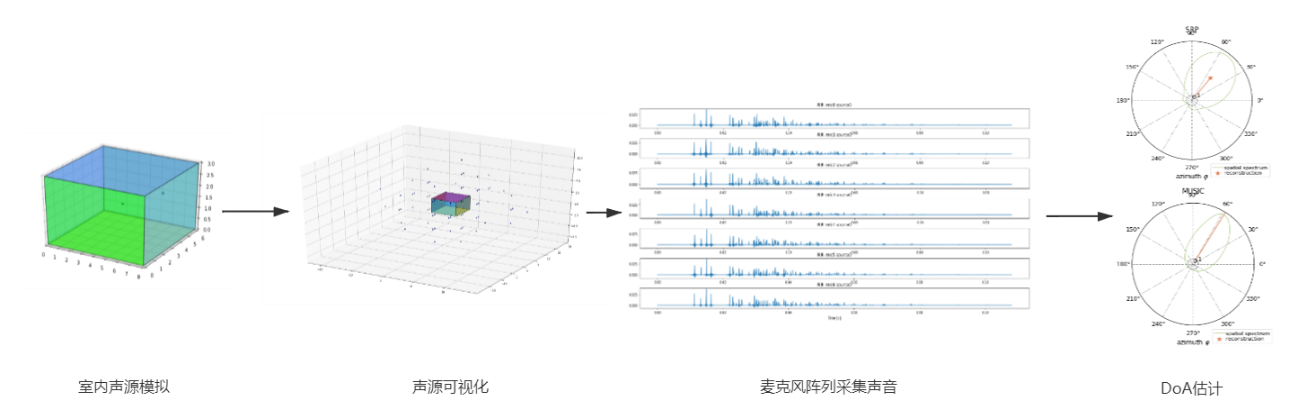


图2-3 DoA估计仿真流程

由于使用七麦克风环形阵列为二维阵列，对z轴方向感知能力有限，故进行了二维仿真，其中仿真实验中涉及变量详见表2-1。

表2-1变量定义及取值

|  |  |
| --- | --- |
| 参数 | 定义 |
| 房间尺寸(size) | 二维室内房间定义长、宽。三维室内房间定义长、宽和高。也可以自定义房间形状，设定角落坐标。 |
| 墙面最大反射次数(max\_order) | ISM允许最大反射次数 |
| 墙面吸收率(absorption) | 即反射波每击中一面墙就×(1 –吸收率) |
| 采样率(fs) | 麦克风阵列采样声源的频率 |
| 信噪比(SNR) | 即输出信号的功率，与噪声功率的比值，单位dB。 |
| 方位角(azimth) | 声源相对麦克风阵列的二维平面角度，波达方向角，DoA算法估计输出值。 |
| 距离(distance) | 声源相对麦克风阵列的距离。 |
| 声源数(num\_src) | 同时存在声源数 |

* **二维仿真**

环境参数：size = 10m×10m，max\_order = 2，absorption=0.1

1. 单声源DoA估计(num\_src=1)：MUSIC算法于SRP算法估计精度都较高（error = ±2°）
2. 多声源DoA估计：

a) **DoA角度对于多声源位置估计的影响：**num\_src = 2，SNR=30dB，固定一声源位置（距离为3m，DoA为0°），另一声源（距离为3m，DoA，实验结果如图2-4：

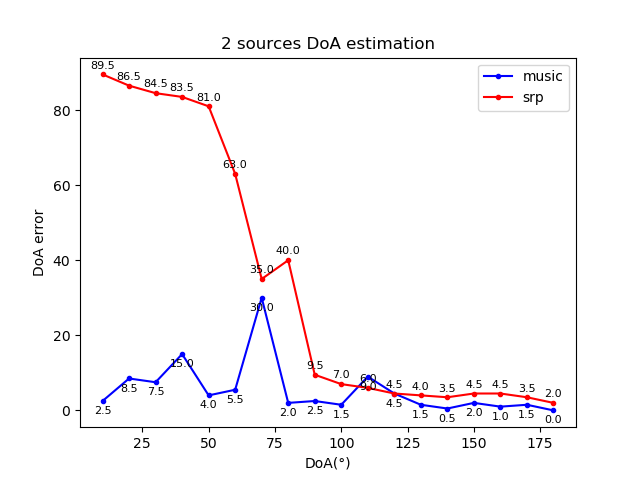
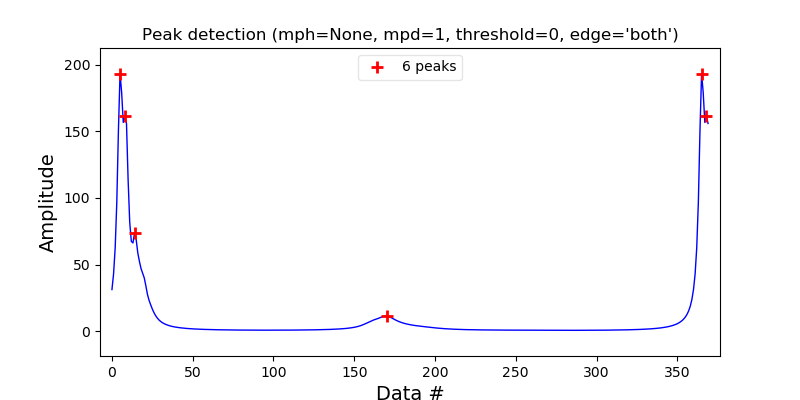
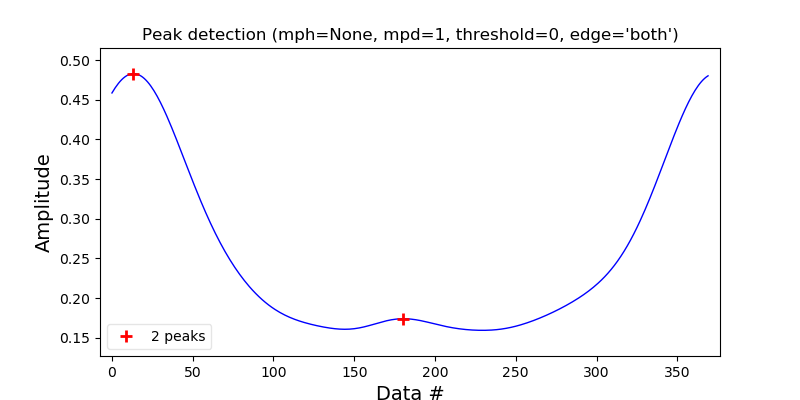
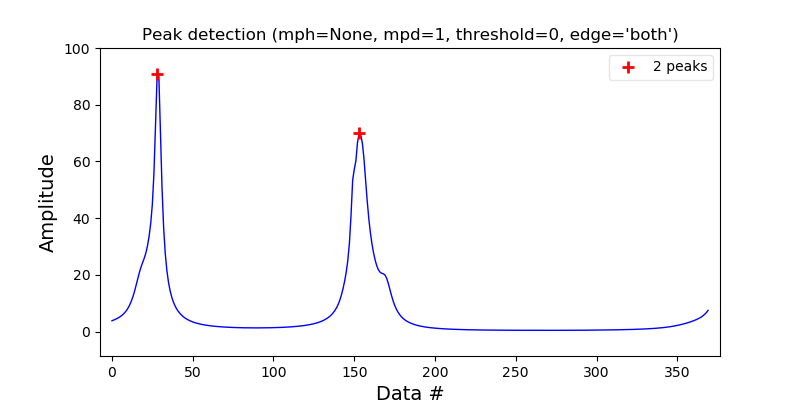
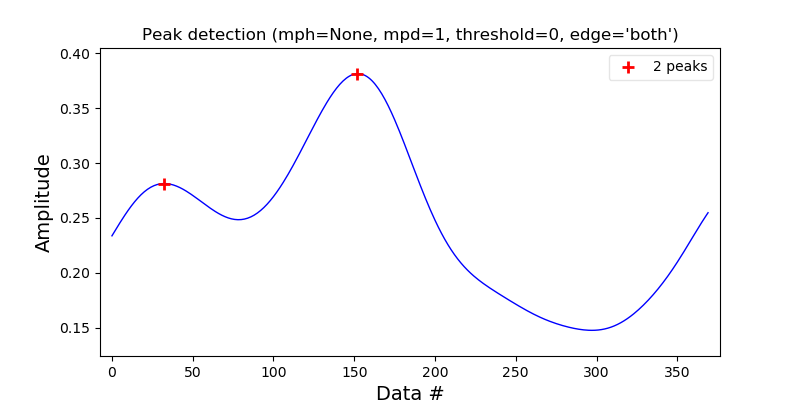


图2-4 DoA在多声源定位中的影响

**实验中发现，SRP-PHAT算法在两声源DoA相近时，误差较大，而MUSIC算法表现良好。**进而对比网格化搜索峰值时的情况，如图2-5 a)，MUSIC算法在空间中不同位置的谱值尖锐，形成峰值明显；如图2-5 b)，两声源位置相近时SRP值似乎较平滑，合并为一个峰值，则导致DoA估计误差很大，分辨率不够的问题。在实验中尝试提高网络搜索分辨率，但峰值仍然不明显。作为对比，图2-5 c) 和d)是两声源位置分离情况下的峰值，在MUSIC和SRP中两波峰都较明显。

a) MUSIC峰值(两声源位置相近) b) SRP峰值(两声源位置相近)

c) MUSIC峰值(两声源位置分离) d) SRP峰值(两声源位置分离)

图2-5网格搜索峰值图

b) **SNR对于多声源位置估计的影响：**num\_src=2,

表2-2 SNR实验中的参数设置

|  |  |
| --- | --- |
| 参数 | 取值 |
| SNR(dB) | [-20,30] |
| DoA1(°) | 30 |
| Distance1 | 3 |
| DoA2(°) | 150 |
| Distance2 | 3 |

由于声源位置相近会使DoA影响SRP的估计结果，故设置两声源位置较分离，探究SNR对DoA估计的影响，实验结果如图2-6，可见在低信噪比环境下DoA估计误差大，噪声严重影响声源位置估计：

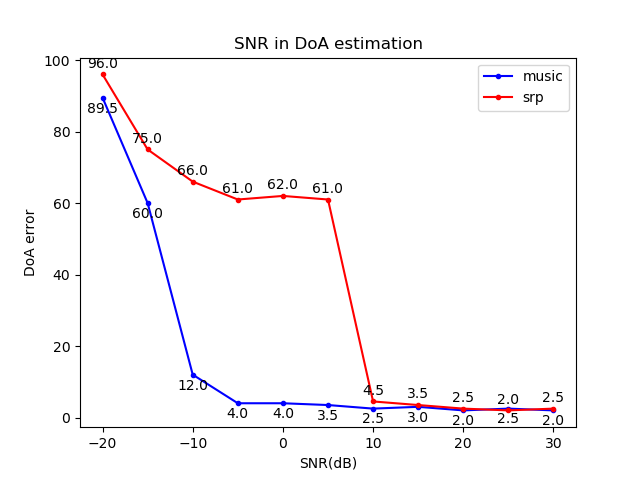


图2-6 SNR对DoA估计的影响

由多声源估计结果可以观察到：1）提高信噪比有利于声源位置估计。2）在声源位置相近，要求多声源估计分辨率高时，MUSIC算法具有超分辨率的优势，峰值明显，而 SRP-PHAT峰值较平滑。

2.4.2波束形成仿真

波束形成仿真流程如图2-7，首先仿真室内房间环境，放置麦克风阵列，感兴趣的声源，干扰声源的位置，然后通过波束形成，将多通道数据滤波后合并为单通道。

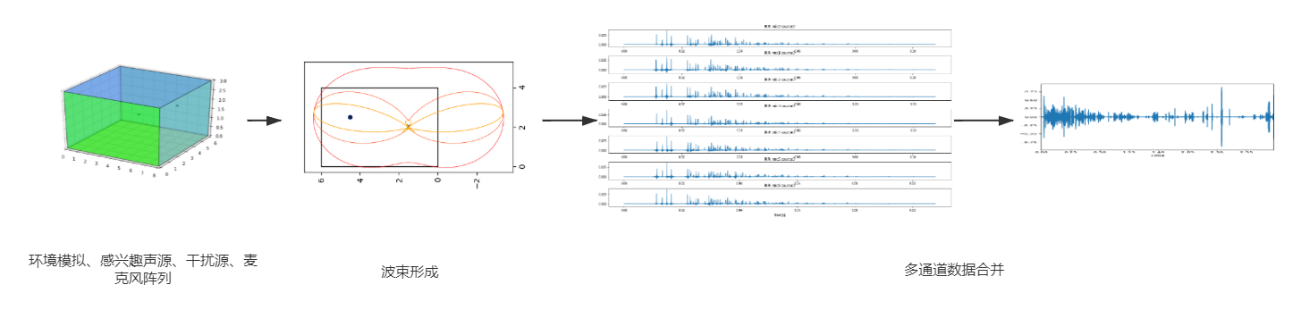


图2-7 波束形成仿真实验流程

首先，定义评价指标——信噪比（SNR）。定义为第j个声源信号，定义第i个麦克风接受的信号，为源到麦克风的滤波器，为噪声，则有：

对于通过波束形成滤波器处理后得到的信号，可以定义并分解为：

其中表示感兴趣声源，表示干扰，表示噪声，则SNR可以定义为：

由于固定波束形成鲁棒性较差，基于自适应波束形成算法MVDR进行仿真实验。仿真环境参数：Size=20m×20m，max\_order = 2，absorption=0.1。

感兴趣声源（咳嗽）DoA = 0°，相对麦克风阵列距离为3m/5m/10m；干扰声源（说话）DoA从30°至180°，分辨率为30°，相对阵列距离从1m到10m，分辨率1m。

固定感兴趣声源的DoA和距离，干扰源在不同距离下的SNR值为其在该距离下，不同角度的SNR值的均值：

由于波束形成的作用为提高信噪比，则在实验中，定义信噪比增益（SNR Gain）为滤波器处理后的SNR，相对处理前SNR的增加值。

如图2-8，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，定向拾音，便于咳嗽识别。

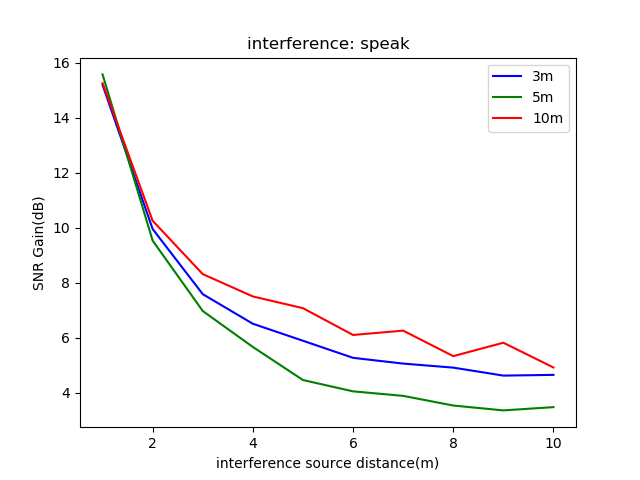


图2-8 不同距离信噪比增益对比（干扰源为说话）

将干扰声源设置为白噪声，重复仿真实验，得到信噪比增益结构，如图2-9，可以发现，干扰源为白噪声，信噪比增益较干扰源是说话的情况更高。

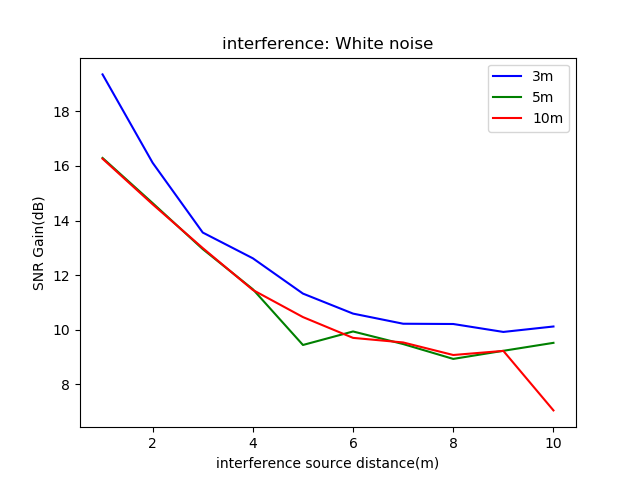


图2-9 不同距离信噪比增益对比（干扰源为白噪声）

2.5 本章小结

本章首先对麦克风阵列模型进行建模，然后介绍声源位置估计（DoA）和波束形成（Beamforming）两个阵列信号处理中的基本任务。详述推导了DoA估计算法中经典的MUSIC算法和SRP-PHAT算法，波束形成中简单的延时-求和滤波器和MVDR滤波器。深入理解阵列信号处理的原理，并在仿真环境中对DoA估计算法和波束形成进行实验，验证声源位置，环境参数等对算法的影响。

# 第三章 基于SVM的咳嗽检测方法

3.1数据来源

由于新型冠状病毒疫情期间，无法通过医院收集咳嗽音频数据，项目主要通过Google AudioSet与YouTube网络视频中的咳嗽音频作为数据集。

* Google AudioSet：

Google在2017年发布的大规模音频数据集，包括632个音频事件类，数据集分为三部分：平衡的评估集，平衡的训练集，不平衡的训练集（注：在平衡数据集重，每个类的样本数量较平均）。数据集提供提取好的特征，通过Google VGGish模型以1Hz提取的128维特征。数据集通过Tensorflow的官方数据格式(.tfrecord)。提取AudioSet中关于咳嗽，以及教室办公场所常见声音的数据（例如，说话，电话铃声，笑等）。

在音频领域，强标注是一件很麻烦的事情，声音事件检测（SED）长期没有足够的训练数据。更多的是弱标注的训练数据，其中弱标注数据分为两类[10]：

1. 有无标注：只标注一段录音中有哪些事件，没有任何关于事件的起止信息。

2. 顺序标注：标注一段录音中事件发生的顺序，但是不知道事件的起止时间，事件之间可能存在重叠。

顺序标注较有无标注更强一些，AudioSet数据集提供的是有无标注类型的弱标注数据。实验过程中发现由于只需要训练部分类别（教室办公室常见声音类别）的数据，样本量大约在1000例以下。若采用AudioSet的128维特征数据，通过深度学习训练神经网络，样本数不足以支持参数估计，故通过爬虫抓取AudioSet的音频数据，切割再处理。

* YouTube视频：

AudioSet仅提供咳嗽评估集60个样本，平衡训练集60个样本。咳嗽数据样本较少，通过在YouTube视频上获取咳嗽音频，人工标注标签扩充数据集。

* 本文系统所用数据集：

通过部分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）。

3.2 音频采集及特征提取

3.2.1 音频采集

在系统实际部署时，通过麦克风阵列采集教室，办公室等公共场所的声音数据，阵列信号处理是对天线接收到的信号进行处理（增强，降噪），并对某些信号参数进行估计。目前采用的麦克风阵列结构为环形阵列(半径=42.5mm)，如图3-1：

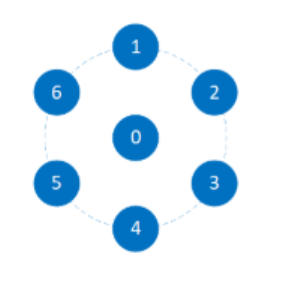


图3-1麦克风阵列结构示意图

3.2.2 音频特征提取

Audioset提供音频时长为10s，YouTube音频时长不等，通过人工截取咳嗽部分。故本文采取三阶段音频特征提取，先提取短期特征（short-term windows），中期帧特征（mid-term windows），最后计算长期特征(long-term windows)。

在声音预处理阶段，通过滤波器消除静音阶段，这确保后续的音频处理只在有声音的信号上进行，也有助于减少算法的运行时间[11]。

在短期特征提取阶段，首先将输入信号分割为帧（frame），音频处理通常在短时频域上，短时傅里叶变换（STFT）要求输入时域信号平稳的且帧足够短，故选取帧长50ms，帧移25ms（50%重叠）。提取每一帧的特征，实验中共提取了34个音频特征，如表3-1，其中包括梅尔倒谱系数MFCC（前13维），过零率等常见音频特征。由于人耳对不同频率的声波有不同的听觉敏感度，从200Hz到5000Hz的语音信号影响比较大，故通过提取MFCC系数增强低频，抑制高频。

表3-1模型评价指标

|  |  |  |
| --- | --- | --- |
| 特征ID | 特征名称 | 描述 |
| 1. 1 | 过零率(Zero Crossing Rate) | 在特定帧的持续时间内信号变化率 |
| 2 | 能量(Energy) | 信号值的平方和，按各帧长度归一化。 |

表3-1（续表）

|  |  |  |
| --- | --- | --- |
| 特征ID | 特征名称 | 描述 |
| 3 | 能量熵(Entropy of Energy) | 子帧的归一化能量的熵。 |
| 4 | 频谱质心(Spectral Centroid) | 频谱一阶距。 |
| 5 | 频谱延展度(Spectral Spread) | 频谱二阶中心矩。 |
| 6 | 频谱熵(Spectral Entropy) | 反应了每一帧信号的均匀程度。 |
| 7 | 频谱通量(Spectral Flux) | 描述相邻帧频谱的变化情况。 |
| 8 | 频谱滚降点(Spectral Rolloff) | 频谱中90%的幅值分布集中的频率。 |
| 9-21 | MFCC | Mel频率倒谱系数形成倒谱表示，其中频带不是线性的，而是按照Mel尺度分布的。 |
| 22-33 | 音阶参数(Chroma Vector) | 12个参数对应12级音阶。 |
| 34 | 音阶参数的标准方差(Chroma Deviation) | 略 |

在中期帧特征提取阶段，对每帧短期特征计算每个特征的均值和标准差。

在长期特征提取阶段，计算中长度帧特征的均值和标准差，最后得到整段音频的特征向量。

3.3 基于SVM二分类的咳嗽检测

3.3.1 SVM理论基础

支持向量机（Support Vector Machine）是常见的二分类模型，通过求解一个超平面对数据特征进行分割，原则是最大化间隔函数，如图3-2。SVM可分为三类：

* 样本线性可分，最大化硬间隔函数，训练线性可分SVM；
* 样本近似线性可分，最大化软间隔函数，训练线性SVM；
* 样本线性不可分，通过核函数和最大化软间隔函数，训练非线性SVM；

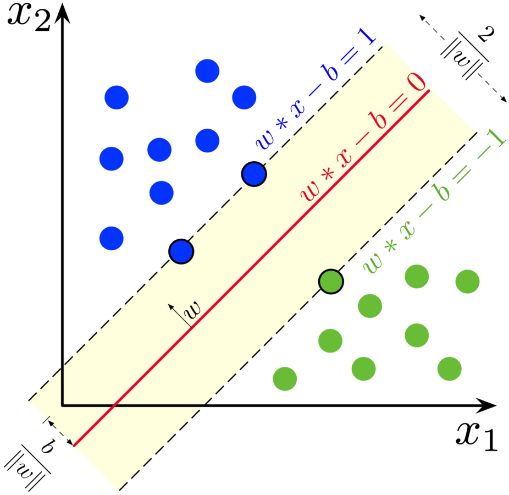


图3-2 二维SVM超平面

实际中大部分数据是线性不可分的，则可以引入核函数，将数据映射到高维空间，解决线性不可分问题，同时解决了映射造成的维度爆炸的问题。对于线性不可分问题，假设超平面：

则有最小化函数：

转化为对偶问题求解：

若直接求解公式(3-3)，先将样本和映射到高维空间中，再做乘积和加和运算，由于特征空间维度可能很高，甚至为无穷维，引发维度爆炸得问题。可以通过核函数巧妙地解决这个问题，定义核函数为：

则公式(3-3)可以重写为：

代入超平面公式求解可得：

核函数技巧在低维空间中进行计算，避免了将特征显示地映射到高维空间后的结果，但将分类效果表现在高维上。通常根据实际情况选择不同核函数和参数，例如：

* 线性核：
* 多项式核（d表示多项式次数）：
* 高斯核：

高斯核可以将原始空间映射到无穷维空间进行计算，通过调节参数，具有比较高的灵活性，是广泛使用的核函数之一。

3.3.1 SVM模型进行咳嗽检测

本文基于scikit-learn库，训练SVM模型，实验SVM\_linear模型和SVM\_RBF模型。

测试不同惩罚系数（0.001, 0.01, 0.5, 1.0, 5.0, 10.0, 20.0），惩罚系数C为SVM软边距，表示对误差的宽容度，C越大，拟合非线性的能力越强。gama参数RBF核函数的参数，。gama参数越大，高斯分布方差越小，则容易过拟合，对未知样本的分类能力下降；gama参数越小，则平滑效应太大，分类器准确率下降，分辨率低。

3.4 模型验证与评价

3.4.1验证集评价

分别构建SVM\_linear模型和SVM\_RBF模型，进行10折交叉验证，实验所设计模型评价指标如表3-2：

表3-2模型评价指标

|  |  |  |
| --- | --- | --- |
| 评价指标 | 定义 | 公式 |
| 1. 精确度(Precision) | 表示预测为正例，实际为正例的样本比例。 |  |
| 召回率(Recall) | 表示实际为正例，预测正确的样本比例。 |  |
| F1指数 | 综合考虑精确度与召回率的评价指标。 |  |
| 准确率(Accuracy) | 预测正确的样本占总样本的比例。 |  |
| 灵敏度(Sensitivity) | 表示所有正例中，预测正确的比例，衡量给了分类器对正例的识别能力。 |  |
| 特异性(Specificity) | 表示所有负例中，预测正确的比例，衡量给了分类器对负例的识别能力。 |  |
| ROC曲线 | 以FPR为横轴，TPR为纵轴形成的曲线。 |  |
| AUC | ROC与坐标轴所围面积大小。 |  |

SVM\_linear模型评价如图3-3，其中最优参数C为0.01，整体模型F1指数为92.7%：

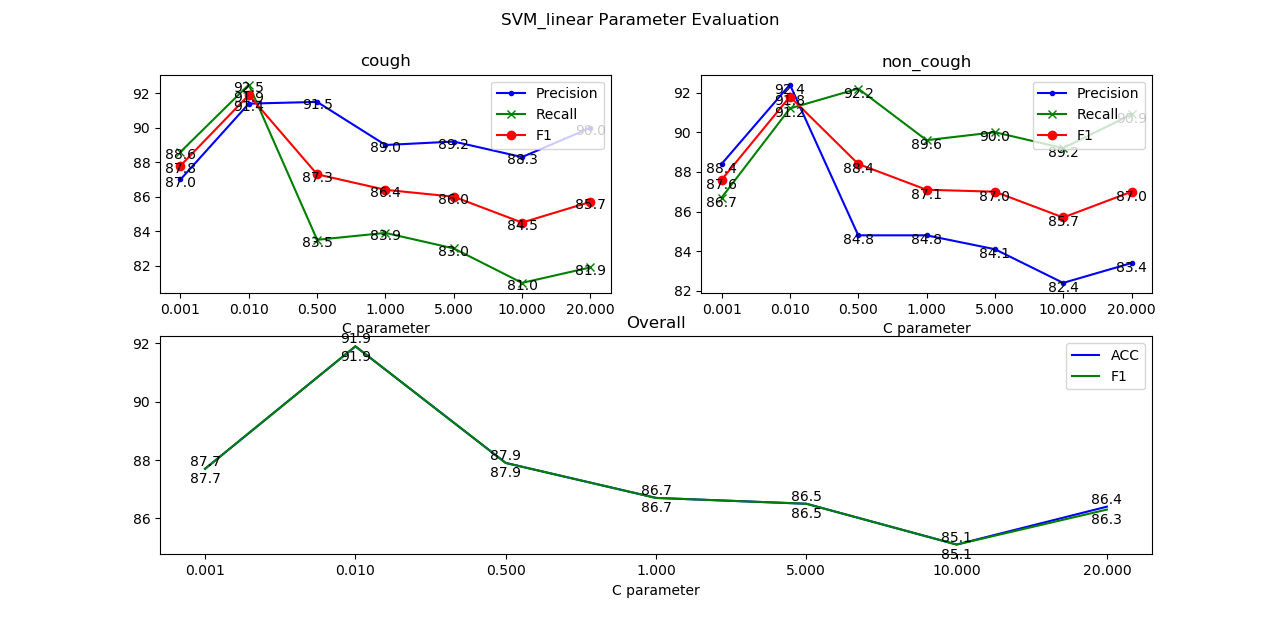


图3-3 线性SVM模型评价

SVM\_RBF模型评价如图3-4，其中最优参数C为1.0，整体模型F1指数为95.3%：

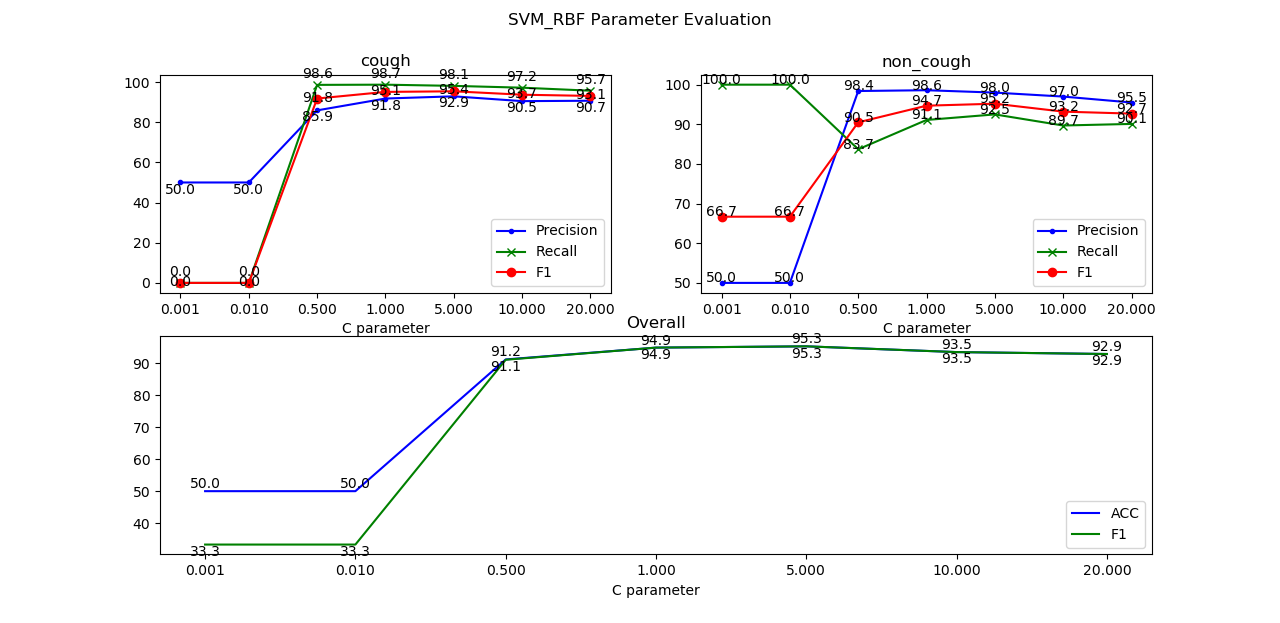


图3-4 SVM\_RBF模型评价

对比SVM\_linear模型和SVM\_RBF模型，如图3-5，SVM\_RBF模型精度更高，分类能力更好，在集成完整系统时，采用SVM\_RBF模型。

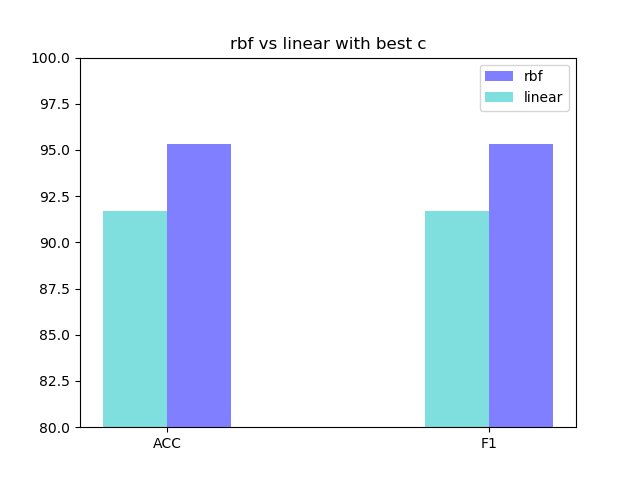


图3-5线性SVM与SVM\_RBF对比

3.4.2测试集评价

基于AudioSet的咳嗽，说话，音乐等音频数据集，构成测试集，包含咳嗽样本46个，非咳嗽样本46个，混淆矩阵如图3-6和表3-3：

表3-3测试集混淆矩阵

|  |  |  |
| --- | --- | --- |
|  | 预测Positive | 预测Negative |
| 1. 实际Positive | TP = 42 | FP = 8 |
| 实际Negative | FN = 4 | TN = 38 |

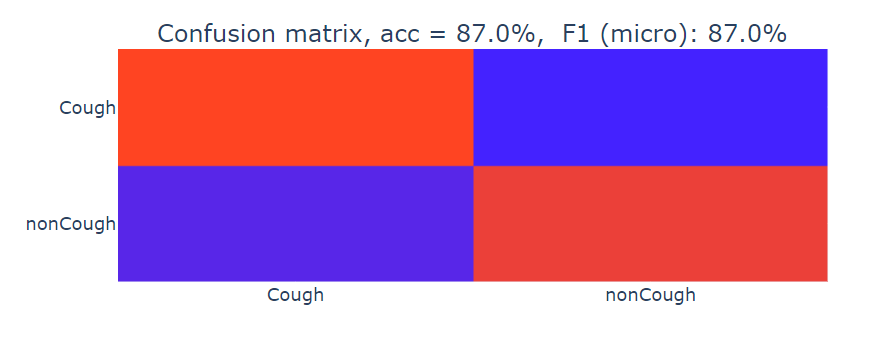


图3-6测试集混淆矩阵

其中测试集各评价指标如图3-7，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%。灵敏度越高表示漏检率越低，特异性越高误诊率越低，咳嗽检测系统对错判宽容度更高，宁可错判不能漏判，故对灵敏度要求更高。总体模型在测试集上表现良好。

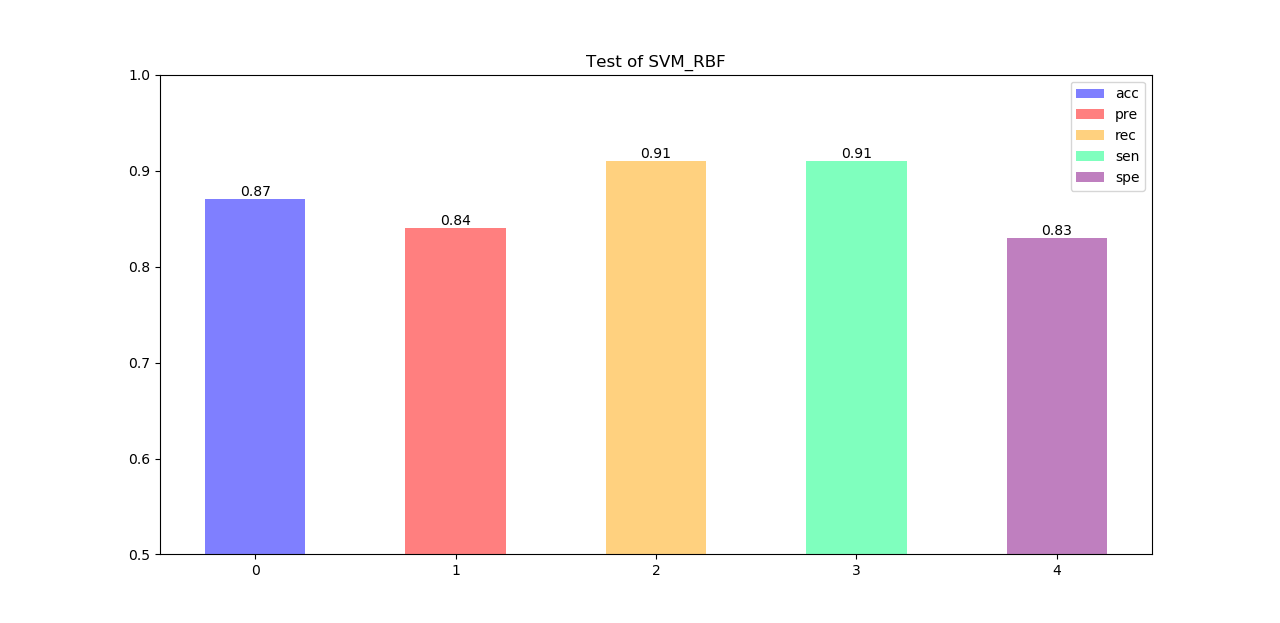


图3-7测试集评价指标

ROC曲线如图3-8，ROC陡峭，AUC值为0.9，分类器效果较好：

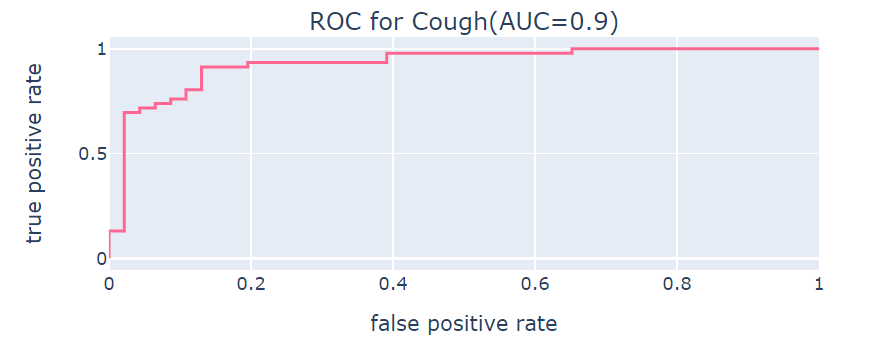


图3-8测试集ROC曲线

3.5本章小结

首先，本章在3.1节详细描述了系统所使用的数据集，选取AudioSet和YouTube咳嗽音频，包含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）。

在3.2节描述了音频采集所用的麦克风阵列结构，然后描述了系统所用的三阶段特征提取模式（短期特征，中期特征，长期特征），包括所提取音频特征。

在3.3节描述SVM模型构建，首先从理论基础推导了SVM模型中核函数的作用与意义，并描述了几种常用核函数。

基于对数据集进行10折交叉验证，对比SVM\_linear模型和SVM\_RBF模型，并进行参数调优，其中SVM\_RBF最优参数C=1，gamma为auto模型时，模型F1指数为95.2%；SVM\_linear模型最优参数C=0.01，F1指数为92.7%。实验结果表明，RBF核函数能更好的拟合非线性数据，SVM\_RBF模型性能优于SVM\_linear模型。

SVM\_RBF模型基于测试集的各评价指标为：准确率87%,精确度84%，召回率91%，灵敏度91%，特异性83%，AUC值为0.9，可表示模型分类性能良好，灵敏度较高，适用于咳嗽自动检测系统。

综上，本章详细阐述了系统中所用数据集，音频采集和特征提取，基于SVM模型进行二分类，检测咳嗽事件的相关工作。

# 第四章 咳嗽自动定位与检测系统集成

4.1 实时系统流程

本文研发系统与智能音箱工作原理与流程有相似的部分，故集成系统前对智能音箱的实时系统流程，应用到本文的系统中如图4-1：



图4-1实时系统流程方案

1. 麦克风阵列实时采集：

一次咳嗽的平均时长大约为2s，系统中采用长度为4s的窗口，如图4-2，为实时可视化波形图，音频以流的方式从右侧流入，系统设定每隔2s对窗口内音频进行后续处理。

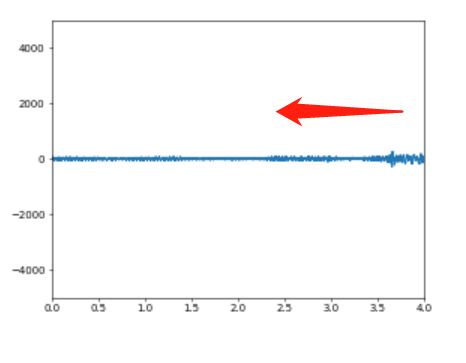


图4-2实时音频流处理

1. 静音检测：

通过简单的幅度阈值检测，判断是否为静音。确保后续的算法作用在非静音片段，降低功耗。

1. 声源定位：

根据基于机器人麦克风阵列的声源定位系统ODAS，论文对比了SEVD-MUSIC算法和SRP-PHAT算法，将声源定位（SSL）任务划分为两个阶段：1）TDOA估计；2）网格化DoA搜索。SEVD-MUSIC与SRP的区别在于第一个子任务，SRP通过GCC-PHAT进行TDOA估计，而SEVD-MUSIC通过奇异特征值分解。并且详细证明了SRP的复杂度更低，较适合在线实时处理[25]。

故在本论文中采用SRP-PHAT算法进行DoA估计，设定活跃声源数量，输出DoA估计值。

1. 波束形成：

在多声源的情况下，对不同声源位置形成波束，采用MVDR滤波，定向提取每个声源，提高信噪比。

1. 咳嗽检测：

对波束形成滤波器处理后的音频进行咳嗽检测，通过SVM模型预测为咳嗽事件的概率。

4.2 实时监测系统集成

通过多线程与异步回调实现实时监测的功能，即一线程实时采集音频，另一线程每间隔2s对窗口（4s）内音频进行后续处理。

在本地控制台日志可观察到静音检测，声源位置估计，咳嗽检测结果如图4-3；通过html和javaScript简易可视化音频波形与DoA估计情况，如图4-4：

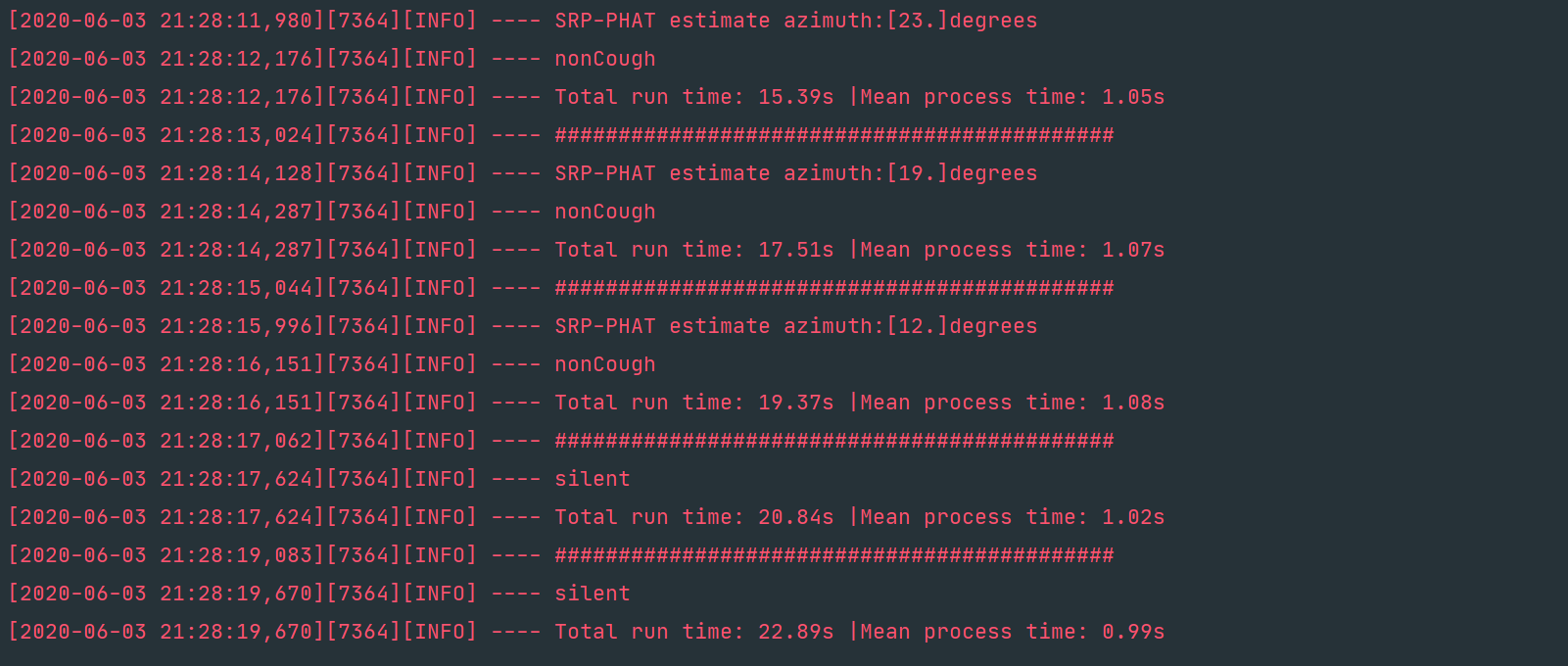


图4-3实时咳嗽检测控制台日志

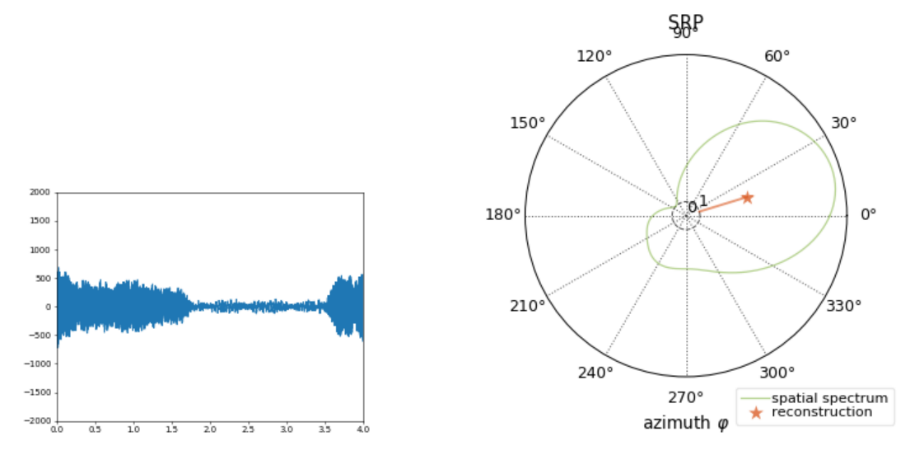
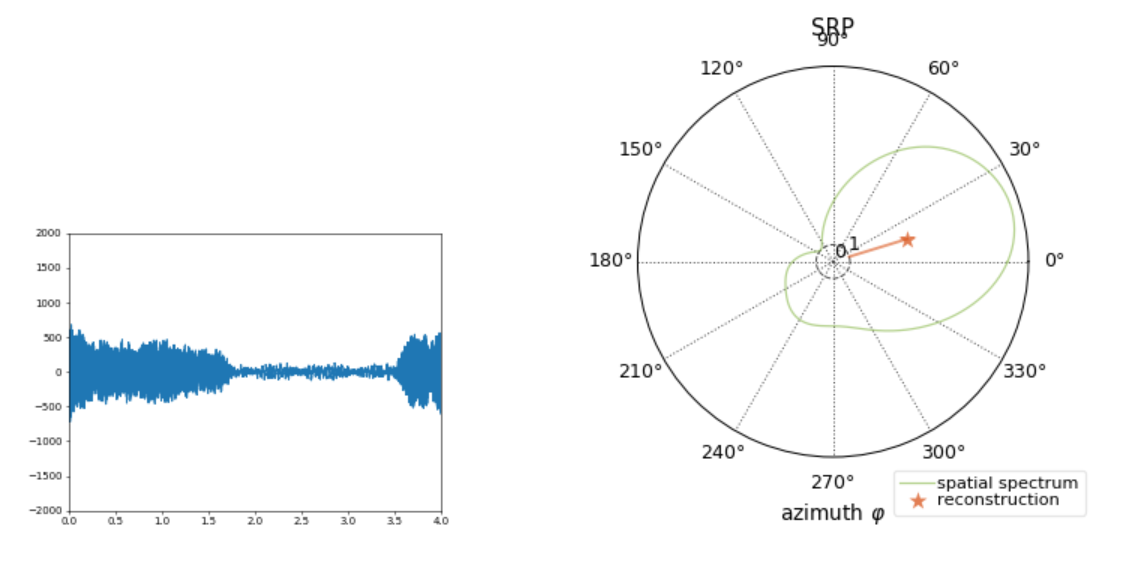


图4-4 波形与DoA估计可视化

并进行了代码块运行时间测试，系统配置如表4-1。

表4-1 测试系统配置参数

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

假设代码块共运行N次，则定义评价指标为，如公式(4-1)，表示系统中代码块总运行时间与代码块运行次数的比值：

系统设定为每间隔2s进行一次检测，由于算法流程中涉及到文件IO读写，绘图等耗时操作，在间隔为2s的情况下可以实时工作，结果如表4-2：

表4-2 系统实时性能测试

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 系统总运行时间 | 测试1 *(s)* | 测试2 *(s)* | 测试3 *(s)* | 均值 | 标准差 |
| 1min | 0.84 | 1.14 | 1.15 | 1.04 | 0.0207 |
| 10min | 0.96 | 1.45 | 0.83 | 1.08 | 0.0712 |

表4-2（续表）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 系统总运行时间 | 测试1 *(s)* | 测试2 *(s)* | 测试3 *(s)* | 均值 | 标准差 |
| 30min | 0.91 | 1.42 | 0.88 | 1.07 | 0.0614 |
| 60min | 0.87 | 1.34 | 0.91 | 1.04 | 0.0453 |

由表4-2三次测试可观察到，测试2较高，测试2中一直是非静音状态，而测试1与测试3静音与非静音时间较平均。若为静音片段，则算法后续步骤将不执行，可降低功耗。于是，对静音片段和非静音片段，不同声源数量，进行算法的性能测试，实验结果如表4-3：

表4-3 静音与非静音片段性能测试

|  |  |  |  |
| --- | --- | --- | --- |
| 静音片段*(s)* | 单声源非静音片段*(s)* | 两声源非静音片段*(s)* | 三声源非静音片段*(s)* |
| 0.63 | 1.17 | 1.33 | 1.48 |

由表4-3可得：1）静音检测减少算法运行时间，降低功耗。2）随着声源数量增加，算法后续步骤耗时增加。

4.3本章小结

本章首先在4.1节总述了系统流程与分块功能，在4.2节展示系统集成的运行结果，单声源情况下DoA估计与检测较稳定，但系统在多声源位置估计与检测方面仍存在不足。并进行了实时性能测试，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0.63s，声源数量增加则算法耗时增加。综上，论文实现了咳嗽自动定位与检测的基线系统，在实时性能和多声源情况下系统仍有很大优化空间，可以在实际部署时做更多测试实验进行调优。

# 结论

积极响应疫情科研攻关的号召，本文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，监测潜在病例。系统可作为一种自动辅助诊断，解决复工复学情况下人员聚集空间中不易排查的问题，提供除发烧症状以外的早期新冠症状监测，减少监测过程中人力物力消耗，为呼吸系统疾病诊断与复工复学监控提供帮助。

在声源定位方面，本文对比MUSIC算法与SRP-PHAT算法，通过仿真实验发现：1）当两声源位置相近时，SRP算法的峰值较平滑，无法分辨出两个波峰。而MUSIC峰值明显，分辨率高。2）SNR较低影响DoA估计，噪声影响波峰，提高SNR有助于DoA估计。在波束形成方面，本文通过仿真，量化了MVDR滤波器的SNR增益。在咳嗽检测模块，训练SVM\_RBF模型，10折交叉验证F1指数为95.3%。并最终集成音频采集，静音检测，特征提取，声源定位，波束形成，咳嗽检测为实时系统，实时性能测试中，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0.63s，声源数量增加则算法耗时增加。

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：1）自动预测活跃声源数量。2）优化SRP-PHAT算法，解决在声源位置相近时，峰值平滑的问题。3）根据DoA估计算法复杂度，量化其对实时系统性能的影响，并优化算法速度。4）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。5）部署到嵌入式终端设备，进行实际测试。

# 参考文献

[1] ZU Z Y, JIANG M D, XU P P, 等. Coronavirus disease 2019 (COVID-19): a perspective from China[J]. Radiology, Radiological Society of North America, 2020: 200490.

[2] 中国-世界卫生组织新型冠状病毒肺炎（COVID-19）联合考察报告[R]. .

[3] 朱春媚, 黎萍. 基于支持向量机的咳嗽自动识别[J]. 计算机与现代化, 2016, 000(7): 111–114.

[4] 石锐, 王博, 何庆华. 基于高斯混合模型的咳嗽音检测方法[J]. 计算机工程与应用, 2011, 047(32): 151–154.

[5] 尹永. 基于Mel倒谱参数的咳嗽声识别[D]. 华南理工大学.

[6] 郑晓平. 基于HMM-ANN混合模型的咳嗽音识别研究[D]. 重庆大学.

[7] MORICE A, FONTANA G, BELVISI M, 等. ERS guidelines on the assessment of cough[J]. European respiratory journal, Eur Respiratory Soc, 2007, 29(6): 1256–1276.

[8] BARRY S J, DANE A D, MORICE A H, 等. The automatic recognition and counting of cough[J]. Cough, 2006, 2(1): 8.

[9] BIRRING S, FLEMING T, MATOS S, 等. The Leicester Cough Monitor: preliminary validation of an automated cough detection system in chronic cough[J]. European Respiratory Journal, Eur Respiratory Soc, 2008, 31(5): 1013–1018.

[10] CROOKS M G, HAYMAN Y, INNES A, 等. Objective measurement of cough frequency during COPD exacerbation convalescence[J]. Lung, Springer, 2016, 194(1): 117–120.

[11] DRUGMAN T, URBAIN J, BAUWENS N, 等. Audio and contact microphones for cough detection[J]. arXiv preprint arXiv:2005.05313, 2020.

[12] REN S, CAI M, SHI Y, 等. Influence of bronchial diameter change on the airflow dynamics based on a pressure-controlled ventilation system[J]. International Journal for Numerical Methods in Biomedical Engineering, Wiley Online Library, 2018, 34(3): e2929.

[13] KVAPILOVA L, BOZA V, DUBEC P, 等. Continuous Sound Collection Using Smartphones and Machine Learning to Measure Cough[J]. Digital Biomarkers, Karger Publishers, 2019, 3(3): 166–175.

[14] TRACEY B H, COMINA G, LARSON S, 等. Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis[C]//2011 Annual international conference of the IEEE engineering in medicine and biology society. IEEE, 2011: 6017–6020.

[15] SWARNKAR V, ABEYRATNE U R, AMRULLOH Y, 等. Neural network based algorithm for automatic identification of cough sounds[C]//2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2013: 1764–1767.

[16] AMRULLOH Y A, ABEYRATNE U R, SWARNKAR V, 等. Automatic cough segmentation from non-contact sound recordings in pediatric wards[J]. Biomedical Signal Processing and Control, Elsevier, 2015, 21: 126–136.

[17] LIU J-M, YOU M, LI G-Z, 等. Cough signal recognition with gammatone cepstral coefficients[C]//2013 IEEE China Summit and International Conference on Signal and Information Processing. IEEE, 2013: 160–164.

[18] PRAMONO R X A, IMTIAZ S A, RODRIGUEZ-VILLEGAS E. A cough-based algorithm for automatic diagnosis of pertussis[J]. PloS one, Public Library of Science, 2016, 11(9).

[19] SMITH J, WOODCOCK A. New developments in the objective assessment of cough[J]. Lung, Springer, 2008, 186(1): 48–54.

[20] BIRRING S S, KAVANAGH J, LAI K, 等. Adult and paediatric cough guidelines: ready for an overhaul?[J]. Pulmonary pharmacology & therapeutics, Elsevier, 2015, 35: 137–144.

[21] SPINOU A, BIRRING S S. An update on measurement and monitoring of cough: what are the important study endpoints?[J]. Journal of thoracic disease, AME Publications, 2014, 6(Suppl 7): S728.

[22] BIRRING S S, SPINOU A. How best to measure cough clinically[J]. Current opinion in pharmacology, Elsevier, 2015, 22: 37–40.

[23] KAPKA S, LEWANDOWSKI M. Sound source detection, localization and classification using consecutive ensemble of CRNN models[J]. arXiv preprint arXiv:1908.00766, 2019.

[24] ZHANG J, DING W, HE L. Data augmentation and prior knowledge-based regularization for sound event localization and detection[R]. Tech. Report of Detection and Classification of Acoustic Scenes and Events …, 2019.

[25] GRONDIN F, MICHAUD F. Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations[J]. Robotics and Autonomous Systems, Elsevier, 2019, 113: 63–80.

# 哈尔滨工业大学本科毕业设计（论文）原创性声明

本人郑重声明：在哈尔滨工业大学攻读学士学位期间，所提交的毕业设计（论文）《咳嗽自动定位与检测系统》，是本人在导师指导下独立进行研究工作所取得的成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明，其它未注明部分不包含他人已发表或撰写过的研究成果，不存在购买、由他人代写、剽窃和伪造数据等作假行为。

本人愿为此声明承担法律责任。

作者签名： **** 日期： 2020 年 6月 7 日

# 致 谢

受新冠肺炎疫情影响，在中期更换了毕业设计题目，是我第一次接触信号处理、音频处理与分析，学习相关知识的过程很艰难，但收获满满。这段学习经历锻炼了我搜寻文献、看论文、总结前人工作并思考什么值得借鉴。由于对相关领域了解不多，过程中疑问很多，总是千方百计去找解决方案，看不懂的理论推导每次看都有新的理解。最后拼凑的系统回头看有很大的改进空间，存在不少待解决的问题，希望在以后的学习中，能在论文的基础上探索更多，孜孜不倦。感谢完成这个基线系统的自己，未来一步一步，做大做强。

非常感谢我的导师，JIE LIU老师，对我的指导和帮助。刘老师科研经历丰富，视角非常具有前瞻性，在思路和理论上启发了我很多思考。每周一次的组会能和导师实时汇报进展探讨问题，同时实验室提供一流的科研平台供我们求知探索，非常荣幸在即将到来的研究生阶段能继续跟随刘老师学习。以及，尤其感谢郑榕老师的悉心指导，每次组会，郑老师都很认真地了解我的进展并给出建议。初识信号，很多问题不懂，郑老师都非常耐心地解释，提供参考文献和解决方案，切实地提供帮助，真的很感动。老师们的指导不局限于项目课题，他们的思考方式和想法也让我受益匪浅，良师益友，人生所幸。

想不到本科毕业时是家里蹲，让这一切结束地毫无仪式感。感谢在工大的四年，母校给了我很大空间，让我体味大学生活，找到兴趣所在，结交了很多好朋友和优秀的老师，在课程中积累知识，课余探索了许多周边美食，也曾为考试揪心难眠，所有事，迷茫过，清晰过。给我很大的感受是，很想做的事一定要去做，持续地探索，然后不要脸，不害怕，不着急。毕业是结束也是开始，开始更坚定目标，更不遗余力，去实现未完成的梦。

最后，再次感谢在毕业设计期间给予我帮助的父母、老师和同学们。我会继续追寻自己心之所向，下一个阶段启程！