

分类号：
密 级：

单位代码：10019
学 号：B20153080285

中國農業大學

学位论文

多媒体播放机器人语音控制方法研究

**Research on Method of Voice Control for Multimedia Player
Robot**

研 究 生：_____宋越_____

指 导 教 师：_____杨邦杰 研究员_____

合 作 指 导 教 师：_____

申请学位门类级别：_____工学博士_____

专 业 名 称：_____农业信息化技术_____

研 究 方 向：_____计算机农业应用技术_____

所 在 学 院：_____信息与电气工程学院_____

2020 年 04 月

独 创 性 声 明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的成果。尽我所知，除了文中已经注明引用和致谢的内容外，论文中不包含本人和其他人已经发表或撰写过的研究成果，也不包含为获得中国农业大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：

宋越

时间：2020 年 8 月 25 日

关于学位论文使用授权的说明

本人完全了解中国农业大学有关保留、使用学位论文的规定。本人同意中国农业大学有权保存及向国家有关部门和机构送交论文的纸质版和电子版，允许论文被查阅和借阅；本人同意中国农业大学将本学位论文的全部或部分内容授权汇编录入《中国博士学位论文全文数据库》或《中国优秀硕士学位论文全文数据库》进行出版，并享受相关权益。

(保密的学位论文在解密后应遵守此协议)

学位论文作者签名：

宋越

时间：2020 年 8 月 25 日

导师签名：

杨邦杰

时间：2020 年 8 月 25 日

摘 要

现代通讯系统中,语音交互已经成为一种最直接的人机交互方式。其不仅极大的提高了信息传递效率而且相比于传统的交互方式也表现出操作简单且正确率较高的特点。但在采用语音交互技术进行语音控制时,其效果容易受到周围环境噪声的干扰及回声路径变化的影响。同时,在语音控制的应用过程中,文本匹配也是影响语音指令执行正确性的关键问题之一。因此,本文针对语音控制应用过程中涉及到的关键问题,包括语音降噪技术、回声消除方法和文本匹配方法,进行了深入的研究,并搭建了语音多媒体播放机器人系统,以进行实验验证。本文的主要研究工作内容如下:

(1) 针对环境噪声对语音控制多媒体播放机器人噪声干扰问题,本文提出了基于改进的阈值函数小波包语音降噪算法。其典型特征在于保持了硬阈值函数的真实性,同时优化了硬阈值函数算法的过渡去噪问题。在语音降噪的应用过程中,该方法表现为不仅能保留带噪语音中的纯净语音信号,同时也可对噪声信号进行合理有效的滤除。仿真验证实验表明,相比与传统阈值函数方法,本方法在平稳与非平稳噪声环境下 SNR 和 PESQ 均提升了 10%,且 WSS 降低了 5%。因此表明本文提出的改进方法具有更加优异的语音降噪性能。

(2) 针对语音控制多媒体播放机器人的回声干扰问题,本文对回声路径进行了研究,并基于自适应滤波理论提出了一种改进的无参数可变步长子带自适应声学回声消除算法。该算法在回声消除的实际应用中可不依赖外部参数,因此其极大的降低了算法的计算复杂度。最后通过对本文提出的改进算法与传统的 NLMS 算法、可变步长算法和子带自适应滤波算法进行实验对比可知,在平稳环境和非平稳环境中,该改进的回声消除算法在收敛速度、稳态误差及抗干扰能力等方面均优于其它三种传统的回声消除算法。

(3) 为了提高语音控制多媒体播放机器人的语音指令执行准确率,本文对语音文本匹配在语音控制中的应用方法进行深入研究,并基于语音识别技术对语音文本进行文本匹配,同时提出了基于局部匹配的文本信息检索方法。该方法采用矢量检索方法计算语音文本并提交查询之间的相似度,并通过将相似度从大到小进行排序以进行文本检索。最后仿真与实验表明,与传统方法相比,本文所提出方法的召回率和准确率均在 99.5%以上,平均检索时间在 4~6s 之间,平均检索效率在 99%以上。因此可见,该方法的各项评价指标均优于传统方法。

(4) 在上述各研究内容的基础上,搭建了语音控制多媒体播放机器人原型系统。实现了语音控制多媒体播放机器人的语音交互、传统交互、多媒体播放和行走等功能。根据上述各功能的实现要求,设计了机器人的软件系统和硬件结构。并针对上述关键问题,对语音控制多媒体播放机器人整体进行了实验验证和评估。结果表明该原型系

统在非平稳环境下具有抗噪声干扰能力及对环境的应变能力,同时提高了语音指令的执行能力。

关键词: 语音控制, 语音降噪、回声消除、文本匹配

Abstract

In modern communication systems, voice interaction has become the most direct method for human-computer interaction, which can greatly improve the efficiency of information transmission. Furthermore, compared with traditional information interaction methods, the voice interaction approaches show the characteristics of simple operation and high accuracy rate. However, when the voice interaction technology is used for voice control, its performance is easily affected by ambient noise and variation of the echo path. Meanwhile, in the application process of voice control, text matching is also one of the key problems. Therefore, in this paper, the application of voice noise reduction, echo cancellation and text matching in voice control are deeply studied. Besides that, a voice multimedia robot prototype system is built for experimental verification. The main works of this paper are shown as follows:

(1) Aiming at the problem of noise interference from environment noise to voice control multimedia playing robot, this paper proposed a speech denoising algorithm based on improved hard threshold function wavelet packet. The typical characteristics of the algorithm is both remaining the authenticity of hard threshold function and optimizing the transition denoising of hard threshold function algorithm. In the application process of speech denoising, the method can not only preserve the pure speech signal from the noisy speech, but also filter the noise signal reasonably and effectively. The simulation results show that compared with the traditional threshold function method, the SNR and PESQ of the proposed method are both improved by 10%, meanwhile the WSS is reduced by 5%. It can be seen that the proposed method has better performance for speech noise reduction.

(2) Aiming at the problem of echo interference of voice controlled multimedia player robot, an improved non-parametric variable step length subband adaptive acoustic echo cancellation algorithm based on adaptive filtering theory is proposed in this paper. In practical applications of echo cancellation, the proposed algorithm does not depend on parameters, so the computational complexity of the algorithm is greatly reduced. The comparison experiments between the proposed algorithm and traditional NLMS algorithm, variable step size algorithm and subband adaptive filtering algorithm show that the proposed algorithm has faster convergence speed, lower steady-state error and high anti-noise ability than other echo cancellation algorithms.

(3) In order to improve the accuracy of speech command execution of voice control multimedia player robot, this paper deeply studied the application of speech text matching

in speech control, meanwhile, based on the speech recognition technology, the text matching is carried out on speech text; and a text information retrieval method based on local matching is further proposed. The proposed method adopts the vector retrieval approach to calculate the similarity between speech texts and submit queries, and then the text is retrieved by sorting the similarity from large to small. The simulation and experimental results indicate that the recall rate and accuracy rate of the proposed method are all above 99.5%, the average retrieval time is between 4~ 6s, and the average retrieval efficiency is more than 99%, which show that all evaluation indexes of the proposed method are superior to the traditional methods.

(4) On the basis of all the above research contents, a prototype system of voice control multimedia player robot was built, in which, the functions of voice interaction, traditional interaction, multimedia play and walking are realized. According to the realization requirements of the functions, the software system and hardware structure of the robot are designed. Meanwhile, focusing on the above key problems, the overall experimental verification and evaluation of the voice controlled multimedia player robot are conducted. The experimental results show that the prototype system has the high ability of anti-noise interference, the better adaptability to the non-stationary environment, and strong execution ability for voice instructions.

Key words: Voice control, Speech denoising, Echo cancellation, Text matching

目 录

| | |
|------------------------------|----|
| 第一章 绪论 | 1 |
| 1.1 研究的背景与意义 | 1 |
| 1.2 国内外研究现状 | 2 |
| 1.3 研究目标与研究内容 | 14 |
| 1.4 技术路线与论文组织结构 | 15 |
| 1.5 本章小结 | 17 |
| 第二章 语音控制机器人环境噪声降低算法研究..... | 18 |
| 2.1 引言 | 18 |
| 2.2 语音信号特性及降噪算法 | 19 |
| 2.3 改进阈值函数小波包语音降噪方法 | 26 |
| 2.4 实验结果与分析 | 31 |
| 2.5 本章小结 | 37 |
| 第三章 回声特征与回声消除方法研究..... | 38 |
| 3.1 引言 | 38 |
| 3.2 回声研究及基本回声消除算法 | 39 |
| 3.3 改进的可变步长子带自适应算法 | 50 |
| 3.4 实验结果与分析 | 52 |
| 3.5 本章小结 | 58 |
| 第四章 文本匹配在语音控制中应用方法研究..... | 59 |
| 4.1 引言 | 59 |
| 4.2 文本匹配的方法研究与概述 | 60 |
| 4.3 基于局部匹配的数字化文本信息检索方法 | 63 |
| 4.4 实验结果与分析 | 65 |
| 4.5 本章小结 | 76 |
| 第五章 语音控制多媒体播放机器人搭建及实验研究..... | 78 |
| 5.1 引言 | 78 |
| 5.2 语音控制多媒体播放机器人介绍 | 78 |
| 5.3 系统结构 | 79 |
| 5.4 总体实验验证 | 84 |
| 5.5 本章小结 | 92 |
| 第六章 结论与展望 | 93 |

| | |
|---------------|-----|
| 6.1 结论 | 93 |
| 6.2 创新点 | 94 |
| 6.3 展望 | 94 |
| 参考文献 | 95 |
| 致 谢 | 108 |
| 作者简介 | 109 |

插图清单

| | |
|---------------------------------------|----|
| 图 1-1 技术路线图..... | 16 |
| 图 2-1 谱减法框图..... | 21 |
| 图 2-2 子空间算法基本结构..... | 23 |
| 图 2-3 深度神经网络结构图..... | 25 |
| 图 2-4 语音降噪系统流程图..... | 26 |
| 图 2-5 小波包语音降噪原理图..... | 26 |
| 图 2-6 两尺度小波包树变换..... | 27 |
| 图 2-7 三层小波包分解树图..... | 27 |
| 图 2-8 三层小波包分解谱特性图..... | 28 |
| 图 2-9 小波包系数的硬阈值函数..... | 29 |
| 图 2-10 小波包系数的软阈值函数..... | 30 |
| 图 2-11 改进阈值方法..... | 31 |
| 图 2-12 三种阈值函数在白噪声环境下 SNR 增长图..... | 33 |
| 图 2-13 三种阈值函数在白噪声环境下 WSS..... | 34 |
| 图 2-14 三种阈值函数在非平稳声环境下 SNR 增长..... | 35 |
| 图 2-15 三种阈值函数在非平稳噪声环境下 WSS..... | 36 |
| 图 3-1 声学回声模型图..... | 39 |
| 图 3-2 电路回声模型图..... | 40 |
| 图 3-3 自适应滤波结构框图..... | 41 |
| 图 3-4 子带滤波器开环结构..... | 46 |
| 图 3-5 子带滤波器闭环结构..... | 47 |
| 图 3-6 四种算法的稳态误差比较, SNR=30dB..... | 54 |
| 图 3-7 四种算法的稳态误差比较, SNR=10dB..... | 55 |
| 图 3-8 四种算法的 ERLE 比较, SNR=30dB..... | 55 |
| 图 3-9 双音讲话稳态误差评估, SNR=30dB..... | 56 |
| 图 3-10 双音讲话稳态误差评估, SNR=10dB..... | 57 |
| 图 3-11 双音讲话的 ERLE 评价标准, SNR=30dB..... | 58 |
| 图 4-1 深度学习文本匹配框图..... | 63 |
| 图 4-2 文本检索结果分布图..... | 67 |
| 图 4-3 三种方法召回率对比实验..... | 68 |
| 图 4-4 加入 10%干扰信息三种方法召回率对比实验..... | 68 |

| | |
|--------------------------------------|----|
| 图 4-5 加入 20%干扰信息三种方法召回率对比实验..... | 69 |
| 图 4-6 加入 30%干扰信息三种方法召回率对比实验..... | 70 |
| 图 4-7 三种比较方法检索准确度..... | 72 |
| 图 4-8 局部匹配方法和最大权重文本匹配的所需要信息检索时间..... | 74 |
| 图 4-9 三种方法检索效率在第一阶段中的比较..... | 75 |
| 图 4-10 三种方法检索效率在第二阶段中的比较..... | 75 |
| 图 4-11 三种方法检索效率在第三阶段中的比较..... | 76 |
| 图 5-1 语音控制多媒体播放机器人..... | 79 |
| 图 5-2 语音控制多媒体播放机器人主板..... | 80 |
| 图 5-3 语音控制多媒体播放机器人麦克风阵列..... | 81 |
| 图 5-4 语音控制多媒体播放机器人显示设备..... | 81 |
| 图 5-5 语音控制多媒体播放机器人行走系统..... | 82 |
| 图 5-6 语音控制多媒体播放机器人系统操作界面..... | 82 |
| 图 5-7 CAN 总线图..... | 83 |
| 图 5-8 机器人语音控制程序框图..... | 84 |
| 图 5-9 机器人原型系统客观评估..... | 90 |
| 图 5-10 机器人原型系统主观评估..... | 91 |

附表清单

| | |
|--|----|
| 表 2-1 三种阈值函数在白噪声环境下 PESQ 评均值 | 34 |
| 表 2-2 三种阈值函数在非平稳噪声环境下 PESQ 评均值 | 36 |
| 表 3-1 LMS 算法总结 | 43 |
| 表 3-2 NLMS 算法总结 | 44 |
| 表 3-3 非参数可变步长 NLMS 算法总结 | 49 |
| 表 3-4 广义 Sigmoid 可变步长 NLMS 算法总结 | 50 |
| 表 3-5 NPVSS NASF 算法总结 | 52 |
| 表 4-1 基于最大权重匹配计算信息检索方法的准确率 | 71 |
| 表 4-2 基于双语义空间的信息检索方法的准确率 | 71 |
| 表 4-3 基于本文方法的信息检索方法的准确率 | 72 |
| 表 4-4 三种方法文本信息检索时间 | 73 |
| 表 5-1 不同人群对语音控制多媒体播放机器人语音控制的影响 | 85 |
| 表 5-2 不同识别距离对对语音控制多媒体播放机器人语音控制的影响 | 86 |
| 表 5-3 不同噪音对语音控制多媒体播放机器人语音控制的影响 | 87 |
| 表 5-4 不同回声路径对语音控制多媒体播放机器人语音控制的影响 | 88 |
| 表 5-5 不同指令相同语义对语音控制多媒体播放机器人语音控制的影响 | 89 |
| 表 5-6 用户动静状态对语音控制多媒体播放机器人语音控制的影响 | 90 |
| 表 5-7 志愿者体验调查表 | 91 |

第一章 绪论

1.1 研究的背景与意义

当今的智能多媒体设备与传统的多媒体设备有着本质的区别,传统的多媒体设备一般只是媒体的载体,然而智能多媒体设备是在传统多媒体设备的基础上应用了计算机技术、物联网技术及各种人工智能算法,从而实现了智能多媒体设备的自动化、智能化和远端操作化。然而现有的智能多媒体控制设备大多数还是传统的遥控器、各种材质的触摸屏、鼠标和键盘等,这些控制方式虽然能够达到控制目标,但操作繁琐,对于农民、中老年人及儿童不太适用,操作的时候往往容易进入误区,无意修改设备的设置选项等功能,从而带来不便。

在 21 世纪,语音识别技术已经应用到人们日常生活中的各个角落,例如微信、百度、有道翻译和淘宝这些常用的手机软件中都具有语音识别功能,它让人们可以更快、更便捷地获取有效信息,从而大大地提高了人们的办事效率。现在不仅仅手机,越来越多的智能多媒体设备也引进了语音控制技术,用户可以通过语音指令控制家中的热水器、电视机、灯及空调等人们经常使用的家电,代替了传统的遥控器和手动开关。因此,语音智能多媒体设备的出现,极大的改善了人们的生活,让人们的生活更加便捷及智能化。目前市场中也有一些多媒体设备是支持语音控制的,但是它们的控制方式比较单一,语音指令都是预先设定好的,它们不能随着外界环境改变而改变。例如,当语音控制设备预先设置“1+1 等于 2”,但是当用户语音指令为“1+1 是多少”的时候,语音控制系统就会失效,缺少智能化。综上所述,拥有自我学习、自我分析判断、自我预测功能的智能语音控制系统亟待研发。

语音控制的应用噪声干扰和文本语音匹配一直是智能多媒体设备应用两大技术难题。现在生活中的大部分语音控制智能多媒体设备在实验中,语音识别准确率都可以达到 95%以上,但是在实际应用中它们的语音识别率会急剧降低,甚至无法实现语音识别功能。比如说,Dautrich^[1]等人所演示的语音识别系统在非噪声环境下识别准确率高达 90%以上,然而加入一些噪声之后,系统语音识别准确率会急剧下降到 60%,甚至更低。尽管在噪声环境下训练的语音识别系统,随着环境的变化,语音识别系统的准确率也会急剧降低。例如,研究人员在实验室中模拟车站的噪声环境训练语音识别系统,使其语音识别准确率达到 95%以上,然而当环境变成嘈杂的餐厅时,系统的语音识别准确率会降到 50%以下。综上所述,尽管在安静的环境下研究人员将语音识别系统训练地非常精准,但是当环境发生改变时,语音识别系统却无法达到人们的需求。综上所述噪声干扰是语音控制实际应用中不可忽略的难题。语音降噪的目的就是

通过某种方法从带噪语音信号中尽可能地去噪声干扰信号,最大化地提取人们需要的语音信号。人们往往通过语音降噪技术手段提高智能多媒体设备的语音质量。

语音文本匹配技术是在现有的语音识别技术为基础实现的,首先是由语音识别系统模块将语音转化成文字,从而获得语音与文本之间相互作用的关系。在不同的领域中,训练模型是截然不同的,从而导致在文本匹配的过程中结果也是存在巨大的差异。语音文本匹配技术中的文本信息是已知的,匹配过程就是将已知文本在不同的训练模型下分割成不同的单元。文本语音匹配技术的应用非常广泛,在线翻译、字幕生成、多媒体检索和语音控制等方面。在语音控制智能多媒体设备中,语音文本匹配技术是语音控制指令执行成功的关键之处。

本文主要针对语音控制多媒体播放机器人环境噪声干扰、自身背景音干扰和语义理解问题,在对相关理论和方法比较分析的基础上,结合语音控制多媒体播放机器人应用需求,基于小波包理论和自适应滤波技术降低环境噪声的干扰和回声消除,并能准确地执行语音指令,播放出用户所需要的视频、音频和文本,最后设计并完成语音控制多媒体播放设备原型系统。

1.2 国内外研究现状

自 2000 年以后,语音技术发展迅速,自 2011 年苹果公司推出 iPhone 4S 语音对话功能之后,各个语音科技公司推出了各种各样的智能语音产品。Amazon 发布了 Echo 智能音箱,而且在美国成为一款最热销的智能音箱产品,受到了不同年龄段人们的青睐。美国科技巨头谷歌公司也参与到语音技术产品争夺的战争中,它收购了许多语音技术公司和相关专利。谷歌并且开放了自己的 API,它包含了 80 多种语言,方便广大用户进行语音识别和实时翻译。美国另一大科技巨头苹果公司收购 Siri 等语音技术公司,并且投入大量资金建立了自己的语音识别团队。在国内,科大讯飞几乎成为全世界众所周知的语音技术公司,它作为我国智能语音技术产业的龙头企业,它的语音识别、语音合成等众多关键技术在国际上获得领先的地位。我国还拥有“灵云科技”,“云知声”等一些优秀的智能语音技术公司。并且各种智能语音产品也在市场上琳琅满目。

在未来的 5 至 10 年,智能语音技术将涉及到更广的领域,汽车、医疗、家用电器等各个行业都会看到智能语音技术的身影。在不久的将来,随着语音技术不断的更新发展,人们面对互联网以及一些硬件设备,无需再使用“键盘”去敲打文字,用人们最习惯的语音交流便能进行交互。语音交互的使用,会让用户更加方便直接。所以说语音控制技术的兴起,提高了人类对智能设备、互联网站使用效率,为人类的生活

带来了便捷。

1.2.1 语音降噪技术研究

在上世纪 60 年代,许多学者开始对语音降噪算法进行初步研究,到了 70 年代左右诞生了众多重要的研究成果,并且对之后的语音降噪算法研究起到了至关重要的作用,对近些年来新的语音降噪算法奠定了基础。在过去的半个世纪中,许多语音降噪算法出现在人们的视野里,这些算法主要为:维纳滤波法^[2-4]、谱减法^[5-6]、最小方误差估计法^[7]和基于信号子空间法^[8-9]等。

谱减法是语音降噪算法中最经常使用算法之一,它在上世纪 70 年代就被广泛的应用^[10-13]。它的技术核心是“谱”和“减法”,“谱”是指信号的功率谱或者幅度谱,“减法”是指从带噪语音信号中减去无用的语音信号得到有用的语音信号。由于加性噪声信号与纯净语音信号是无关的,谱减法正是利用了这一特征进行语音降噪。它是假设噪声是一个平稳或缓慢变化的过程,并且在更新周期之间噪声谱没有显著变化,噪声信号中有且仅有加性噪声,然后从带噪语音信号谱中将噪声信号谱去除,从而得到纯净语音信号谱。该算法就计算复杂度而言,谱减法相对简单,并且容易实现,对计算机要求也非常小,所以在早期可以广泛地被应用。

在上世纪刘丹亭^[14]等人提出了多模型谱减法,它针对环境噪声的的多样化从而对不同种类噪声进行训练与分类,如果从带噪语音信号中发现与训练模型相同的噪声模型,便从带噪语音信号中减去相同类型的噪声信号谱从而获得纯净语音信号,提高了系统的鲁棒性。金薛冬^[15]等人提出了基于递归算法的改进谱减算法,该改进的算法去噪性能优异,鲁棒性很强,提高了语音系统的语音质量。郑成诗^[16]等人对噪声谱结构特性进行了分析与实验验证,他们提出了基于噪声谱结构特性的谱减法,该算法大大地提高了传统谱减法的信噪比提高和语音降噪效果。

Kamath^[17]针对有色噪声的特殊性提出了多波段谱减法,因为有色噪声具有多样性和复杂性,它在不同频率下会对语音谱产生不同的变异效果。该方法优于标准的功率谱相减方法,具有更好的语音质量且大大降低音乐噪声。Lebart^[18]提出了一种新的基于谱相减的单声道抑制语音信号后混响的方法,该算法在各种混响环境中的自动语音识别分数方面带来了显著的改进。Radu^[19]等人提出了一种改进的谱减法,以减少直升机座舱或汽车引擎等噪声环境中的噪声,该方法利用人耳听觉系统的掩蔽特性可以屏蔽残余噪声。

虽然谱减法语音降噪技术计算简单,容易实现,但是由于噪声的随机变化,谱减法会导致短时幅度或功率谱的负估计。当纯正语音信号与噪声语音信号是非正交的,

传统的谱减法语音降噪效果不能达到理性的效果，处理结果误差比较大。且随着信噪比的降低，处理失真变得更加明显。

在上世纪中期，Lim 等人^[20]提出了维纳滤波（Wiener Filtering）语音降噪算法，它是基于最小均方误差准则实现语音降噪。该方法主要是通过设计一种线性滤波器，让该滤波器的纯净语音输出信号与带噪语音信号的均方误差为最小，它主要应用与平稳噪声语音信号。使用维纳滤波语音降噪算法一般遵循两点，一是输入语音信号是平稳的，二是语音信号输入过程中统计特性是已知的。因此，使用维纳滤波语音降噪算法要对输入语音信号进行预处理。它的优点在于语音降噪后的残留噪声信号与白噪声比较相似，它的缺点就是不能在非平稳随机过程中进行应用。

奚吉等人^[21]提出一种改进的维纳滤波语音降噪算法，针对传统维纳滤波语音降噪算法噪声估计偏差大的问题，改进算法结合了人耳听觉特性将全带带噪语音信号分解成多个子带语音信号，先在每个子带进行语音降噪处理，然后通过综合所有子带信号完成语音降噪处理。该方法有效地提高了语音降噪算法性能，并且同时提高了可懂度。靳立燕等人^[22]提出了改进的维纳滤波算法和奇异谱分析算法相结合的语音降噪算法，新的算法主要是改善了传统维纳滤波语音降噪算法在非平稳环境下容易存在信号失真，并且提高了带噪语音信号的信噪比。从仿真实验结果中可以看出，该算法可以在不同的非平稳环境下可以有效地进行噪声消除，并且应用范围非常广。彭健新和唐云飞^[23]提出了谱减法和维纳滤波器相结合的算法，在仿真实验中降噪效果优于传统的谱减语音降噪算法和维纳滤波降噪算法。

Kim 等人^[24]提出了一种基于可变加权因子的语音失真加权多通道维纳滤波器（SDW-MWF）的多通道降噪算法，由于传统的 SDW-MWF 算法使用固定的权重因子在噪声降低和语音失真之间进行权衡从而容易造成语音失真，该算法提出了一种在降噪和语音失真之间进行更灵活权衡的解决方案，在降低语音失真的情况下提高了信噪比。Cornelis 等人^[25]基于语音失真加权多通道维纳滤波器降噪技术提出了两种新的双耳助听器的降低带宽噪声的方法，该方法对声音定位和提高语音可懂度的效果非常好。

子空间的信息提取一般有两种方法：一种是特征值分解法，它分解后由特征值和特征向量表示；另一种为奇异值分解法，它在某方面与对称矩阵的对角化相似，但又完全不相同。这两种方法在实验仿真都是非常优异的，但是它们都存在一个共同缺点，计算程度比较复杂，不适合在实际问题中应用。Dendrinos 等人^[26]最早提出了基于子空间语音降噪算法，它的主要思想是将带噪语音信号分成两部分：语音信号子空间和噪声信号子空间。该算法的最终目的就是丢弃噪声信号子空间，保留纯净的语音信号子空间，但是在语音信号子空间中仍然存在噪声信号，从语音信号子空间尽可能去除

的噪声信号从而提取纯净语音信号。

吴冠宁等人^[27]提出了基于组合算法的语音降噪研究,该算法利用了子空间算法和自适应算法的优点并将二者相结合,首先利用子空间算法将带噪语音信号进行分解,然后利用自适应滤波降噪算法分实现语音增强,实验表明该组合算法在语音降噪方面性能优异且强于两种单一语音降噪方法。肖毅^[28]针对低信噪比情况下的语音降噪,提出了基于滑动窗口子空间语音降噪算法,该方法主要运用空间重组和多次迭代计算去解决低信噪比问题。该方法与传统子空间语音降噪算法相比,大大提高语音降噪的精度,减少纯净语音信号失真。裴俊华^[29]等人提出了子空间自适应小字典语音降噪算法,该算法良好地解决了语音信号失真问题,大大提高了语音质量。

Ephraim 和 Van 等人^[30]等人提出了一种基于感知的线性信号估计器,用于增强由不相关加性噪声退化的语音信号,该估计器采用信号子空间方法实现。该算法与传统子空间算法一样将带噪声信号的向量空间分解为两个部分:信号子空间和噪声子空间,但是该算法的噪声子空间是互补正交的,它在语音降噪方面表现更加优异。由于在语音应用中,大多数降噪方法的主要缺点是被称为音乐噪声的恼人的残余噪声,Firas^[31]为解决这一问题将人类听觉模型与子空间语音降噪算法相结合,该方法在语音降噪方面强于传统的子空间语音降噪算法。Hermus^[32]对信号子空间语音增强进行了广泛的综述,且推导了这些技术性能的上界,全面研究了子空间滤波在增强自动语音识别器对平稳加性噪声失真的鲁棒性方面的潜力,最后还证明了子空间滤波比已知的谱相减技术更有优势。

传统子空间语音降噪算法是根据理论假设提出的,在实际的应用中并没有理论仿真实验效果好。由于现实环境因素的限制传统子空间语音降噪算法会残留残余噪声及造成语音失真。最近几年来,随着小波包算法和神经网络算法的广泛应用,许多学者将它们应用到语音降噪中并取得了良好的效果。小波包分解算法是小波分解算法的一种改进,小波包可以更精细地分解信号,它与小波分解算法相比,它将信号的时频分解更细致,尤其在高频部分,小波分解算法只在低频部分进行再分解,忽略了高频部分,所以会失去信号中重要信息,小波包分解算法解决了这一问题。各种基于小波包语音降噪的算法也被提出。

田玉静^[33]等发现了语音信号被小波包分解成各个子带信号后,它们之间存在着语音共振峰,结合人类听觉感知系统,提出了基于共振峰的小波包语音增强算法,通过控制每个子带阈值进行语音降噪。实验结果表明,该算法比传统的小波包语音降噪算法在低信噪比情况下具有更好的语音失真度且可以更好地提取纯净语音信号。

由于小波包软阈值和硬阈值小波包语音降噪算法都存在固有缺陷,任永梅^[34]提出了一种新的阈值函数小波包语音降噪算法,该新阈值函数结合了小波包的软阈值

和硬阈值的优点,实现新阈值函数连续性且平缓过渡。新的阈值函数结合人耳听觉感知特性在仿真实验中与传统阈值函数语音降噪算法相比具有较好的语音降噪效果。

Sungwook^[35]考虑了小波阈值法对非平稳或有色噪声的估计,提出了在有色或非平稳噪声中自适应的节点相关阈值。然后在仿真实验中又提出了一种基于强度直方图的谱熵噪声估计方法,代替了基于中值绝对偏差(MAD)的估计方法。实验结果表明,新的阈值方法与传统的基于MAD的水平相关阈值方法相比具有较大地提升。

Sanam^[36]对带噪语音信号的Teager能量进行了特征分析,并且基于Teager能量改进了硬阈值函数,提出了基于小波包阈值函数语音降噪算法,将新的阈值函数用于带噪语音信号WP系数。Sanam对上述算法进行了大量仿真实验,实验结果表明该方法对语音信号中的白噪声信号和彩色噪声信号具有良好的抑制效果,从而提高了语音质量。多个标准的客观测量和主观观察表明,该方法在从高到低的SNR中都优于传统阈值小波包语音降噪方法。

Mohamed和Aicha^[37]研究了含噪语音的增强和降噪问题,提出了一种基于小波变换系数重构的去噪方法。该技术将小波去噪方法的直接多元推广与主成分分析技术相结合。在不同类型的噪声下进行了大量的仿真,通过客观测量和主观评价表明,该方法在低信噪比下对噪声干扰语音的降噪效果优于比较的语音降噪方法。

基于神经网络的语音降噪算法在2013年开始出现在大众的视野里,并在语音降噪方面取得了良好的效果。这为之后学者研究基于神经网络的语音降噪算法打开了大门,随后基于神经网络语音降噪算法的大量研究成果不断出现在人们的视野里。

郑鲲等人^[38]提出了采用RBF神经网络在语音信号频域进行降噪处理,传统的RBF神经网络语音降噪算法是时域上进行降噪处理。通过大量仿真实验比较,频域语音降噪处理的效果要优于时域上的语音降噪,更好地提高了语音质量。

赵鸿图^[39]基于小波神经网络的特性提出了一种改进的粒子群优化算法进行语音降噪。由于小波分解算法在信号时频域上具有良好地处理效果,神经网络算法拥有良好地自我学习功能,小波神经网络算法具有二者的优点。将粒子群算法与小波神经网络算法相结合对语音信号进行降噪处理,通过实验证明语音降噪效果可以有效提高,并且具有良好的鲁棒性。

文仕学^[40]将渐进学习语音降噪算法应用到语音识别中,代替了传统神经网络语音降噪方法,很好地提高语音信噪比。在语音识别系统进行了语音信号预处理,在语音识别准确率有10%地提高。

Morten^[41]研究了基于深度网络的单麦克风语音增强技术,构建了语音增强系统在背景噪声类型、目标说话人性别和信噪比方面通用模型。利用已经训练好的模型与最小均方误差算法进行比较,通过实验证明,基于深度神经网络的麦克风语言增强技术

在语言质量和语言可懂度有较大的改善。基于深度学习神经网络语音降噪方法技术在降噪效果取得了很好的效果，但是它需要大量的训练数据且结构复杂，对于简单的智能语音设备应用效果不佳。

综上所述，语音降噪是语音信号处理的重要问题，在过去的 50 年里一直受到广大学者的关注。为了提高语音降噪能力，国内外众多学者提出一些有效方法，但是随着人们对智能语音设备语音质量的高要求，传统的语音降噪算法不能满足人们的需求。一个计算简单、去噪性能好的语音降噪算法是人们语音信号处理急迫需要解决的问题。

1.2.2 回声消除技术研究现状

回声现象是人们现实生活中经常发生的，它严重地影响了通信设备的语言质量。尤其是在语言或者视频会议中，讲话者会通过自己的扬声器听到自己的声音，让人们感到极其不舒适，降低了会议语言质量。所以在现代通信设备中，声学回声是影响语音质量的常见问题。因此，回声消除技术近年来受到了广泛的关注。许多自适应滤波算法^[42-47]被应用于声学回声消除系统，其基本原理是建立一个回波路径脉冲响应信号，自适应滤波器可以在扬声器和麦克风之间识别出该模型，从而生成回声的电子估计信号，通过从麦克风参考信号中减去回声信号来消除回声。

在上世纪 60 年代，Widrow 和 Hoff 在大量前人研究的基础上提出了自适应最小均方（LMS）算法^[48]。由于该算法的结构简单并且鲁棒性强，在上世纪中后期回声消除中得到了广泛的应用。然而，LMS 算法的最优步长是固定的，在实际应用中还是有较大的局限性。赵新波和杨苹^[49]提出了一种改进可变步长 LMS 算法，该算法结合了误差控制步长和误差积累的方法在实际中进行应用。通过大量的仿真对比实验，结果表明该算法比传统的 LMS 算法具有更快的收敛速度和较低的稳态误差，即使环境发生突然变化，也能保持收敛速度。但是 LMS 算法的收敛速率较慢，在回声消除中不能完全满足人们的需要。

归一化最小均方（NLMS）回声消除算法是 LMS 的一种特殊形式，它将 LMS 算法的步长因子进行了归一化。与 LMS 回声消除算法相比，它在收敛速度和稳定性方面都有巨大的提升。但是它与 LMS 有着一个共同的缺点，它们的步长因子都是固定的，不能在同一时刻满足较快的收敛和低的稳态误差。所以为了解决这些问题，有许多学者对 NLMS 算法进行改进，提出了改进的 NLMS 算法。

王正腾和谢维波提出了一种改进的 NLMS 算法^[50]，该算法根据滤波器原理模拟回声信号，通过阈值调节滤波器活跃区域的权重因子，抛弃非活跃区域。改进的算法不仅仅保留了 NLMS 算法原有的结构简单，还解决了快速收敛和较低的稳态误差的

矛盾, 通过仿真实验表明, 该算法具有较优异的性能。

由于非高斯噪声具有良好的滤波性能, 具有误差非线性的自适应滤波器近年来受到越来越多的关注。Zhnag^[51]等人将现有的基于最小均方 (LMS) 的误差非线性自适应算法分为 V 形、 \wedge 形和 M 形三类。提出了一类新的鲁棒 M 形误差加权 LMS (NFRMS-LMS) 算法。对其稳态均方差 (MSD) 进行了分析, 揭示了误差非线性的学习能力。通过 Monte Carlo 仿真验证了理论结果, 并在不同环境下验证了算法的有效性。Ciochina^[52]提出了一种优化的 NLMS 算法, 该算法在归一化步长和正则化参数上都遵循一个联合优化问题, 以最小化系统失调。实现了即快速收敛和稳态误差适当的折衷。

为了加快收敛速度, Duttweiler^[53]提出了一种比例归一最小均方 (PNLMS) 自适应算法, 该算法初始阶段的收敛速度较快, 并且较好地保持较低的稳态误差。与传统 NLMS 算法相比较, 加快了收敛速度但是也增加了计算复杂度。并且当回声路径发生改变时, 该算法性能会极度衰减。李跃明和侯楚林针对 PNLMS^[54]快速收敛和低稳态误差之间的矛盾, 提出了一种变步长比例仿射投影算法, 该算法利用后验误差补偿稳态误差对系统的负作用, 通过实验表明在回声消除方面, 该算法可以实现快速收敛和较低稳态误差。

为了更好解决快速收敛和稳态误差之间的矛盾, 许多学者提出了可变步长的 NLMS 算法。Shin^[55]提出了一种改进的可变步长归一化最小均方投影和仿射投影算法, 通过仿真实验结果可以得出该方法收敛速度和稳态误差都得到提升。分数抽头长度归一化最小均方 (FT-NLMS) 自适应算法与现有的变抽头长度 LMS 算法相比, 具有鲁棒性强、计算复杂度低等优点, 但是在 FT-NLMS 算法中, 参数增量控制着收敛速度和最佳抽头长度的偏差, 由于其固定值, 无法同时满足二者之间的矛盾, Mayyas^[56]提出了一个调整良好的时变步长因子, 可以提供快速的收敛速度和小的稳态偏差。该算法在噪声干扰不相关的情况下, 自适应滤波器能逼近最优值, 可以同时满足快速收敛和较小的稳态误差。Yu^[57]为了克服传统的归一化最小均方算法在快速收敛和低稳态偏差之间的折衷, 利用输入信号功率和输入信号与误差信号的互相关来估计真实跟踪误差功率提出了可变步长 NLMS 算法, 不停更新步长因子。经过实验仿真发现, 改进的算法比其它可变步长算法具有更低的稳态误差。

可变步长归一化最小均方 (VSS-NLMS) 算法的目的是为了同时满足较快的收敛速度和低的稳态误差, 解决二者之间的矛盾。在上文中提到的大多数 VSS-NLMS 算法都有一个共同缺点, 它们在实际应用中完全依赖参数, 然而这些参数在实际应用中不容易调整, 从而导致算法在实际应用中不能达到满意的效果。Benesty^[58]等人提出了一种改进的 VSS-NLMS 算法, 该算法不受任何参数的影响, 在实际应用中易于控

制，在回声消除的背景下具有良好的性能。以上所有的这些 VSS-NLMS 算法都可以分为两类。第一类算法基于声路的状态空间，第二类算法基于误差信号。虽然 VSS-NLMS 算法具有良好的收敛速度和较低的稳态误差，但是当输入信号为彩色信号时，上述的 VSS-NLMS 算法收敛速度会急剧下降。

为了克服彩色输入信号的问题，Lee 和 Gan 提出了一种归一化子带自适应滤波 (NSAF) 算法^[59]。NSAF 算法将每个子带中的输入信号转换为多个子带以进行白化。尽管 NSAF 算法在面对有色输入信号时具有较快的收敛速度，但 NSAF 算法和 NLMS 算法都存在步长固定的缺点。它们都需要面对快速收敛和低稳态误差之间的矛盾。为了解决这一矛盾，一些文献^[60-65]提出了一些可变步长 NSAF 自适应滤波算法。

Jae^[66]和 Ni^[67]几乎同时提出了一种可变步长 NSAF 自适应滤波算法，该算法通过在每次迭代中最小化最优权值向量与估计权值向量的均方偏差来实现。通过大量的仿真对比实验可以看出，尽管在高噪声环境下，该子带自适应滤波器仍然可以保持较低的稳态误差。

Shams 和 Shafiee^[68]提出了一种基于子带滤波器动态选择的变步长归一化子带自适应滤波器算法。执行选择策略以在每次迭代中实现连续均方偏差之间的最大减少。该算法具有较快的收敛速度和较低的稳态误差。由于 DS 的特性，VSS-DS-NSAF 算法计算复杂度远远低于 VSS-NSAF 算法，并且通过仿真实验证明了该算法的优异性能。以上 VSS-NSAF 算法虽然能很好地解决快速收敛和低稳态误差的矛盾，但它们都有很高的计算复杂度。Ni 从另一个角度提出了一种变步长矩阵 NSAF (VSSM-NSAF)^[69]，即从自适应滤波器的子带误差信号中恢复子带系统噪声的功率，以进一步提高 NSAF 的性能。该方法也可以应用于建模不足的情况，因此不需要事先知道子带系统噪声的功率。此外，分析了该算法的稳态均方特性，从理论上证明了 VSSM-NSAF 可以获得较低的稳态误差。仿真结果表明，与其它 NSAF 类算法相比，该算法具有良好的性能。Yu 和 Zhao^[70]等人为了提高改进的归一化子带自适应滤波器 (INSAF) 和比例 INSAF 算法在非平稳环境中的性能，提出了它们的集隶属度版本。该算法相比其它 NSAF 算法具有更小的稳态误差，而且计算复杂度也大大地得到了降低。为了保持更低的稳态误差，利用平滑因子更新子带步长，在仿真实验结果中证明了算法的优异性能。

在一些较复杂的情况下，制定一个自适应滤波算法进行回声消除是无法满足人们的需求，最近几年，为了更好地解决快速收敛和低稳态误差之间的矛盾，许多学者对组合滤波器回声消除进行了研究^[71-76]。Chien 等人将稳健凸组合方法应用于回声消除问题，根据最小均方误差准则，推导出双分量组合自适应滤波器的最优组合因子，然后提供了一个逼近不可及最优组合因子的方法。最后，将此近似因子推广到一个稳健

因子,使其适用于强环境脉冲噪声和回声信道突变的情况。实验仿真结果表明,该方法在归一化均方误差和回波回波损耗增强方面优于同类算法。

除了采用自适应滤波算法外,还有一部分学者采用了神经网络算法、小波算法、WebRTC 及麦克风阵列回声消除等算法。例如, Hyeji^[77]利用多层深层结构的深神经网络(DNN),提出了一种基于回归的声回波与背景噪声综合抑制算法。Michael 提出一种在时频域中工作的算法,假设只有在近端信号不活跃的频率下才能估计各自的传递函数,利用混合信号训练的深层神经网络检测近端信号的活动性,在没有检测到活动的频率中,使用传统的频域最小二乘法估计声学传递函数^[78]。Mohammd^[79]基于多约束优化准则,建立了小波变换 LMS 自适应算法。姚力和刘强^[80]提出了一种基于 WebRTC (WebReal-TimeCommunication) 开源框架的回声消除算法。刘德全采用麦克风阵列进行回声消除和语言增强^[81]。

上述算法都是对声学回声消除理论算法的研究,在实际的应用场景中,理论算法要与硬件进行结合,是以硬件为基础进行应用推广的。近些年来,不仅仅涌现出大量回声消除算法,硬件方面也是发展极其迅速。在美国,全球领先的半导体 TI 公司提出了可以在同一时间消除 512 个通道回声信号。位于蒙特利尔的 Octasic 科技公司研发出一款 OTC9600 在回声消除应用中性能表现优异。在国内,科大讯飞生产的六麦环形麦克风阵列在机器人、智能家居等一些电子产品中应用实现良好的回声消除及语言增强。此外,如百度、腾讯、华为等国内的一些科技公司都有自己的回声消除硬件产品^[82]。

综上所述,回声消除方法已经发展了许多年,各种各样的算法层出不穷,在实际的回声消除应用中,自适应滤波算法仍然是众多学者的首选。虽然 LMS 自适应滤波算法、NLMS 和 NSAF 这些传统的自适应滤波算法结构简单,但是它们的共同缺陷就是拥有固定的步长因子导致快速收敛速度和较低的稳态误差不能同时满足。可变步长因子的引入又会带来参数的干扰导致在实际应用中系统不够稳定。无论哪种自适应滤波方法都会面临着快速收敛和较低稳态误差及计算复杂度之间的矛盾。

1.2.3 语音文本匹配技术研究现状

语音文本匹配技术涉及到语言识别和文本匹配两个领域,在国外上世纪 80 年代, Joseph 和 Kathleen 两位学者就对语音文本匹配技术进行了研究^[83]。在上个世纪,语音识别技术还不够成熟,在处理语音信号时需要对语音数据进行人工分割与标注,整个任务是相当繁冗的,由于人为干扰因素的存在,精度也不是非常的高^[84]。由于近些年来语音识别的技术日益成熟,语音文本匹配技术基于语音识别技术进行语音文本匹

配,既人们利用语音识别系统将语音转换成文本与参考文本进行匹配^[85]。所以语音文本匹配技术的研究重点逐渐转移到文本匹配技术的研究,文本匹配技术也在语音控制中得到广泛的应用。文本匹配技术可以应用到多个自然语言处理任务中,例如:信息检索^[86] (Information retrieval)、机器翻译^[87] (MT)、自动问答^[88] (Automatic question answering) 和复述问题^[89] (Rehearsal question) 等等。在这些自然语言处理任务中,信息检索与机器翻译是人们最经常使用的^[90]。

文本匹配技术最早在国外开始进行研究的,从字面意义上可以知道它是计算两个文本的相似度。相似度问题中最难处理的就是如何在语义方面上找到一个文本匹配算法建立一个最优匹配模型。在中文中有非常的多同义词、多义词,这些词语在不同的语言环境中表达的意思也是不一样的。例如,“这苹果多少钱?”,这句话在水果市场里面的大概意思就是“这苹果多少钱一斤? ”。然而在电子市场里阐述这句话时,它往往表达的是“这个电子产品多少钱? ”。在文本匹配中,语义层面的匹配是其中的难点,它主要有向量空间文本匹配技术^[91-95]、矩阵分解文本匹配技术^[96-100]、主题模型文本匹配技术^[101-105]和深度学习文本匹配技术^[106-112]。

上世纪中后期,Salton^[113]等人提出了向量空间模型,并且成功的将其应用到文本检索系统当中,获得了显著的效果。该算法把文本内容匹配简单地转换成空间向量运算,从而以空间的相似度获取文本的相似度。

马玉玲^[114]针对 XML 文档不准确匹配搜索存在效率不够高、准确率较低等一些问題,提出了一种改进的向量空间 XML 文档搜索方法,该方法用文档树的表现形式将 XML 文档表示出来,且把 XML 文档映射为矩阵空间。利用矩阵变换缩小文档的搜索空间提升了文档的搜索效率,通过仿真实验验证,该改进的方法比传统的向量空间文本匹配方法具有更高的搜索效率和准确率。

李海荣^[115]提出了一种改进的基于向量空间文档语义分类算法解决传统文档分类和语义分类问题,该算法将原文档的高维向量空间进行降维处理,简化了搜索模型从而大幅度提高文档分类和语义分类的性能,提高了文档搜索效率。

张佩云和陈传明等人^[116]利用了语义词典和统计方法的长处,提出了一种文本匹配模型来降低文本向量维度。该算法基于子树匹配的文本相似度对原文本特征向量与子树进行文本匹配。通过仿真实验表明,新算法提高了文本相似度计算的准确率,并且加强了文本语义理解能力。

由于文档中同一个词在句子或段落共现是文本匹配的一个极其有用的特征,然而这个特征往往被忽略掉。Eriks^[117]利用了共现单词这种特征提出一种向量空间匹配模型对文档进行查询和筛选,通过向量之间的长度得出文档地相似度。通过实验验证了该向量空间模型的优异性能。

Watters 针对检索过程中文档语义和上下文信息, 对传统信息检索方法的进行扩展, 基于称为概念空间的信息结构的检索的替代视图提出了一个逻辑框架来定义一个语义模型, 该语义模型包含了当前的检索模型, 用于使用包含在这样一个概念空间中的知识。一个统一的检索框架或理论允许人们形式化书目检索的语义, 以提供专家系统的功能需求^[118]。

Hubert 等人^[119]基于跨语言组合分布语义的最新进展, 提出了一个基于 web 的交互式写作辅助系统。在给定日语或英语查询的情况下, 可以从高质量的英语语料库中检索语义相关的句子。通过使用跨语言约束的向量空间模型来表示短语, 自然地避开了直接的词与文本匹配所带来的一些困难, 并且能够提供新的功能, 如可视化语言间和语言内短语之间的语义关系。

向量空间文本匹配时用高维稀疏向量代替文本内容, 结构简单且易懂, 现在仍然有许多检索系统采用该方法。但是向量空间模型的维度会随着字典的增大而增大, 并且处理同义词和多义词的问题效果不佳。为了解决向量空间维度问题, 在上世纪 80 年代 Di^[120]将高维稀疏矩阵分解为两个低维矩阵且这两个矩阵包含文本中的语义信息。这种矩阵分解文本匹配方法很好地解决了向量空间高维的问题, 但是该方法的物理含义比较模糊。为了解决此问题, 基于主题模型的文本匹配方法出现了。该模型摒弃了矩阵分解, 为了解决多义词的分布问题采用了概率分布^[121]。

Brants^[122]提出了一种新的基于主题的文档分割方法, 即识别不同主题的文档部分之间的边界。该方法将概率潜在语义分析 (PLSA) 模型与基于相邻块对间相似度值的分割点选择方法相结合。通过组合同一模型的不同实例, 使用不同的随机初始化或不同数量的潜在类, 提高了分割性能。常用数据集的结果明显优于其他最新系统的结果。

Kong^[123]提出了一套新的方法来探索嵌入在口语文档中的主题信息, 并将这些信息用于口语文档的自动摘要。通过引入一组潜在的主题变量, 概率潜在语义分析 (PLSA) 有助于发现文档和术语之间潜在的概率关系。并且在 PLSA 模型的基础上, 提出了两种有用的度量方法, 即主题意义和术语熵, 来确定对文档有重要意义的术语和句子, 从而可以用来构建摘要。对普通话广播新闻故事的初步测试结果表明, 与现有的一些方法相比, 该方法的性能有所提高。

基于主题模型文本匹配的物理含义比较清晰, 而且很好地处理了同义词、多义词的问题。在文本匹配时, 基于主题模型可以采用度量方式, 也可以使用引入 KL 等度量分布, 而且应用广泛。但是基于主题模型文本匹配对短文本处理效果不好, 需要引入额外的设置参数等一些问题。

以上这些文本匹配算法都是相对比较传统的方法, 它们都是对一段文本进行概率

的预测,其语言模型就是要求概率的大小,这种语言模型结构简单,但是无法解决人类历史上所有的语言样本这样巨大的数学统计。然而基于深度学习文本匹配算法解决了大数据集问题,2013年,Tomas Mikolov 发表了两篇划时代的论文,这对 NLP 来说是一个时代的进步。第一篇首次提出了两个模型 CBOW 和 Skip-gram^[124]来表示大数据集词的连续向量,通过实验仿真证明该算法精度有了很大的提高,计算成本大大降低,并且证明了这些向量在测试集上提供了最先进的性能,可用于测量语法和语义词的相似性。第二篇主要是阐述了几种优化训练的模型,并提出了一种结构简单的文本短语查找模型,并且通过实验验证了学习百万量级短语的向量表示是可行的^[125]。随后 Tomas Mikolov 提出了 word2vec 工具,它的出现得到了广泛的关注并且极大的促进了自然语言处理的发展,也让更多学者研究深度神经网络在自然语言处理的应用,因为该工具同时将文本匹配的模型和算法都进行了优化。随后,有大量的 word embedding 方法涌现出来。

Levy 和 Goldberg^[126]分析了 Mikolov 等人提出的带负采样的 skip-gram (SGNS) 嵌入方法,并证明了它是隐式分解一个词上下文矩阵,然他们提出了一种嵌入方法 (NCE),该方法是隐式分解一个相似的矩阵,其中每个单元格是给定上下文的单词的对数条件概率。通过实验仿真证明使用稀疏移位的正 PMI 词上下文矩阵来表示词,可以改善两个词相似性任务和两个类比任务之一的结果。

由于自然语言处理系统都依赖于单词嵌入作为基本特征,而单词嵌入以前是在大型语料库上以无监督的方式训练的。然而,为更大的文本块获取嵌入并没有那么成功 Alexis 等人^[127]展示了如何使用斯坦福自然语言推理数据集的有监督数据训练的通用句子表示在广泛的传输任务中始终优于无监督方法,如 skipthought 向量。

在 2018 年 Subramanian 等人^[128]发现无监督方式训练的词的分布式量表示应用到学习单词序列的表示仍然是存在问题的,他们提出了一个简单、有效的句子表征多任务学习框架,在一个单一的模型中结合了不同训练目标的归纳偏差。通过大量的实验表明,在弱相关任务之间共享一个递归语句编码器,可以比以前的方法得到一致的改进。

上面的方法都是一个共同点,获取句子的表征然后在新的分类任务进行训练。为了训练模型更简单,Salesforce 的 Bryan McCann 等人^[129]只取其中的 Embedding 层和 Encoder 层作为输入,他们用一个深度 LSTM 编码器从一个注意序列到序列模型,训练机器翻译 (MT) 中上下文的词向量。实验表明,添加这些上下文向量 (CoVe) 比在各种常见的 NLP 任务中仅使用无监督的词和字符向量提高了性能:情感分析 (SST, IMDb)、问题分类 (TREC)、蕴涵 (SNLI) 和问题回答 (SQuAD)。对于细粒度的情感分析和蕴涵,CoVe 将基线模型的性能提高到最先进的水平。

后来的学者对输入层和输出层进行通用的设计，只要中间层的 Transformer 表征能力足够强在 NLP 任务处理效果可以足够好。Subramanian^[130]等人提出了 GPT，他们使用抽取摘要粗略地识别显著信息，并使用 neuralive 模型生成文章，该模型只引入了一种只支持解码器的体系结构，这种体系结构可以伸缩地处理非常长的序列，比用于序列转换的典型编解码器体系结构要长得多。通过实验表明，在给定参考文献的情况下，可以证明它能够提取出困惑和人为评价中反映的相关事实信息。随着 GPT 的出现，在处理多项 NLP 任务中效果获得一致好评。

然而仅仅过了四个月，BERT 的出现轰动了整个自然语言处理界，它可以被称为划时代的产物。Google AI 的 Jacob Devlin^[131]等人提出了一个新的语言表示模型 BERT，它利用了双向的 Transformers。对 BERT 提供一个额外的输出层就可以微微的改变预先训练的 BERT 模型，就可以建立新的应用模型不用对特定于任务的体系结构进行大量修改。BERT 理念容易理解，功能强大，许多人都用“一切过往，皆为序章”来形容 BERT 的出现。之后又众多的学者都对 BERT 进行了大量研究并发表了大量的文章^[132-137]。

随着深度学习的逐渐盛行，处理 NLP 任务的方法不断更新，文本匹配是 NLP 处理的核心问题，信息检索、机器翻译、自动问答和复述问题这些多项 NLP 处理任务中是相互关联的，方法是可以互相借鉴的，然而每个 NLP 任务中也有自己的独特之处。例如语音文本检索，它只是一般的关键词，结构简单不具有完整的语法结构，文本长度较短。在语义匹配方面也不太相同，在语义文本检索中并不是两个文本片段整体语义上的匹配，而是两个关键词匹配即可^[138]。

1.3 研究目标与研究内容

1.3.1 研究目标

本研究针对语音控制多媒体播放机器人降低环境噪声、回声干扰和语音控制问题，在对相关理论和方法比较分析的基础上，结合本研究的应用需求，研究基于小波包阈值语音降噪、自适应滤波回声消除算法和文本匹配技术，降低环境噪声干扰和消除回声并能正确理解语音指令。实现通过语音指令播放出视频和控制多媒体播放机器人，最后设计并完成语音控制多媒体播放设备原型系统。

1.3.2 研究内容

(1) 小波包阈值降噪算法研究

针对语言信号和噪声信号的特性，分析研究了小波包变化分解系数的特点，并对

小波包软阈值函数和硬阈值函数进行研究。针对两种阈值函数的优点和缺点,基于硬阈值函数对其进行改进,保持了软阈值函数在端点的连续性,同时也保持硬阈值函数的不失真性。通过语音降噪仿真实验,验证本文提出改进硬阈值函数小波包语音降噪算法的优异性。

(2) 自适应回声消除算法研究

针对语音控制多媒体播放机器人回声干扰的问题,研究回声信号的种类及产生机理,通过改进的NSAF算法构建自适应回声消除滤波器,保证其较快收敛的速度和较低的稳态误差优点,准确地模拟出当前环境的回声信号。最后进行回声消除仿真实验,对实验结果进行分析及对比。

(3) 语音文本匹配研究

针对如何让多媒体播放机器人正确理解语音指令的问题,将语音短语命令对应的文本信息和机器人语言控制指令文本信息进行文本匹配,构造文本树和查询树,生成查询中的祖先后代关系,并将其映射到语音指令文本中。根据局部匹配的检索方法,采用矢量检索方法计算语音指令文本并提交查询之间的相似度。将相似度从大到小进行排序,使文本树依次输出语音指令文本信息。

(4) 语音控制播放多媒体机器人原型系统搭建

在上述研究的基础上,设计并完成语音控制多媒体播放体机器人原型系统,实现语音控制功能及人机对话功能,实现语音播放器开始、暂停、声音大、声音小等动作,并且可以语音控制机器人的行走,使人机交互更加人性化。

(5) 语音控制播放多媒体机器人原型系统验证

对语音控制多媒体播放机器人原型系统进行实验验证,主要对语音控制系统进行了语音控制训练实验、个性化语音识别率实验及环境噪声干扰实验,分别测试语音指令执行成功率、语音控制系统针对不同人的语音识别率、环境噪声抗干扰能力、回声路径变换适应力及语音文本匹配能力,通过实验的对比分析验证语音控制播放多媒体机器人原型系统的优异性。

1.4 技术路线与论文组织结构

1.4.1 技术路线

本课题根据研究内容和目标,制定技术路线如图 1-1 所示。

通过国内外文献调研与分析,了解和掌握相关的语音信号进行特征模型研究,解决语音控制应用中环境噪声、回声干扰和语音理解的难题,并且调研了目前控制相关理论和方法的优势和不足之处,进行对比分析,为后续研究工作奠定基础。

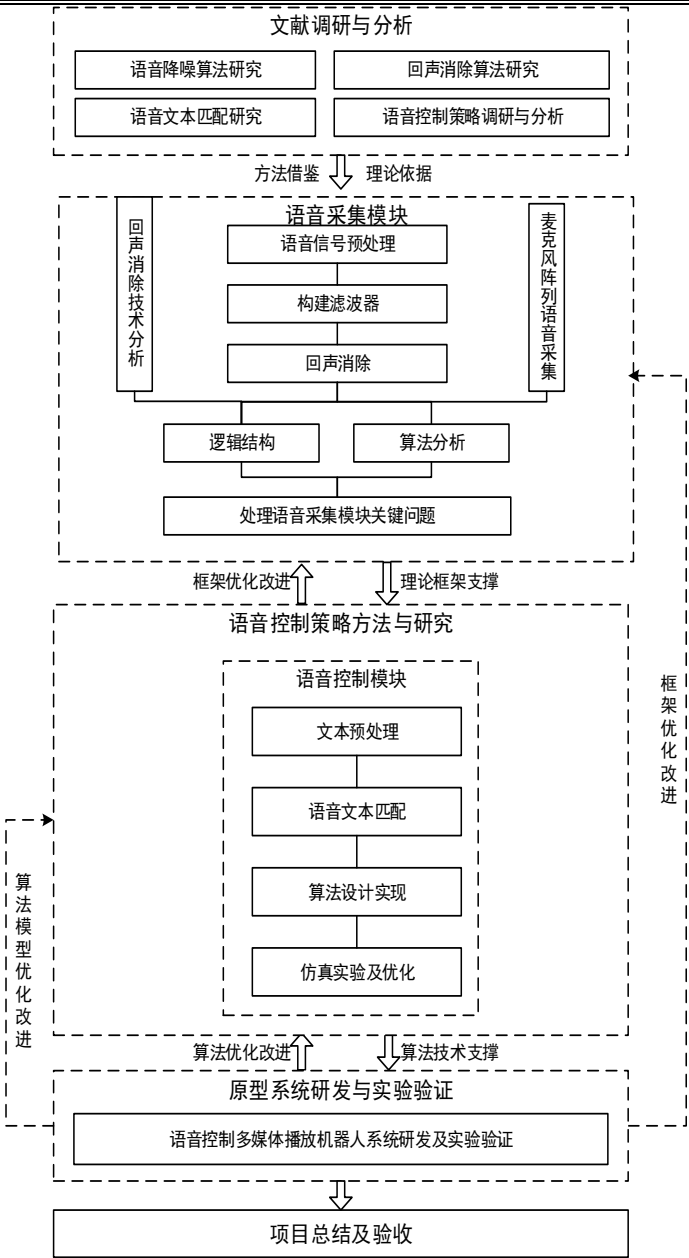


图 1-1 技术路线图

1.4.2 论文组织结构

本文一共有六章内容，分别是“第一章绪论”、“第二章语音控制播放机器人降低环境噪声研究”、“第三章回声特征与回声消除方法研究”、“第四章文本匹配在语音控制应用方法研究”、“第五章语音控制播放机器人原型系统搭建与实验验证”、“第六章总结与展望”。

第一章简要描述了语音控制多媒体播放机器人基本概念和特点，以及智能语音技术对于未来生活的研究背景和意义。同时也给出了本文研究内容的国内外研究现状和

本文的研究目标和内容,介绍了语音降噪技术、回声消除研究技术和语音文本匹配技术,并且通过研究目标和研究内容,制定了研究方法和技术路线,最后给出了文章结构。

第二章通过对语音信号特性和降噪算法的研究,了解了语音信号的模型和噪声信号的模型,并且掌握了一些常用的语音降噪算法。在本章中,提出了改进的语音降噪算法,通过改进硬阈值函数小波包语音降噪算法进行语音降噪,保留了软硬阈值函数的优点,在文章的最后对算法进行了仿真实验。

第三章主要对回声特征和回声消除算法进行了深入研究与分析。其中,对几种常见的自适应算法进行了详细的推导与介绍。并对NSAF算法进行了改进,同时保证了改进算法的较高收敛速率和较低稳态误差,然后对改进回声消除算法进行了仿真验证实验,对结果进行了详细的分析与比较。

第四章主要针对语音文本在语音控制中应用方法进行了深入研究。首先对文本匹配的方法进行了研究与概述,了解了多种传统文本匹配模型,然后提出了基于局部匹配的数字化文本信息检索的方法,根据局部匹配的检索方法,采用矢量检索方法计算语音指令文本并提交查询之间的相似度。将相似度从大到小进行排序,使文本树依次输出语音指令文本信息并对其结果进行了分析与比较。仿真实验表明该方法检索效率高,检索结果全面准确,可用于文本信息的快速检索。

第五章通过对第二章、第三章和第四章的研究,进行了语音控制多媒体播放机器人原型系统的搭建,在本章里介绍了语音控制多媒体播放机器人的功能和结构,并对该系统进行多组整体实验验证。

第六章总结了本文的研究内容与创新点,并且对未来的研究方向与发展进行了展望。

1.5 本章小结

本章主要通过语音控制多媒体播放机器人的调研与研究引出了研究的背景和意义,通过大量阅读文献阐述了语音降噪技术、回声消除技术、语音文本匹配技术的国内外研究现状,并从中找到现有方法的不足之处,并且根据国内外研究现状在本章中制定了本文的研究目标和研究内容。通过调研文献、相关技术的比较分析及仿真实验的验证,制定了本文的技术路线,在本章的最后给出了本文的组织结构。

第二章 语音控制机器人环境噪声降低算法研究

本章对语音控制多媒体播放机器人的去噪算法进行了研究,通过对语音信号模型与降噪算法的分析,基于小波包变化函数提出了改进阈值函数小波包语音降噪方法,并进行了实验验证与分析,为后续章节语音控制多媒体播放机器人回声消除进行了语音预处理。

2.1 引言

在人们的生日常活中,周围常常伴随着各种各样的噪音,比如非平稳的汽车鸣笛声、键盘的敲打声、以及人们吵闹声等。不同的噪声在不同的路径上,产生的效果又是不一样的。所以,一个更加准确、迅速的语音降噪算法来是人们日常生活中所需要的。噪声大概可以分为加性噪声、乘性噪声和周期噪声。在语音控制智能设备中,语音信号是指人类发声器官发送出来的语音信号,它具有非平稳性和随机性的特点。人类的声带与声道决定了语音信号的特性,所以说不同年龄、不同性别的人语音信号的特性是完全不相同的。

为了更好的获得纯净语音信号,许多的降噪算法已经被国内外学者提出来。在上世纪 60 年代,就已经对语音降噪算法进行了研究。语音降噪技术在语音信号处理的过程中是十分重要的,语音降噪技术的好坏决定了语音信号处理的成败。在将近半个世纪的研究中,神经网络降噪算法、麦克风阵列降噪算法、维纳滤波算法、最小均方误差估计法、小波及小波包降噪算法已被提出^[2-9]。

小波包变换作为小波变换的提高和改进,可以对信号在不同的层次上进行分解,每个层次上的分解系数代表着信号在各个分辨率上的信息,而且纯净语音信号和噪声信号在每个小波包分解层上根据小波分解系数会表现出不同的分布特点。所以通过一个阈值对带噪语音信号的小波包分解系数进行划分,保留比这个阈值大的小波包系数,去除比这个阈值小的小波包系统,根据阈值函数的划分重构原始语音信号。但是由于语音信号的复杂性及非平稳性,固定阈值不能满足语音信号去噪的要求,许多学者通过自适应阈值的方法进行小波包语音降噪去解决语音信号的非平稳性及复杂性。

本章提出了一种改进的自适应阈值降噪的小波包降噪算法,对改进的算法进行了推导与分析。并根据语音降噪的评价方法对提出的算法进行验证与仿真,最后对结果进行比较与分析。

2.2 语音信号特性及降噪算法

2.2.1 语音信号特性

语音信号的传递方式主要靠声波进行传播^[138]，声波的震动方向就是语音所传播的方向，声波的波长与频率是成反比的^[139]。语音信号是一种非平稳的并且具有较强的随机性，所以人们对语音信号进行处理时往往都会做出一个假设，它在短时间内是一个平稳的信号，由众多的微小平稳语音信号构成一个完整的非平稳语音信号，这样人们可以更好地处理语音信号^[140]。

语音信号一般具有两大特性：随机性和非平稳性。由于语音信号具有随机性的特点，人们一般采用数理统计的方法对其进行建模与分析。此外，对于语音信号的非平稳性特点，人们会采用上述假设的方法，假设语音信号是暂时平稳的，将语音信号分割成多段处理，进行语音信号分帧处理，最后进行数据统计。

2.2.2 噪声信号特性

噪声信号是指人们日常生活中不需要的信号，也可以认为是打扰人们工作、看书和生活一种信号，它并不是指某一类单一的信号。比如说，音乐可以让人们心情舒缓及身体放松起来，但是在类似考试这种特定的场合中音乐可能就是一种不需的信号，它会打扰到考生做题的思路。也有一些信号从本质上来讲就是噪声信号，例如大型机器的摩擦声、装修中的电钻等声音信号。噪声信号可以分为三大类，它们分别为加性噪声、乘性噪声及周期噪声。

加性噪声根据定义可以理解为噪声信号和纯净语音信号是在时域频域上为互相叠加的，两种声音信号相对比较独立。在人们的生活中，加性噪声随处可见，比如说在工地上人们敲打金属发出的响声，在办公室内人们敲打键盘的声音及电脑主机发出的响声，在室外汽车和火车的喇叭声等等常见的物理声响都是属于加性噪声。加性噪声也常常被叫做系统背景噪声，在实际的信号采集过程中发现语音信号的幅度值和声音信号的强度为线性关系^[141]，所以说加性噪声和语音信号是相加的关系。它的来源主要有三大方面：一种是人为制造的噪声，还有一种是自然界产生的噪声，最后一种就是内部噪声。在这些噪声信号中有些信号是已知的，在理论上这些加性噪声是可以从带噪语音信号中消除的。

乘性噪声是指噪声信号在时域上与语音信号以卷积的形式表现出来，在频域上与语音信号以乘积的方式表现出来，也可以称作为非加性噪声，卷积噪声属于乘性噪声的一种^[141]。在人们的生活中，乘性噪声一般是由信道引起的，无线电通道传输中表

现比较明显,比如麦克风采集的语音信号与传输过程中会产生乘性噪声。人们一般处理乘性噪声信号时都采用滤波处理,可以有效的去除乘性噪声。

周期噪声根据定义可以理解为具有一定的规律性,不是随机的噪声,从信号处理方面可以认为周期噪声信号是在频域上表现出一定的周期性并且从频谱图中可以呈现出来。人们的日常生活中常见的周期噪声信号主要来源于具有周期性的机械设备(风扇、汽车发动机和电机等),一些电气设备的干扰也会引起周期性噪声。然而实际生活中的周期噪声信号并不是完全的理论周期信号,也可以把它当作一个不会简单的有线谱分量组成,而是由大量的窄谱构成的,并且不会是定常的^[141]。

2.2.3 语音降噪常用算法

假设噪声信号用 $b(n)$ 表示,纯净的语音信号用 $s(n)$ 表示,那么带噪语音信号 $x(n)$ 可以表示为:

$$x(n) = s(n) + b(n) \quad (2-1)$$

其中 n 代表的是时间索引。

(1) 谱减法语音降噪算法

谱减法是一种最常使用而且应用较早的一种语音去噪算法^[142-143],因为它运算量较小、复杂度不高且算法简单。该算法在加性噪声信号中发现的幅度谱,通过从带噪语音信号谱中减去噪声信号谱的估计值,从而获得了纯净语音信号谱的估计值。该算法一般是在三种假设条件下进行的,首先,假设所有噪声信号是加性噪声信号,并且语音信号是平稳的与噪声信号是相对独立的。其次,在语音降噪的过程中所有噪声在功率谱上是没有差异的。最后,假设可以利用语言信号幅度谱熵的相位信息得到纯净的语音信号^[141]。在时域上获取纯净语音信号时,将瞬时幅度谱的估计值与噪声信号的相位相结合,然后通过离散傅立叶逆变换变换到时域。但是该算法随着信噪比的降低,降噪后的失真效果也较为明显。

谱减法实现的过程首先将语音信号进行傅里叶变换,将公式 2-1 中三个语音信号变换成频域,表达式如下所示。

$$x(f) = s(f) + b(f) \quad (2-2)$$

其中 $x(f)$ 、 $s(f)$ 和 $b(f)$ 分别是带噪信号 $x(n)$,纯净的语音信号是 $s(n)$ 和噪声信号是 $b(n)$ 的傅里叶变换。在谱减法中,输入信号 $s(n)$ 被缓冲并分成 N 个样本长度的段。利用 Hanning 或 Hamming 窗对每一段进行加窗,然后通过离散傅立叶变换(DFT)将其变换为 N 个光谱样本。窗口减轻了每个段端点处不连续性的影响。加窗函数可以表达为:

$$\begin{aligned}
 x_w(n) &= w(n)x(n) \\
 &= w(n)[s(n) + b(n)] \\
 &= s_w(n) + b_w(n)
 \end{aligned} \tag{2-3}$$

加窗函数在频域中表示为:

$$X_w(f) = w(f)x(f) = S_w(f) + B_w(f) \tag{2-4}$$

谱减法的方程可以表示为:

$$|\widehat{S}_w(f)|^c = |X(f)|^c - \alpha \overline{|B(f)|^c} \tag{2-5}$$

$|\widehat{S}_w(f)|^c$ 是原始语音信号的频谱估计, $|X(f)|^c$ 和 $\overline{|B(f)|^c}$ 是时间平均噪声频谱。假设噪声是宽平稳随机过程。当指数 $c=1$ 时, 公式 2-5 就是幅度谱减法; 当指数 $c=2$ 时, 公式 2-5 就是功率谱减法。参数 α 就是控制从语音带噪信号中减去噪声的量。当 $\alpha = 1$ 时, 它是全噪声减法; 当 $\alpha > 1$ 时, 它是过噪声减法。最后通过离散傅立叶逆变换 (IDFT) 获得语音信号的时域估计, 谱减法框图如图 2-1 所示。

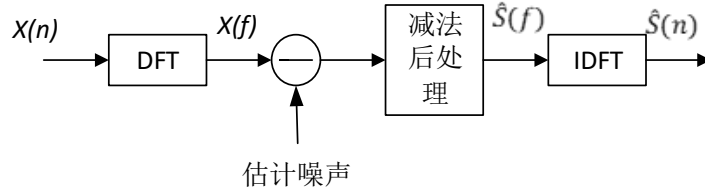


图 2-1 谱减法框图

时间平均噪声谱是当原始语音信号不存在且只有噪声存在的时段获得的, 它的表达式如下:

$$\overline{|B(f)|^c} = \frac{1}{K} \sum_{i=0}^{K-1} |B_i(f)|^c \tag{2-6}$$

$|B_i(f)|$ 是第 i 个噪声帧的频谱, 并且假设仅在噪声周期内有 K 个帧, 其中 K 是可变的。或者, 平均噪声谱可以作为一阶数字低通滤波器的输出来获得, 其表达式如下:

$$\overline{|B(f)|^c} = \rho \overline{|B_{i-1}(f)|^c} + (1 - \rho) |B_i(f)|^c \tag{2-7}$$

其中低通滤波器系数 ρ 通常设置 0.85 和 0.99 之间。对于时域信号的恢复, 将幅度谱估计 $|\hat{S}(f)|$ 与噪声信号的相位相结合, 然后通过逆离散傅里叶变换变换到时域, 表达式如下:

$$\hat{S}(n) = \sum_{k=0}^{N-1} |\hat{S}(k)| e^{j\theta_x(k)} e^{-j\frac{2\pi}{N}km} \tag{2-8}$$

其中 $\theta_x(k)$ 是噪声信号频率 $X(k)$ 的相位。

(2) 子空间语音降噪算法

近些年来, 基于子空间的语音降噪算法被众多学者提出来, 并且在语音降噪的实际应用中得到广泛的认可^[144-148]。这些语音降噪算法都是基于奇异值分解, 它们的核心思想就是将信号空间分解为两个正交的子空间: 信号子空间与噪声子空间。信号子

空间主要有干净的语音信号与少量的噪声信号，噪声子空间全为噪声信号。这种方法的主要目的就是去除噪声子空间从而保留信号子空间，并从信号子空间内提前出更干净的语音信号完成语音降噪。

带噪语音信号可用 $\mathbf{x}(l) = [x(hl - K + 1), x(hl - K + 2), \dots, x(hl - 1), x(hl)]^T$ 表示，纯净语音信号可用 $\mathbf{s}(l) = [s(hl - K + 1), s(hl - K + 2), \dots, s(hl - 1), s(hl)]^T$ 表示，噪声语音信号用 $\mathbf{b}(l) = [b(hl - K + 1), b(hl - K + 2), \dots, b(hl - 1), b(hl)]^T$ 表示，它们之间的关系为：

$$\mathbf{x}(l) = \mathbf{s}(l) + \mathbf{b}(l) \quad (2-9)$$

其中 l 表示帧号， h 表示样本中预先帧大小， K 表示向量维度。在本文中，定义 R_x 为带噪语音信号的自相关矩阵， R_x 的特征分解由下式表示：

$$R_x = U \Lambda_x U^T \quad (2-10)$$

其中 $\Lambda_x = \text{diag}[\lambda_{x,0}, \dots, \lambda_{x,K-1}]$ 是特征值对角矩阵， $U = [u_0, \dots, u_{K-1}]$ 是相应的正交特征向量矩阵。

带噪语音信号按照特征值可以分为两部分，一部分可以划分成 M 个特征值，它是由纯净语音信号和部分噪声信号构成， $U_1 = [u_0, \dots, u_{M-1}]$ ；另一部分可以划分成 $K-M$ 个特征值，它由全被噪声信号构成， $U_2 = [u_M, \dots, u_{K-1}]$ 。那么矩阵 $U = [u_0, \dots, u_{K-1}]$ 变成：

$$U = [U_1 U_2] \quad (2-11)$$

根据子空间的定义，人们习惯把 U_1 称为信号子空间， U_2 称为噪声子空间。由于两个空间具有正交性，可以得到：

$$UU^T = [U_1, U_2] \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} = U_1 U_1^T + U_2 U_2^T = I \quad (2-12)$$

所以可以得到：

$$U_1 U_1^T = I - U_2 U_2^T \quad (2-13)$$

从上式中可以看出将含噪语音信号的向量空间通过洛伊变换（KLT）分解为信号子空间和噪声子空间，通过去除噪声子空间并且从信号子空间中提取出纯净的原始语音信号。

在子空间方法的运算过程中，为了降低子空间方法运算的复杂度，通过 KLT 估计变换出的增益矩阵 $\mathbf{G} = [g_1, g_2 \dots g_k]$ ，其中 k 为向量的维数，最后在通过洛伊逆变换（IKLT）获取纯净原始语音信号，子空间算法结构如图 2-2 所示。

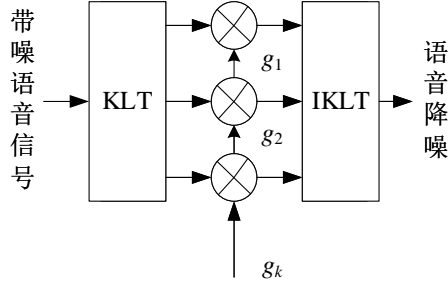


图 2-2 子空间算法基本结构

子空间语音降噪算法可以分成时域约束估计器(TDC)和频域约束估计器(SDC)两种线性估计器。

时域约束估计器是在时域上是语音失真能量达到最小。在本文中定义 R_s 和 R_b 分别为纯净语音信号和噪声信号的自相关矩阵，然后假设 n 时刻带噪自相关矩阵为：

$$R_x(n) = R_s(n) + R_b(n) \quad (2-14)$$

在本文中对原始语音信号进行估计，其估计值为 \hat{s} ，那么误差信号为：

$$e(n) = s(n) - \hat{s}(n) \quad (2-15)$$

由子空间定义可知，误差信号是由语音信号失真 $e_s(n)$ 和残余噪声 $e_b(n)$ 构成，可以表示为：

$$e(n) = e_s(n) + e_b(n) \quad (2-16)$$

当估计语音信号 \hat{s} 与原始语音信号 s 越接近，即 $e(n)$ 越接近零，语音失真就越小。也就说语音信号失真 $e_s(n)$ 和残余噪声 $e_b(n)$ 尽可能的小。在本文中分别定义语音信号失真的能量值 $\overline{e_s}^2$ ，残余噪声能量为 $\overline{e_b}^2$ ，它们的表达式为：

$$\overline{e_s}^2 = \text{tr}\{[H(n) - I]R_s(n)(H(n) - I)^T\} \quad (2-17)$$

$$\overline{e_b}^2 = \text{tr}\{H(n)R_b(n)H(n)^T\} \quad (2-18)$$

其中 $H(n)$ $K \times K$ 维线性估计矩阵。那么在本文中只需求最优估计矩阵就可以得到最小 $\overline{e_s}^2$ 和最小 $\overline{e_b}^2$ 。

本文用拉格朗日法求最优估计矩阵 Hs 。

$$Hs = R_s(n)(R_s(n) + \mu R_b(n))^{-1} \quad (2-19)$$

其中 μ 为拉格朗日乘子。

降噪后的估计语音信号可以表示为：

$$\hat{s}(n) = Hs(n)x(n) = U(n)G(n)U^T(n)x(n) \quad (2-20)$$

其中 $G(n)$ 为 Ephraim 所提出的增益矩， $G(n) = \Lambda_s(\Lambda_s + \lambda\Lambda_\mu)^{-1}$ 。

频域约束估计器是在每个频谱上获得最小信号失真能量。其核心思想是将信号帧分为语音帧和噪声帧两类，频域约束估计器的优化过程与时域约束估计相似，同样采用拉格朗日方法进行获得一个最优估计矩阵 Hs ，最后求得最小信号失真能量。

上面的时域约束估计器和频域约束估计器只适用于白色噪声环境下。如果是有色噪声环境下,需要对噪声信号进行白化,将噪声自相关矩阵 Cholesky 分解 $R_b = R^T R$,得到白化矩阵 R 。

(3) 基于深度神经网络语音降噪算法

近年来,由于深度学习的广泛应用受到了广大学者的关注,它是浅层神经网络学习的一种拓展,并在 2006 年由 Hinton 首次提出^[149],深度学习广泛被应用到图像处理、自然语音处理及语音处理等领域^[150-153]。深度学习方法与其它方法相比,它的特别之处在于面对非平稳信号时仍然具有较强的抗干扰能力,该方法在语音去噪方面的应用取得了巨大的成功。

基于深度学习的语音降噪算法可以分为训练阶段和测试阶段。在训练阶段中,将带噪语音信号的特征和相应的干净语音信号特征分别反馈给深度神经网络的输入和输出节点,并对网络进行优化,最小化降噪后语音特征和干净语音特征的均方误差。训练阶段结束后,通过深度神经网络从带噪语音信号特征中估计干净的语音特征,并采用 GV 均衡后滤波器对过平滑的输出轨迹进行补偿。

在训练阶段,分别从带噪语音信号和相应的干净语音中提取深度神经网络的输入和输出特征。深度神经网络的输入输出特性在输入网络之前通常被标准化为零均值和单位方差。对于输入输出特征,众多学者们提取了噪声和干净语音的对数功率谱^[154-156]。近些年来的研究比较了基于映射的直接估计干净语音的方法和基于掩码的产生二进制或比率掩码目标的方法的性能^[157-159]。一般人们习惯分别用 z_t 和 y_t 表示在 T 帧上的带噪语音信号和纯净语音信号的 F 维归一化对数功率谱。然后,输入特征向量 v_t^0 一般构造如下:

$$v_t^0 = [z_{(t-K)}^T, z_{(t-K+1)}^T, \dots, z_{t+K}^T]^T \quad (2-21)$$

其中K是输入内容扩展参数。

图 2-3 示出了具有三个隐藏层的典型深度神经网络的结构。深度神经网络由一个输入层、几个隐藏层和一个输出层组成,它们与相邻的层完全相连。为了表示简单,假定隐藏层的数目为 L,并且输入层和输出层分别表示为深度神经网络的第 0 层和 (L+1) 层。

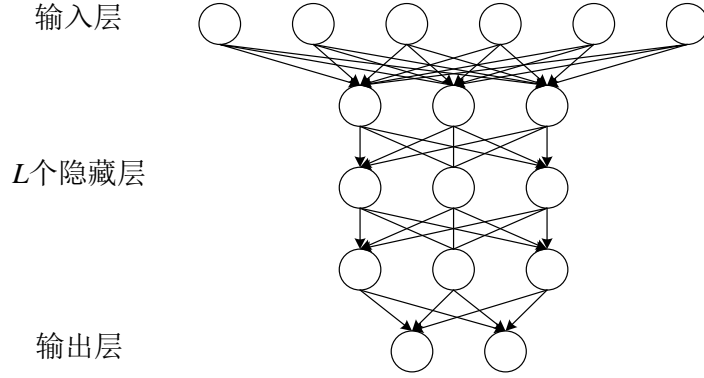


图 2-3 深度神经网络结构图

第 L 层的节点数用 n_l 表示。 n_l 维激活向量 \mathbf{v}_t^l 生成为：

$$\mathbf{v}_t^l = g(\mathbf{a}_t^l) = g(\mathbf{W}^l \mathbf{v}_t^{l-1} + \mathbf{b}^l) \quad (2-22)$$

其中 \mathbf{a}_t^l 表示为 n_l 维预激活向量， \mathbf{W}^l 表示为 $n_l \times n_{l-1}$ 维权重矩阵， \mathbf{b}^l 表示为 n_l 维偏移矢量， g 表示原始激活函数。当所有隐藏层激活因子被计算出来后， F 维输出向量 \mathbf{x}_t 可以表示为：

$$\mathbf{x}_t = \mathbf{W}^{L+1} \mathbf{v}_t^L + \mathbf{b}^{L+1} \quad (2-23)$$

在测试阶段，通过标准的前向处理从 \mathbf{v}_t^0 中获得干净语音估计 \mathbf{x}_t 。在没有 GV 均衡的语音增强算法中， \mathbf{x}_t 被反标准化为 $\bar{\mathbf{x}}_t$ ，如下所示

$$\bar{\mathbf{x}}_t = \mathbf{x}_t \otimes \mathbf{s} + \mathbf{m} \quad (2-24)$$

其中 \mathbf{m} 和 \mathbf{s} 分别是用于规范化深度神经网络的输出特征的平均值和标准偏差向量， \otimes 表示两个向量之间的元素相乘。传统的基于深度神经网络的语音降噪算法的一个显著缺点是，通常会导致降噪后语音的谱轨迹过于平滑。为了缓解这种现象，通常采用 GV 均衡后滤波器来修正 \mathbf{x}_t 的方差以匹配 \mathbf{y}_t 的方差。

在与频率无关的 GV 均衡中， \mathbf{x}_t 和 \mathbf{y}_t 的全局方差计算如下

$$GV(\mathbf{x}) = \frac{1}{TF} \sum_{t=1}^T \sum_{f=1}^F (x_t(f) - \frac{1}{TF} \sum_{t=1}^T \sum_{f=1}^F x_t(f))^2 \quad (2-25)$$

$$GV(\mathbf{y}) = \frac{1}{TF} \sum_{t=1}^T \sum_{f=1}^F (y_t(f) - \frac{1}{TF} \sum_{t=1}^T \sum_{f=1}^F y_t(f))^2 \quad (2-26)$$

基于上面公式，频率无关的 GV 因子 α 可以表示为

$$\alpha = \sqrt{\frac{GV(\mathbf{y})}{GV(\mathbf{x})}} \quad (2-27)$$

在 GV 均衡后滤波器中，将 GV 因子与输出特征相乘可被视为在线性谱幅度域中施加指数因子。通过该滤波器后，谱轨迹的方差随 α 值的增大而增大或减小。

在测试阶段，纯净语音估计信号可以通过前馈处理从中获得。深度神经网络语音

降噪系统流程图如图 2-4 所示。

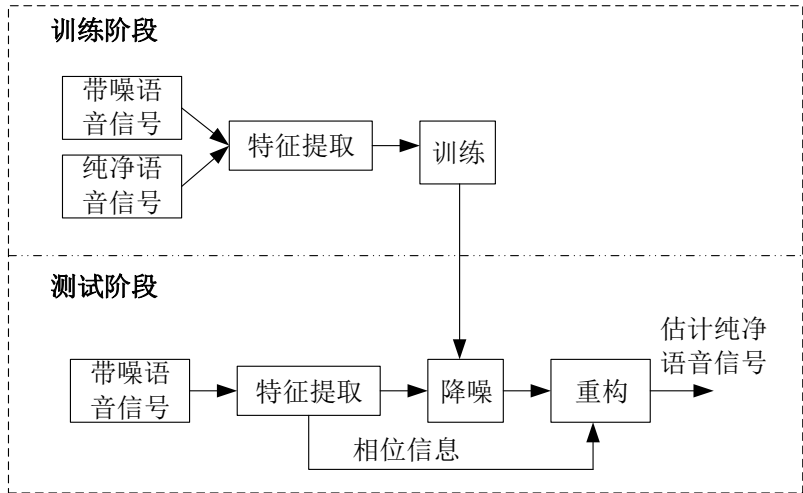


图 2-4 语音降噪系统流程图

如图 2-4 所示，在测试阶段将带噪语音信号变换成特征表示域，通过降噪过程后的信号重构，估计出纯净的语音时序信号。

2.3 改进阈值函数小波包语音降噪方法

Coifman 首先提出了小波包变换的数学工程。由于小波包是从小波变换中推广而来的，因此小波包变换也可以通过滤波器组结构来实现。实际上，构造离散小波变换的有效方法之一是在低通尺度函数分支上迭代一个双通道完全重构滤波器组。这种方法也叫做 Mallat 算法。然而，该算法将产生对数频率分辨率，这对大多数信号（但不是所有信号）都有效。此外，由正则小波变换产生的子带的带宽与心理声学模型的带宽不匹配。为了克服上述缺点，可以迭代 Mallat 算法树的高通小波分支和低通尺度函数分支。这种由任意子带树产生的小波分解称为小波包分解。

基于小波包语音降噪算法可以概括为以下三个步骤^[37]：

- (a) 分解嘈杂语音信号划分小波系数。
- (b) 采用合适的阈值的方法来在 (a) 中获得的小波系数。
- (c) 合成 (b) 中得到的这些阈值小波系数以实现语音降噪。

其小波包降噪原理如图 2-5 所示。

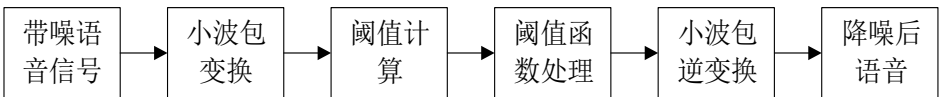


图 2-5 小波包语音降噪原理图

图 2-6 给出了用于两尺度小波包分解和重构的全二叉树的示例。在图中， $h_0(n)$ 和 $h_1(n)$ 分别是分解低通尺度滤波器和高通小波滤波器，而 $g_0(n)$ 和 $g_1(n)$ 分别是合成低

通尺度滤波器和高通小波滤波器。

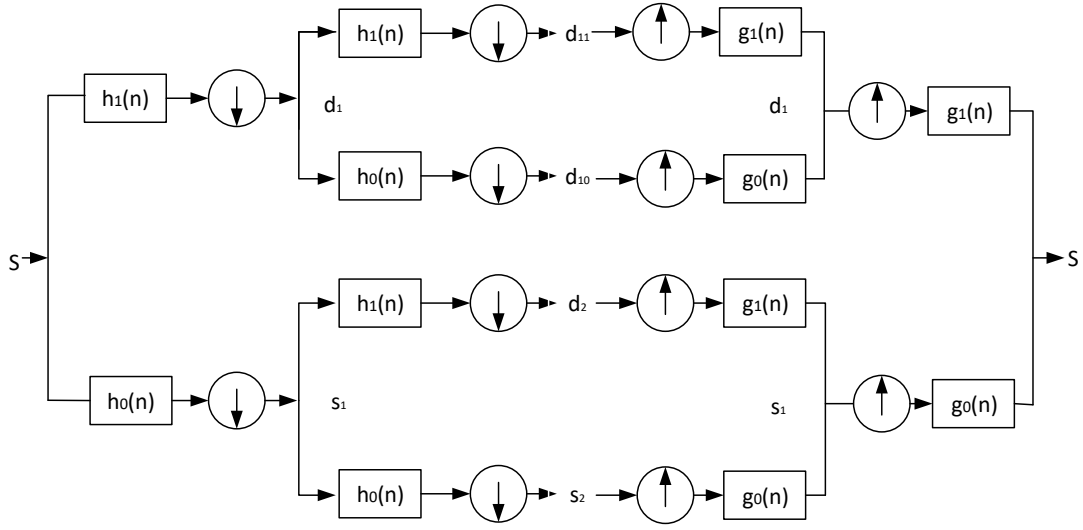


图 2-6 两尺度小波包树变换

$$h_1(n) = (-1)^n g_0(1-n) \quad (2-28)$$

$$g_1(n) = (-1)^n h_0(1-n) \quad (2-29)$$

此外,图2-6中所示的符号 \uparrow 和 \downarrow 分别表示上采样和下采样的过程。如图2-7所示,小波包变换可以由分解和重构单元来组成。

在小波包分解的过程中,人们会发现小波包分解系数的排序并不是按照频率有规律排序的。因为小波包在分解的时候,低通滤波分解是按照频率从低到高排序的,而高通滤波前会出现翻转现象,从而在分解的过程中是按照从高到低排序的,所以小波包分解后会造成频率顺序混乱,需要对分解后的分支要进行重新排序。如果对小波包进行 n 层分解,那么最后分解的层数是 2^n , 假设 $n=3$, 其小波包分解树图如图2-7所示,小波包分解谱特性如图2-8所示。

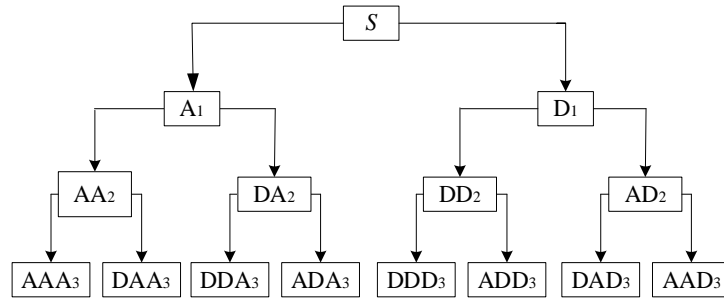


图 2-7 三层小波包分解树图

从图中可以看出三层小波包分解后的自然顺序并不是人们想要的频率顺序,图中 S 表示输入语音信号, A 代表低通滤波, D 代表高通滤波, 数字表示层数。所以在本文中第三层小波进行重新排序, $S = AAA_3 + DAA_3 + ADA_3 + DDA_3 + AAD_3 + DAD_3 + ADD_3 + DDD_3$ 。

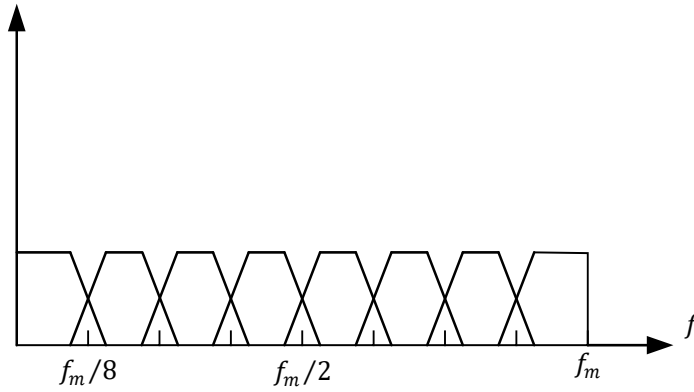


图 2-8 三层小波包分解谱特性图

在小波包语音降噪的时候，阈值的作用是非常大的。由于语音信号的时变性和非平稳性，使用固定阈值会导致语音信号的过阈值。换言之，使用相同的阈值不仅可以抑制不需要的噪声，而且还可以抑制一些类似于清音的语音片段。当阈值过大，纯净的语音信号会被当成噪声处理掉；当阈值过低时，噪音会被当成纯净语音信号。人们常见的阈值选择准则有四种：

固定形式阈值：

$$thr_1 = \sigma\sqrt{2\ln N} \quad (2-30)$$

自适应阈值：

$$thr_2 = \sigma\sqrt{Q_\alpha} \quad (2-31)$$

启发式阈值：

$$thr_3 = \begin{cases} T_1 & \alpha > \beta \\ T_2 & \alpha < \beta \end{cases} \quad (2-32)$$

极小化极大化阈值：

$$thr_4 = \begin{cases} \sigma(0.3936 + 0.1829)\log_2^N & N > 32 \\ 0 & N < 32 \end{cases} \quad (2-33)$$

不同的阈值准则选取的方法是不一样的，每个阈值准则都有自己的缺点与优点。固定阈值与启发式阈值是两个比较相似的阈值准则，它们核心思想就是处理全部小波包系数，这样可以较好的去除噪声，但是也容易将部分纯净语音当成噪声被处理掉，这类方法较适合处理语音信号的高频部分。自适应阈值和极小化极大化阈值两个阈值准则与其它两种阈值准则相比，是一个较为折中的方法。它不像那两种阈值准则容易过度去噪，相反的是它容易去噪不完整，容易将噪声信号当成有用的语音信号。所以遇到不同的情况，选择不同的阈值准则。小波包变化是将信号分成几个不同的频率段，人们可以根据频率范围选取所需要的阈值准则。

在语音信号中，通过对小波包系数进行阈值化进行去除噪声分量，能量主要集中在少数小波包维度上。这些维度的系数与其它维度或任何其他信号（尤其是噪声）相

比相对较大。因此，通过对系数进行阈值化，将较小的系数设置为零，可以从噪声信号中去除噪声成分。基本的降噪算法总结如下：

$$x = s + \text{noise} \quad (2-34)$$

其中 noise 为白色高斯噪声信号。

如果 W 为小波包矩阵，上述方程可以变为：

$$X = S + \text{NOISE} \quad (2-35)$$

其中：

$$X = W \cdot x, S = W \cdot s, \text{NOISE} = W \cdot \text{noise} \quad (2-36)$$

可以从由带噪声语音信号的小波包系数获得估计纯净语音信号 \hat{S} 的小波包系数：

$$\hat{S} = \text{THR}(X, T) \quad (2-37)$$

其中 $\text{THR}(\cdot)$ 代表阈值函数， T 为阈值的标量值。

在基于小波包方法中使用标准阈值函数一般有硬阈值函数和软阈值函数。硬阈值函数的特性就是“取大舍小”，只要比设定阈值小的就舍弃尽管里面包含有用的语音信号。软阈值函数则是将小于预先设定阈值的小波系数置零，并把大于阈值的小波系数向零做收缩。两者的表达公式如公式（2-38）和公式（2-39）所示。

硬阈值函数：

$$\text{THR}_H = \begin{cases} Th, & |Th| \geq T \\ 0, & |Th| < T \end{cases} \quad (2-38)$$

软阈值函数：

$$\text{THR}_S = \begin{cases} \text{sign}(Th)(|Th| - T) & |Th| \geq T \\ 0, & |Th| < T \end{cases} \quad (2-39)$$

其中 $\text{sign}(\cdot)$ 为符号函数，硬阈值函数和软阈值函数效果示意图如下所示。

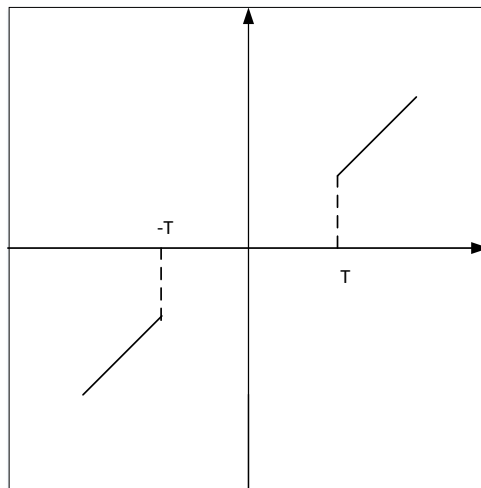


图 2-9 小波包系数的硬阈值函数

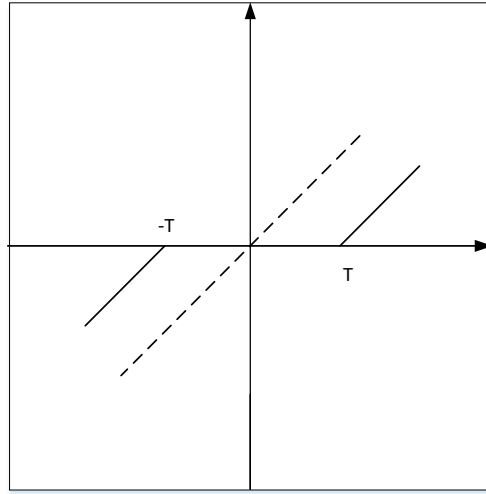


图 2-10 小波包系数的软阈值函数

从上面的图中可以看出，硬阈值方法和软阈值方法都有一定的缺陷。在硬阈值方法中得到小波包系数出现断点，不是连续函数。软阈值方法得到的小波系数虽然连续，但是与实际信号存在偏差，不够准确。所以希望得到一个既连续又与实际信号偏差较小的一个函数，也就是硬阈值和软阈值函数折中的一个函数。

在数学领域可知，任意一个积分值为 1 的光滑函数 $f(x)$ 与一个不连续的函数做卷积运算，可以得到一个光滑连续的函数，并且这个新的函数逼近原不连续函数。根据这个原理对硬阈值函数进行改进，改进的硬阈值函数实现了既连续又与实际信号偏差较小。在数学应用中，改进的阈值函数理想效果就是同时保持信号的真实性和不会对有用信号造成过度损失。

由于

$$\int f(x) = \int \frac{f(cx)}{c} \quad (2-40)$$

所以选取 $f(x) = \frac{e^{-(x)^2}}{\sqrt{\pi}}$ ，从中可以看出 $f(x)$ 的积分值为 1，而且它可以无数次求导，根据公式 (2-40) 本文将 $f(x)$ 变为新的函数：

$$\frac{f(cx)}{c} = \frac{e^{-\left(\frac{x}{c}\right)^2}}{c\sqrt{\pi}} \quad (2-41)$$

本文用硬阈值函数与公式 (2-41) 进行卷积运算，可以得出：

$$\frac{f(cx)}{c} * THR_H = x + \frac{1}{2} \frac{\left(-e^{-\frac{(x+Th)^2}{c^2}} + e^{-\frac{(x-Th)^2}{c^2}} \right)}{\sqrt{\pi}} \quad (2-42)$$

当 c 越趋近于 0 时，新的函数就与原硬阈值越逼近，其改进的硬阈值函数图如图 2-11 所示。

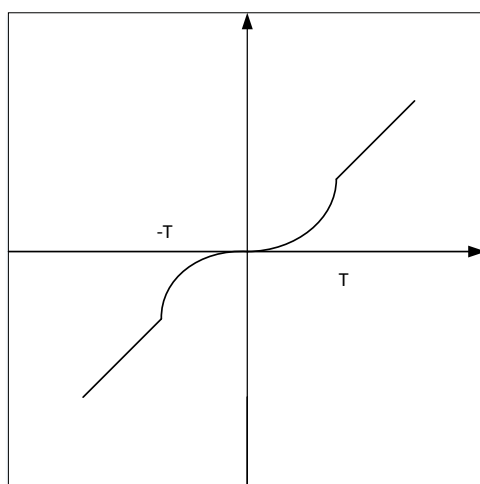


图 2-11 改进阈值方法

从图 2-11 中可以看出,改进的硬阈值函数在端点处是连续的而且对硬阈值函数的缺点进行了改进,在实际应用中它不会对有用的信号造成过度损失。对于小波包分解系数来说,通过改进硬阈值函数处理,都可以更好的保留原始信号的细节部分。

2.4 实验结果与分析

2.4.1 实验设置

为了验证改进的硬阈值函数小波包语音降噪算法的可靠性和优异性,本文对算法进行了语音降噪仿真对比实验,所有实验都是采用了 TIMIT 语音数据库中的语音数据。为了更好的模拟噪声环境,本文在-15dB、-10dB、-5dB、0dB、5dB、10dB 和 15dB 7 种情况下的信噪比,在纯净语音信号中加入噪声信号。在实验中首先加入平稳的白色高斯噪声信号对提出的改进阈值函数方法进行对比实验,与传统的硬阈值函数和软阈值函数进行语音降噪效果比较。然后从 TIMIT 语音数据库中获取非平稳的汽车噪声加入到纯净语音信号中对改进阈值函数方法进行对比实验,仍然和硬阈值函数、软阈值函数进行语音降噪效果比较。在语音信号处理过程中,语音波形的采集频率为 8kHz,每个波长被设定为 256 个样本点。在小波包变化中,小波包采用 db4 小波,分解层数为 4。

2.4.2 评价标准

语音降噪算法的评价标准就是对语音信号进行降噪后给出的一种评价标准,从而评判各种语音降噪算法的性能优异性和不足之处。总的来说,语音降噪算法的评价标准一般分为两大类,第一类主要为主观评价,它主要是以个人感知为主的一种评价方

式，人们通过听觉对降噪效果后的声音进行主观评价；第二类为客观评价，它主要是依靠一些客观的数据进行推理获得的评价标准。一般来说，两类评价标准缺一不可，语音降噪的目的就是让人们或者机器更容易的获取纯净语音信号，主观评价有时往往带主观色彩从而忽略了语音信号的真实性，客观评价往往过于真实忽略了人们的感受。

(1) 主观评价标准

语音降噪算法的主观评价标准一般应用于对声音的品质进行评判，该标准主要是来源于人们的主观感受，通过人们的听觉感受评价声音的优劣状态。一般音响设备播放出的声音品质评价、网络会议的通话质量评价和音乐品质的评价都经常会使用主观评价标准。一般经常使用的主观评价方法有成对比较法、排序法、等级评价法、幅度调节法和语义细分法等^[160-163]。传统的主观评价方法往往是片面的，它是基于个人感受去进行评价，会受到环境等多种因素的影响，而且还需要耗费大量的人力。然而在声音品质方面的评价，客观评价标准又是不够全面的，声音的品质更侧重于人们的听觉感受。综上所述，即使语音降噪算法主观评价标准是片面及武断的具有一定的局限性，但是在声音品质方面的评价，主观评价是不可或缺的一部分。

主观语音质量（PESQ）评价标准就是将降噪后的语音信号文件与原始语音文件进行比较。这种评价标准一般与 MOS 得分相结合，它的分值一般在 1.0 和 4.5 之间，只有当信号失真时得分会小于 0。语音质量越差，得分就会越低。

(2) 客观评价标准

语音降噪客观评价是指语音降噪后，采用一些科学数据对降噪后的语音信号文件进行评价，并且将这些数据量化处理，客观的描述语音降噪后的效果。该评价标准没有个人主观色彩，不会受到人为地影响，且相对主观评价标准更加公平，也减少了大量的人力相对容易实现。但是客观评价标准只是数据的量化体现，往往对语音质量评价会做出错误的判断，缺乏一定的人性化。在客观评价标准方面，人们常常使用信噪比和对数谱失真度等，它们都是针对人们日常生活中的应用需要所提出的。

信噪比（SNR）评价标准根据字面意义可以理解为信号和噪声的比值，其比值的結果越大说明语音降噪后的效果就越好，也就是说明降噪处理后的语音信号与纯净的语音信号越相似。SNR 的评价标准如公式 2-43 所示。

$$SNR = 10\log\left(\frac{\sum_n^N s^2(n)}{\sum_n^N (s(n)-\hat{s}(n))^2}\right) \quad (2-43)$$

对数谱失真度标准是在时域内进行的，是指在时域中原始输入信号的失真度。其结果越小，说明降噪后得到的输出信号和原输入信号越接近，说明其降噪效果好。LSD 计算公式如下：

$$LSD = \frac{1}{M} \sum_{i=0}^{M-1} \left\{ \frac{1}{\frac{N}{2}+1} \sum_{k=0}^{\frac{N}{2}} [10 \log_m |s(k)| - 10 \log_m \hat{s}(n)]^2 \right\}^{\frac{1}{2}} \quad (2-44)$$

在一些实际的语音降噪应用中,人们会根据不同的需求选择不同的语音降噪评价指标。目前,大多数学者都会采用主观评价标准和客观评价标准相结合的方法,可以更合理并且准确的评价语音降噪系统。

2.4.3 结果与讨论

从图 2-12 中可以看出,三种不同阈值函数方法在-15dB—15dB 情况下获得的信噪比改善情况,所有阈值函数方法都可以达到降噪效果,并且随着信噪比的降低降噪效果提升越明显。例如,在信噪比为-15dB 情况下,软阈值函数方法的 SNR 提升了 3dB,硬阈值函数方法提升了 3.4dB,改进的阈值函数提升了 3.8dB。SNR 提升越高说明阈值函数方法降噪效果越好,图 2-12 中表示了改进的阈值函数重构的语音信号的 SNR 有较大的提高,并且语音降噪效果优异于其它两种传统阈值函数降噪方法。

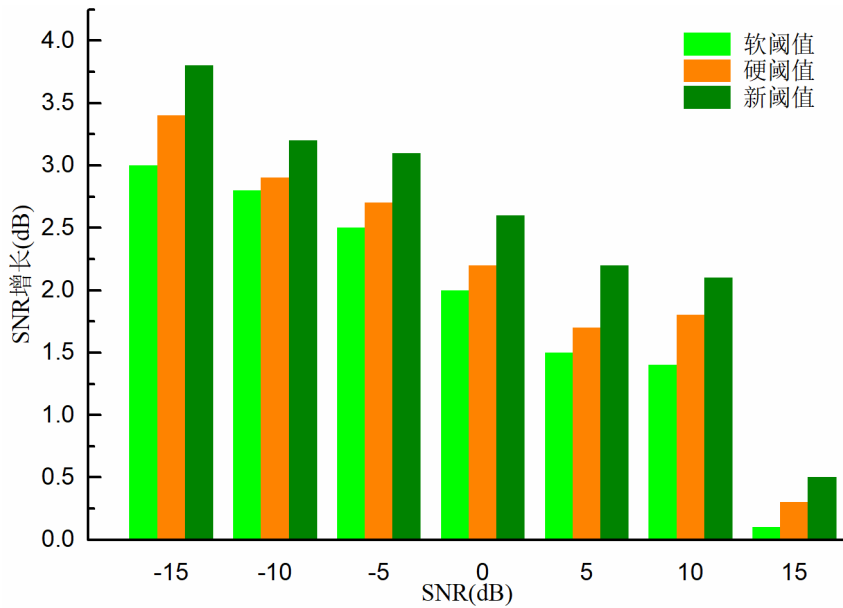


图 2-12 三种阈值函数在白噪声环境下 SNR 增长图

表 2-1 中给出了在白噪声环境下小波包软阈值函数、小波包硬阈值函数和小波包改进阈值函数在不同信噪比下的平均 PESQ 结果。从表中可以看出,在高信噪比的环境下,本文提出的改进阈值函数与传统的两种阈值函数相比,平均有 0.5 个点的 PESQ 提升。在低信噪比的情况下,改进的阈值函数优异性能表现的更加明显,相比其它两种传统阈值函数,平均有 1 个点的 PESQ 提升。通过数据表明,改进的阈值函数不仅在同等环境表现性能更加优异,而且抗噪能力也要强于其它两种传统的阈值函数。

表 2-1 三种阈值函数在白噪声环境下 PESQ 评均值

| SNR (dB) | 软阈值函数 | 硬阈值函数 | 改进阈值函数 |
|----------|-------|-------|--------|
| -15 | 1.14 | 1.15 | 2.24 |
| -10 | 1.21 | 1.25 | 2.37 |
| -5 | 1.34 | 1.48 | 2.57 |
| 0 | 1.42 | 1.56 | 2.68 |
| 5 | 2.55 | 2.78 | 3.12 |
| 10 | 2.69 | 2.85 | 3.45 |
| 15 | 2.78 | 2.95 | 3.81 |

图 2-13 中表示三种阈值方法在白色噪声环境内不同信噪比情况下产生的加权谱倾斜测度 WSS，WSS 值越小说明扭曲越少，越小越好。从图中可以看出，WSS 随着 SNR 的降低而增加。在某一特定的信噪比下，改进的阈值函数方法所获得的 WSS 值要低于其它两种传统阈值函数。例如在信噪比 15dB 的情况下，改进的阈值函数获得 WSS 值为 45，软阈值函数获得 WSS 值为 48，硬阈值函数获得 WSS 值为 60，说明改进的阈值函数在 15dB 情况下优异于其它两种传统阈值函数。即使改进函数方法获得的 WSS 值与软阈值函数方法获得 WSS 值较接近，但是在较低的信噪比情况下，如 -15dB，改进阈值函数获得 WSS 值为 105，然而软阈值函数获得 WSS 值为 125，改进阈值函数方法远远优异与软阈值函数方法。综上所述，改进的阈值函数方法在获取 WSS 值性能优异于其它两种传统阈值函数方法。

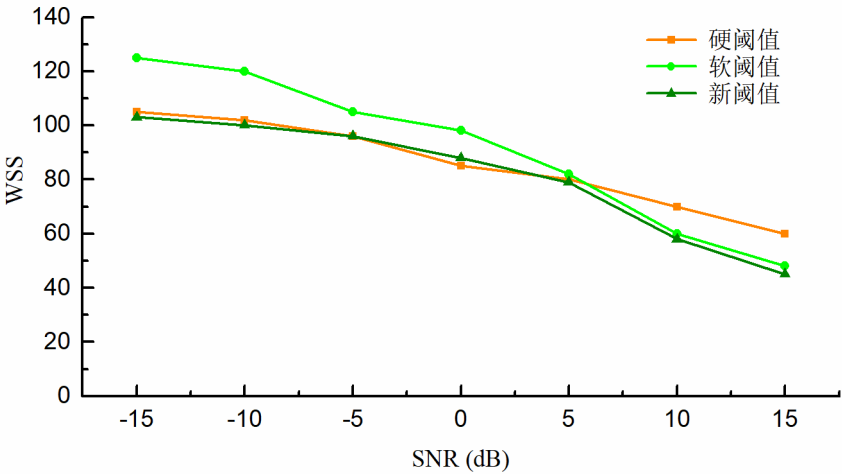


图 2-13 三种阈值函数在白噪声环境下 WSS

在图 2-14 中，在非平稳汽车噪声环境下本文提出的改进阈值函数方法与其它两种阈值函数方法在不同 SNR 水平下进行了比较。从图中可以看出，随着信噪比的降低，信噪比的改善也会增加。在信噪比最低的情况下（SNR=-15dB），本文所提出的

改进阈值函数方法可获得最高的信噪比增加值。在较低的信噪比下，这种较大的 SNR 增加值可以证明所提出的改进阈值函数语音降噪方法对于受到汽车噪声严重污染的语音具有更好的语音降噪能力。这些足以说明改进的阈值函数语音降噪法在面对非平稳噪声噪声环境下具有理想的降噪能力。从图中也可以看出，在特定的信噪比情况下，本文提出的改进阈值函数方法与其它两种阈值函数方法相比，重构的语音信号的 SNR 有更大的提高。例如，在信噪比为-15dB 情况下，软阈值函数方法的 SNR 提升了 3dB，硬阈值函数方法提升了 3.3dB，改进的阈值函数提升了 3.8dB。综上所述，本文提出的改进阈值函数方法在非平稳环境噪声环境下，语音降噪性能优异于其它两种传统阈值函数语音降噪性能。

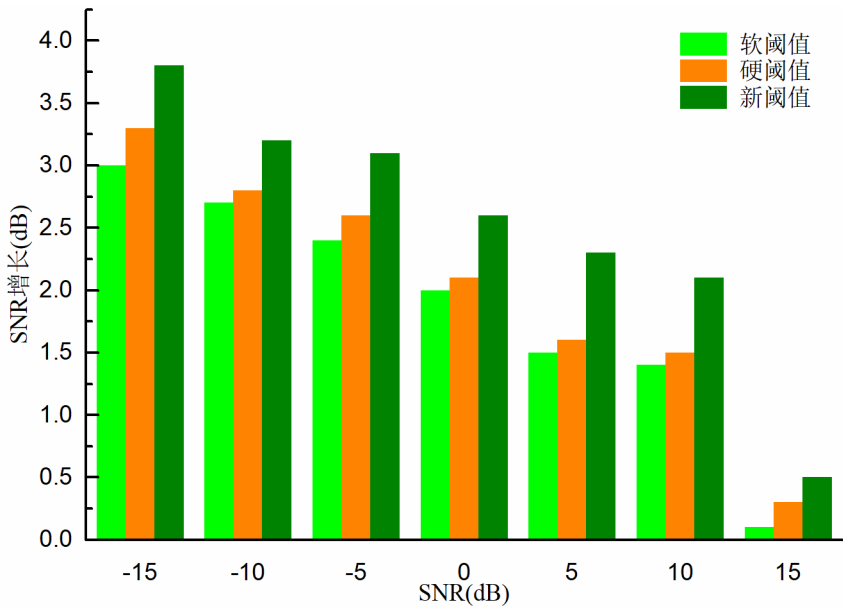


图 2-14 三种阈值函数在非平稳声环境下 SNR 增长

表 2-2 中给出了在非平稳噪声环境下小波包软阈值函数、小波包硬阈值函数和小波包改进阈值函数在不同信噪比下的平均 PESQ 结果。从表中可以看出，在高信噪比的环境下，本文提出的改进阈值函数与传统的两种阈值函数相比，平均有 0.6 个点的 PESQ 提升。在低信噪比的情况下，改进的阈值函数优异性能表现的更加明显，相比其它两种传统阈值函数，平均有 1.2 个点的 PESQ 提升。由于在特定的信噪比情况下，较高的 PESQ 分数表示较好的语音质量。通过数据表明，即使存在车辆噪声改进的阈值函数不仅在同等信噪比下表现性能更加优异，而且抗噪能力也要强于其它两种传统的阈值函数。

表 2-2 三种阈值函数在非平稳噪声环境下 PESQ 评均值

| SNR (dB) | 软阈值函数 | 硬阈值函数 | 改进阈值函数 |
|----------|-------|-------|--------|
| -15 | 1.0 | 1.11 | 2.20 |
| -10 | 1.11 | 1.15 | 2.27 |
| -5 | 1.28 | 1.37 | 2.51 |
| 0 | 2.38 | 2.48 | 3.03 |
| 5 | 2.45 | 2.62 | 3.18 |
| 10 | 2.59 | 2.78 | 3.35 |
| 15 | 2.12 | 2.38 | 2.78 |

图 2-15 中表示三种阈值方法在车辆噪声环境内不同信噪比情况下产生的 WSS。从图中可以看出，WSS 随着 SNR 的降低而增加。在某一特定的信噪比下，改进的阈值函数方法所获得的 WSS 值要低于其它两种传统阈值函数。例如在信噪比 15dB 的情况下，改进的阈值函数获得 WSS 值为 37，软阈值函数获得 WSS 值为 48，硬阈值函数获得 WSS 值为 52，说明改进的阈值函数在 15dB 情况下性能远远优于其它两种传统阈值函数。并且从图中可以看出，本文提出的改进阈值函数方法获取 WSS 值始终低于软阈值函数很硬阈值函数方法，并且平均值保持 10 以上的水平。综上所述，在非平稳噪声环境下，改进的阈值函数方法在获取 WSS 值性能优于其它两种传统阈值函数方法。

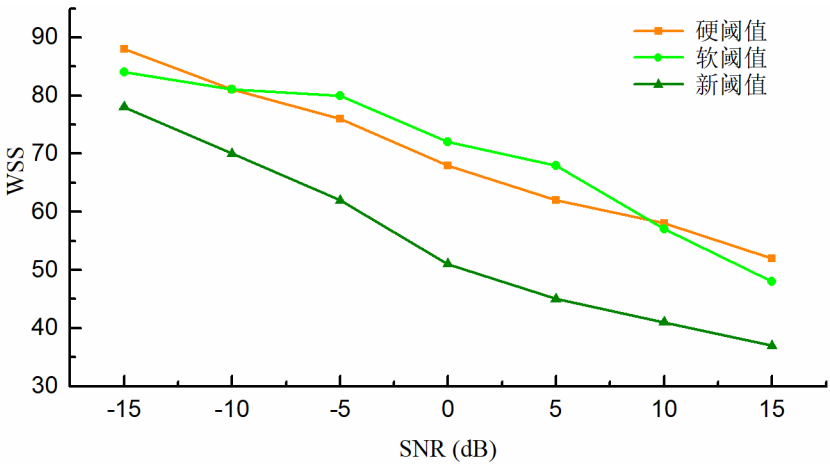


图 2-15 三种阈值函数在非平稳噪声环境下 WSS

通过上面仿真实验可以看出，本文提出的改进阈值函数语音降噪算法不仅仅在平稳的白色噪声情况下能有较好的降噪效果，而且在非平稳的汽车噪音环境下语音降噪效果表现优异。改进的阈值函数语音降噪算法与其它两种阈值函数方法相比，信噪比增加了 0.5dB 以上，PESQ 提高了 0.8 以上，WSS 降低了 10。所以说本文提出的改进

阈值函数语音降噪算法在平稳的噪声环境下和非平稳的噪声环境下的性能都要优于传统阈值函数降噪算法。

2.5 本章小结

本章对语音信号和噪声信号进行了特征分析，并介绍了谱减法降噪算法、子空间降噪算法、有监督降噪算法等传统的语音降噪算法，并且在不同环境下选择不同的语音降噪标准。并根据传统阈值函数的缺点，对硬阈值函数进行改进，在仿真实验中从 SNR、PESQ 和 WSS 三种评价标准进行验证。实验结果表明改进阈值函数语音降噪算法在面对平稳与非平稳噪声环境下，都具有良好的语音降噪性能。

第三章 回声特征与回声消除方法研究

本章主要对自适应回声消除方法进行了研究,首先对回声机理与回声路径进行了简要的概述,并介绍了几种常见的自适应回声消除算法并对其进行推理,通过对传统的回声消除方法进行了修改,提出了改进的可变步长子带滤波回声消除方法,在最后进行了仿真实验验证,为后续章节语音控制多媒体播放机器人语音文本匹配进行了预处理。

3.1 引言

近年来,由于提高语音质量仍然是通信系统最重要的目标之一,因此回声消除已经成为一个极具挑战性的课题。例如:电话会议、免提和移动通信等。它经常被用来消除扬声器和麦克风之间耦合产生的不需要回声,回声消除是自适应滤波器的一个重要应用。它的基本原理是建立一个回声路径脉冲响应信号,并且自适应滤波器可以识别这个模型,该滤波器输出一样的回声信号,并从参考信号中减去回声信号。尽管许多自适应算法理论上都适用于声学回声消除,但是实际应用中的精度和处理能力都是有限的。

因此,在最近几年,声学回声消除(AEC)技术已经吸引了广泛的关注。由于自适应均方(LMS)算法的简单性和鲁棒性,它成为一个应用最早且应用广泛的一种回声消除技术。然而,由于回声路径往往是非常大的,声音输入信号不是静止的,独立的,收敛的效果不是很好。上述问题可以通过标准化固定步长大小,从而提出了NLMS算法来改善。但是固定步长无法解决收敛速度和稳态误差之间的矛盾。当步长因子越大,收敛速度越快,然而稳态误差也随之增加。相反,步长因子越小,稳态误差也就越小,但是收敛速度也会随之减慢^[53]。为了满足更快的收敛速度和较低的稳态误差,许多可变步长的自适应算法已经被提出。一个可变步长算法的核心思想就是人们经常会在初始状态使用一个较大的步长让算法的收敛速度较大,且当误差信号较大时,该算法使用一个较小的步长大小,以减少在稳定状态下的稳态误差。然而,当输入信号是彩色信号时,上述算法的收敛速度会受到极大的衰减。

因此,为了克服彩色信号输入的问题,许多人已经提出了各种子带自适应滤波器(SAF)。虽然SAF可以提高收敛速度,但是由于一个单独的子滤波器在每个子带自适应环路使用,传统SAF的收敛速度是受到混叠和带边缘效应的限制。一种多带(MSAF)的结构中被众多学者提出来^[164-171],它是一种通过使用提取子带信号来解决混叠和带边缘效应更新全频带自适应滤波器。在MASF的基础上, Lee 和 Gan 开发

了归一化的 SAF (NSAF) 算法。但是归一化 SAF (NSAF) 算法仍然不会降低计算复杂度, 越长的滤波器计算度是越复杂的。后来还有学者针对上述问题采用了闭环无延迟的结构解决混叠效应, 因为它是使用最小化全频带的 MSE 代替每个子带的 MES 减少混叠效应^[172-174]。

3.2 回声研究及基本回声消除算法

从回声产生的原因的角度出发, 回声可以分为声学回声和电路回声。声学回声是指设备扬声器的声音多次返回到麦克风所引起的, 大家也可以认为声学在空间内经过反射产生的声音。

从图 3-1 中可以看出, 当说话者 A 的语音信号发送给 B, 当 B 接受到 A 语音信号时会在空气中反射形成了声学回声, 声学回声又会通过 B 的话筒传送给 A, 此时 A 又听到了自己说话的声音。声学回声一般发生在视频会议, 车载电话和免提等一些应用中。

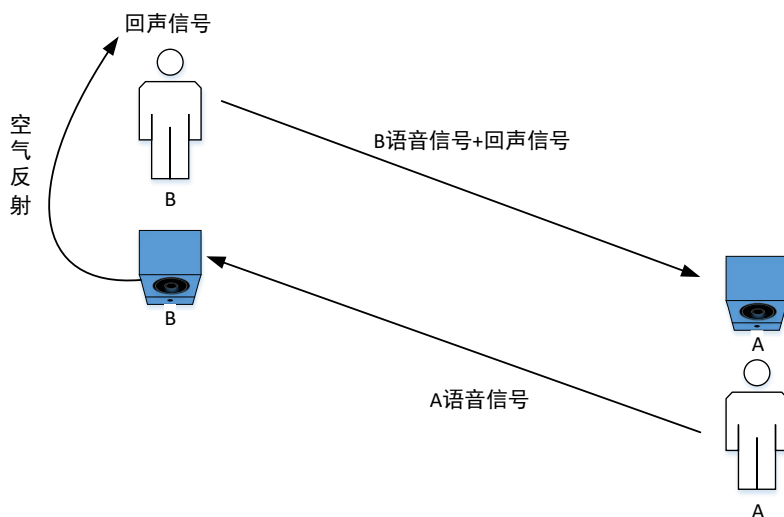


图 3-1 声学回声模型图

电路回声是指由于物理线路原因引起的, 临近的两条线路产生耦合产生电路回声。由于为了节省电线的成本, 用户之间的采用的是两线制, 而电话局是采用的四线制。所以为了两线制和四线制的匹配, 电话局一般会采用交换机进行线路转换。所以一般在交换机上会发生电路回声, 在交换机内部还有 2-4 线转换的电路, 由于线路之间不匹配, 部分的语音信号会无法转换, 从而被反馈原有电路, 最后形成了电路回声, 线路原理图如图 3-2 所示。从图中可以看出, 用一个简单的线性叠加器(消除回声模块)就可以实现电路回声消除。

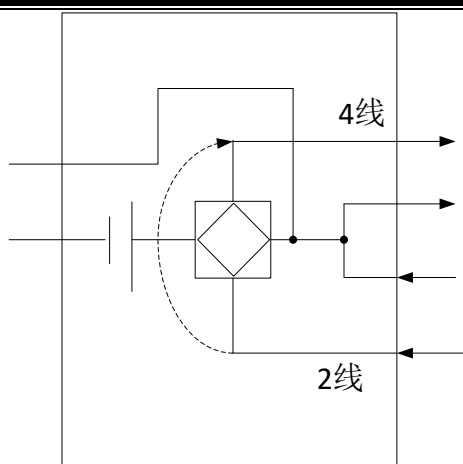


图 3-2 电路回声模型图

两种回声都是人们生活中所常见的，但是电路回声可以通过合理设计消除回声模块消除掉，相对比声学回声消除比较容易。消除声学回声是一件复杂和困难的事情。影响声学回声的因素也比较多，通常会因回声路径的改变，从而回声信号会发生改变。所以在本章内容主要就是讨论一下声学回声消除。

声学回声不是单一路径回声，它多种路径共同产生，声学回声可以根据路径分为直接回声和反射回声。回声路径的不确定性导致回声的多样性，也让回声消除增加了难度。声学回声消除的关键在于构建一个良好的自适应滤波器来消除声学回声。

因为下面三个因素使声学回声消除变成一个非常艰巨的任务：

（1）回声路径通常具有长脉冲响应。声学路径的特性取决于空间的物理条件。比如说在一个很大的房间里，需要一个很长的自适应滤波器去对回声路径去建模。往往却因为步长太小导致了较慢收敛和沉重的计算负荷结果。

（2）所述声学回声路径的特性随时间变化而变化，也可能由于人的移动，一扇门的打开或关闭或者室内温度的变化等原因发生改变。因此，自适应滤波器应该具有快速收敛能力适应环境的变化。

（3）语音信号往往是高度相关且非平稳的。对于较短的语音信号，谱密度在不同的频率是不相同的，这也导致收敛速衰减严重。自适应滤波算法的核心思想就是根据输入信号的特征调整滤波器的系数因子，从而得到一个最优输出^[42-47]，如图 3-3 所示。

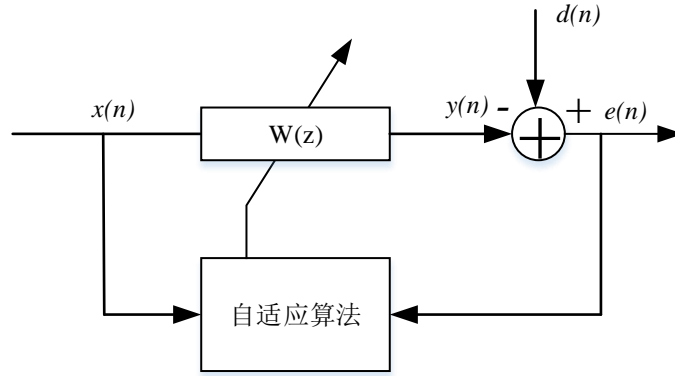


图 3-3 自适应滤波结构框图

$$e(n) = d(n) - y(n) \quad (3-1)$$

其中 $x(n)$ 为系统的输入信号， $y(n)$ 为系统的输出信号， $d(n)$ 为期望信号，估计误差为 $e(n)$ 。由图中可以看出，估计误差决定了滤波器性能的好坏。当 $e(n)$ 越接近 0 时，滤波器的输出信号 $y(n)$ 就越接近人们的期望信号 $d(n)$ ，那么输出信号就是人们想要的信号。

自适应滤波器在许多应用中使用，因为它们拥有在未知的和不断变化的环境中运行的能力。自适应滤波的应用可以分为四类：自适应系统识别，自适应逆建模，自适应预测和自适应阵列处理。其中声学回声消除是自适应系统辨识的一种应用。

3.3.1 LMS 算法推导与分析

在 1960 年，当 Hoff 和 Widrow 霍夫研究自适应线性滤波器，他们提出的最小均方（LMS）算法^[48]。它是一个结构相对简单和广泛应用的自适应滤波算法。众所周知，它在信号平稳环境下具有低计算复杂度和良好的收敛。该滤波器自适应率可以通过调整步长的参数被改变。当步长因子较大时，自适应收敛速度快，但稳定性差。相反，当步长较小，自适应收敛速度慢，但稳定性好。

LMS 方法就是在均方的情况下使误差信号 $e(n)$ 最小。传统的 LMS 算法依据最陡下降法算法实现的，也就是将函数 $\hat{\xi} = e^2(n)$ 替换性能函数 $\xi = E[e^2(n)]$ ，性能函数算法定义如下：

$$\xi(n) = E[e^2(n)] = E[d^2(n) - 2d(n)y(n) + y^2(n)] \quad (3-2)$$

其中：

$$y(n) = \mathbf{X}^T(n) \mathbf{w} \quad (3-3)$$

$$\mathbf{w} = [w_0, w_1, w_2, \dots, w_{N-1}]^T \quad (3-4)$$

$w_0, w_1, w_2, \dots, w_{N-1}$ 为滤波器的抽头权系数， \mathbf{w} 为滤波器抽头权向量。

所以可以推出：

$$\xi = E[(d(n) - \mathbf{w}^T x(n))(d(n) - x^T(n)\mathbf{w})] \quad (3-5)$$

将公式 (3-5) 式右边展开，因为 \mathbf{w} 不是统计变量，最后 3-5 式可以变成

$$\begin{aligned} \xi = E[d^2(n)] - \mathbf{w}^T E[x(n)d(n)] - E[d(n)x^T(n)]\mathbf{w} \\ + \mathbf{w}^T E[x(n)x^T(n)]\mathbf{w} \end{aligned} \quad (3-6)$$

将 \mathbf{P} 设置为互相关向量：

$$\mathbf{p} = E[x(n)d(n)] = [p_0 \ p_1 \ \dots \ p_{N-1}]^T \quad (3-7)$$

将 \mathbf{R} 设置为自相关向量

$$\begin{aligned} \mathbf{R} = E[x(n)x^T(n)] \\ = \begin{bmatrix} r_{00} & r_{01} & r_{02} & \dots & r_{0,N-1} \\ r_{10} & r_{11} & r_{12} & \dots & r_{1,N-1} \\ r_{20} & r_{21} & r_{22} & \dots & r_{2,N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{N-1,0} & r_{N-1,1} & r_{N-1,2} & \dots & r_{N-1,N-1} \end{bmatrix} \end{aligned} \quad (3-8)$$

可得到新的性能函数

$$\xi = E[d^2(n)] - 2\mathbf{w}^T \mathbf{p} + \mathbf{w}^T \mathbf{R} \mathbf{w} \quad (3-9)$$

从公式 (3-9) 可以看出新的性能函数是关于 \mathbf{w} 的二次函数，并且有且仅有一个极限值，为了得到性能函数的最优解，然后需要对该函数 \mathbf{w} 求偏导：

$$\frac{\partial \xi}{\partial w_i} = 0, \quad i = 0, 1, \dots, N-1 \quad (3-10)$$

最后可得到最优解：

$$\mathbf{w} = \mathbf{R}^{-1} \mathbf{p} \quad (3-11)$$

若 $\mathbf{w}(n)$ 为第 n 部迭代的抽头权向量，那么下列方程可以用于更新 $\mathbf{w}(n)$

$$\mathbf{w}(n+1) = \mathbf{w}(n) - 2\mu g(n) \mathbf{e}^2(n) \quad (3-12)$$

其中 $\mathbf{w}(n) = [w_0(n) \ w_1(n) \ \dots \ w_{N-1}(n)]$ ， μ 是补偿因子，梯度向量 $g(n)$ 计算公式如下：

$$\mathbf{g}(n) = \left[\frac{\partial}{\partial w_0} \ \frac{\partial}{\partial w_1} \ \dots \ \frac{\partial}{\partial w_{N-1}} \right]^T \quad (3-13)$$

然后将这结果带入公式 (3-12) 可以得到

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2\mu \mathbf{e}(n) \mathbf{x}(n) \quad (3-14)$$

这就是最小 LMS 算法，它阐述了滤波器迭代的过程，目的是让输出信号 $y(n)$ 与期望信号 $d(n)$ 的差值达到均方值最小化。表 3-1 对 LMS 算法进行了总结。

表 3-1 LMS 算法总结

| | |
|---------------------------------|----------|
| 输入：滤波器抽头权向量 | $w(n)$ |
| 输入向量 | $x(n)$ |
| 期望输出 | $d(n)$ |
| 输出：滤波器输出 | $y(n)$ |
| 滤波器抽头权向量 | $w(n+1)$ |
| 滤波： | |
| $y(n) = w^T(n)x(n)$ | |
| 误差估计 | |
| $e(n) = d(n) - y(n)$ | |
| 抽头权系数向量更新 | |
| $w(n+1) = w(n) + 2\mu e(n)x(n)$ | |

3.3.2 NLMS 算法推导与分析

归一化 LMS 算法是 NLMS 算法一种特殊形式，它是将 LMS 算法的步长因子进行归一化处理。在性能上，它比 LMS 自适应算法更加优异，应用范围也广泛。它比 LMS 算法稳定性更好且可以快速收敛。

将 LMS 方程中的步长因子更新为时变步长：

$$w(n+1) = w(n) + 2\mu(n)e(n)x(n) \quad (3-15)$$

可以看出 NLMS 和 LMS 算法的不同之处在于步长因子，LMS 算法中的步长为一个给定的常数，NLMS 算法中的步长为一个随时间变化的量，定义自适应算法的先验误差、后验误差为

$$e(n) = d(n) - W^T(n)X(n) \quad (3-16)$$

$$\varepsilon(n) = d(n) - W^T(n+1)X(n) \quad (3-17)$$

将后验误差进行最优处理。将 LMS 算法递推公式带入 3-17，可获得：

$$\varepsilon(n) = (1 - 2\mu(n)x^T(n)x(n))e(n) \quad (3-18)$$

对时变收敛步长求偏导，并让其等于 0，可获得：

$$\mu(n) = \frac{1}{2x^T(n)x(n)} \quad (3-19)$$

将上式带入 LMS 算法地推公式可获得 NLMS 算法递推方程，方程如下：

$$w(n+1) = w(n) + \frac{1}{x^T(n)x(n)} e(n)x(n) \quad (3-20)$$

也可以将上式理解为每一个新的输入信号多次运行 LMS 算法直到收敛为止。
NLMS 算法总结如下表所示：

表 3-2 NLMS 算法总结

| | |
|--|----------|
| 输入：滤波器抽头权向量 | $w(n)$ |
| 输入向量 | $x(n)$ |
| 期望输出 | $d(n)$ |
| 输出：滤波器输出 | $y(n)$ |
| 滤波器抽头权向量 | $w(n+1)$ |
| 滤波： | |
| $y(n) = w^T(n)x(n)$ | |
| 误差估计 | |
| $e(n) = d(n) - y(n)$ | |
| 抽头权系数向量更新 | |
| $w(n+1) = w(n) + \frac{\mu}{x^T(n)x(n) + \psi} e(n)x(n)$ | |
| 其中 ψ 为常数，防止分母过小 | |

3.3.3 子带分解自适应算法

在一个典型空间内进行声学回声消除需要高阶自适应 FIR 滤波器，以覆盖长回声巴。这是使用 FIR 自适应滤波器进行滤波的缺点，因为高的计算负荷和较慢收敛的速度。所以提出的子带滤波自适应滤波器（SAF）结构来克服这些问题。在所有的 SAF 中，分析滤波器将彩色输入信号分割成几乎互斥的多个子带信号，然后利用近似于白信号的抽取子带信号更新滤波器的系数向量，从而提高收敛速度。然而，传统的 SAF 的收敛性能受到混叠和带边效应的限制，这主要是由于在每个子带自适应环中使用单独的子滤波器。为了解决这种结构问题，提出了一种 SAF 的多带结构，利用抽取的子带信号更新全带自适应滤波器。在多频带结构的基础上，由 Li 和 Gan 开发了归一化 SAF (NSAF) 算法。与 NLMS 相比，NSAF 对有色输入信号提供更快收敛速度。除了这个优点，NSAF 仍然保持与 NLMS 几乎相同的计算复杂度，尤其是对于长滤波

器。

此外，子带系统还降低了计算复杂度，因为可以用低阶子滤波器以与其窄带宽度相当的低抽取速率处理子带信号。使用 SAF 的主要动机是降低计算复杂度，并改善具有大频谱动态范围的输入信号的收敛性能，即当输入信号是高度相关的。

当输入信号高度相关时，基于梯度的算法（例如 LMS 和 NLMS）遭受缓慢收敛。这个问题与输入自相关矩阵中由于输入信号的大频谱动态范围而引起的大特征值扩展有关。由于子带信号的频谱动态范围小于全带信号的频谱动态范围，因此可以实现更快的收敛。此外，自适应子滤波器的长度比全带滤波器短，这允许使用更大的步长来进一步提高收敛速度。

子带自适应滤波的主要思想是使用一组低阶自适应子滤波器分解高阶自适应滤波任务。根据误差信号的推导，SAF 结构可分为开环结构或闭环结构。在每个子带中，通过自适应子滤波器对抽取的输入信号进行滤波，并将输出与对应的期望响应进行比较。子带误差信号（即，子带期望响应与对应的子滤波器输出之间的差）用于更新相关联的自适应子滤波器。这些子带误差信号用于使用合成滤波器组形成全频带误差信号。这种类型的 SAF 结构被称为开环结构，因为子滤波器的自适应与全带误差信号无关。开环结构在自己的自适应环路中独立地调整单个子滤波器，以使每个子带中的相应本地均方误差（MSE）最小化。它们无需共享与子带之间的其它子滤波器信息（输入和误差信号）独立地操作。开环结构的函数是所有子带的 MSE 的总和，可以用公式（3-21）表示。

$$J_{SB} = \sum_{i=0}^{N-1} E \{ |e_{i,D}(k)|^2 \} \quad (3-21)$$

其中 $|\cdot|$ 代表绝对值。图 3-4 表示子带自适应滤波器的开环结构。

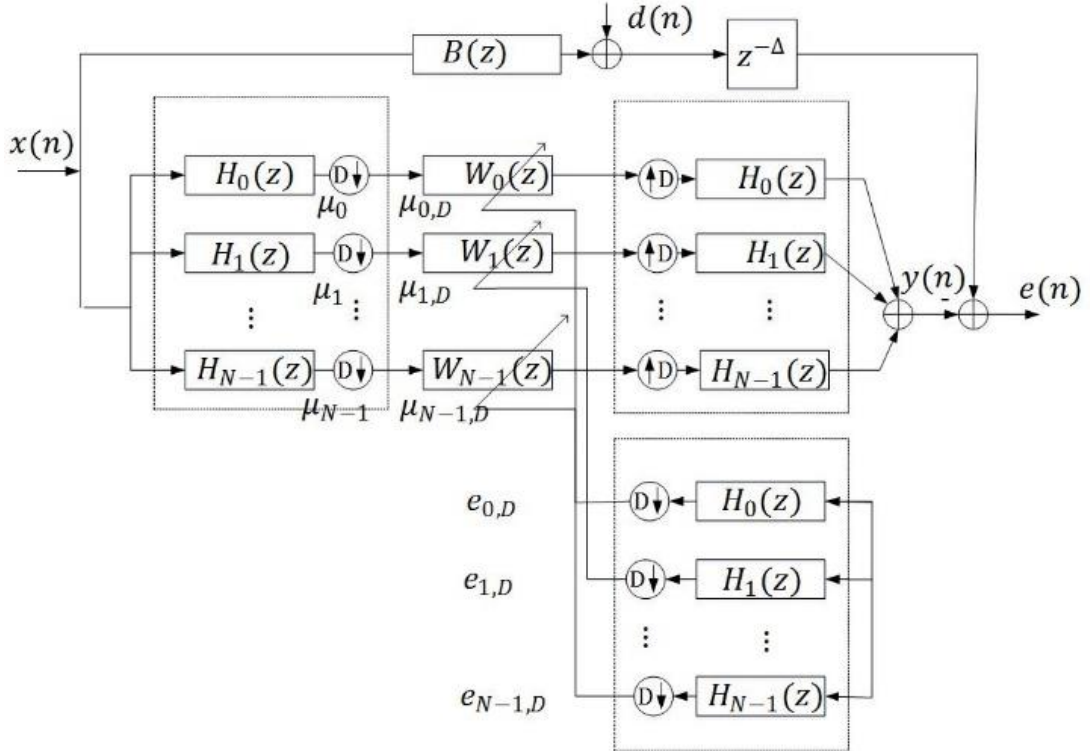


图 3-4 子带滤波器开环结构

闭环结构全频带的输出是使用合成滤波器组进行重新构建的。它的全带误差信号被划分成使用相同分析滤波器子带，并且抽取的子带误差信号被用于各个子带自适应滤波器。期望响应信号是未知系统的输出。闭环结构最小化了全频带 MSE，而不是开环对应的子带 MSE 之和。利用闭环反馈，全带误差信号中的非零频率分量将被反馈以更新所有子带自适应子滤波器，以使那些分量最小化。因此，闭环结构允许全域误差信号的最小化，并保证整个自适应子滤波器收敛到最优维纳解。然而，除了人们有意强加于期望响应的延迟之外，应用于输出信号的合成滤波器组和应用与误差信号的分析滤波器组的组合也产生不希望的延迟。这种延迟减小了步长的上界，因此要求闭环结构使用较小的步长，这导致较慢的收敛速度，子带闭环结构如图 3-5 所示。

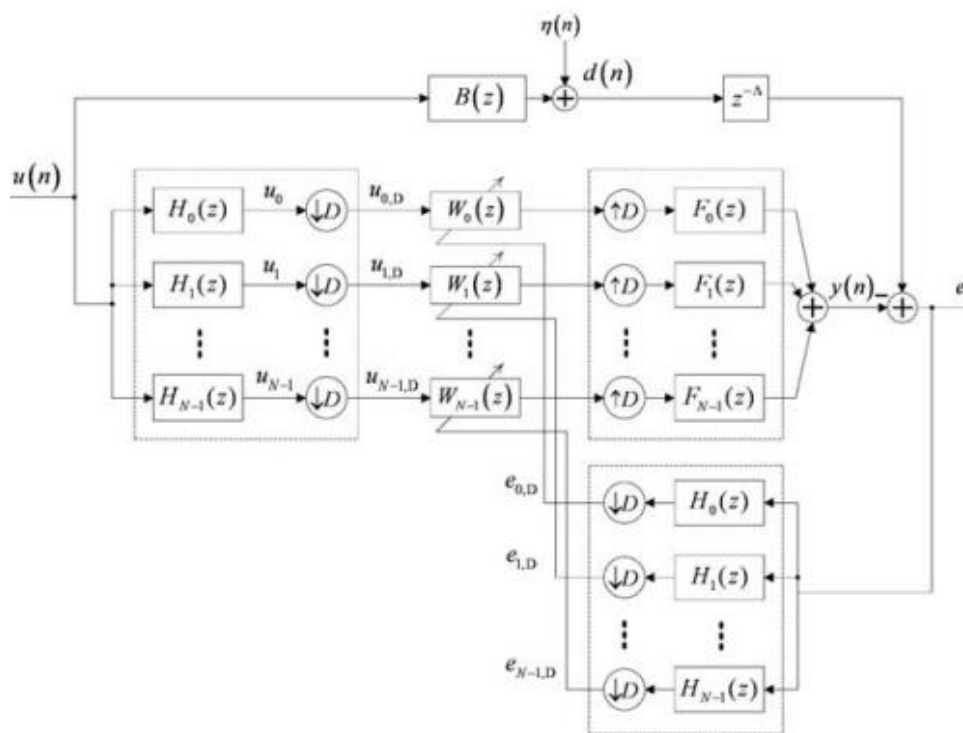


图 3-5 子带滤波器闭环结构

子带滤波器组临界降采样每个子带的频谱的动态范围和加速自适应子滤波器的收敛。然而，采样会引入混叠成分到子带信号。这些混叠成分作为附加的噪声扰乱了自适应滤波，因此在重构全频带信号会导致更高的 MSE。即使混叠成分可以通过重建分析-合成滤波器组去消除，但是混叠是在个别子带不可避免的。

为了减小子带系统中的混叠效应，通常采用过采样结构来代替临界采样结构。利用过采样 DFT 滤波器组具有很好的效果。这种过采样结构虽然减少了严重的混叠成分。但是，由于分析滤波器的频带边缘，子带信号的过采样会在子带自相关矩阵中产生一些小的特征值，又会带来边缘效应。

已经有许多方法可以克服临界采样的 SAF 遇到的混叠效应，方法如下：

- (1) 由相邻的子带之间产生的频谱间隙减少分析滤波器的重叠。
- (2) 掺入相邻子带之间的自适应交叉滤波器来补偿混叠效应。
- (3) 采用的多频带结构 SAF (MSAF)。
- (4) 采用闭环无延迟结构。

可以使用闭环子带结构来减少混叠效应，该子带结构最小化全带 MSE，而不是单个子带 MSE。最小化的闭环子带结构被降低。

3.3.4 可变步长自适应算法

归一化最小均方(NLMS)算法的性能受两个重要参数的影响,即标准化的步长和正则化项。标准化步长参数反映了算法的收敛速度和收敛精度之间的权衡。正则化项参数对于所有不良和病态的问题都是起着重要作用的,例如在自适应滤波器中,它取决于系统的信噪比(SNR)。这些参数都可以控制(即,使之与时间相关),以解决快速收敛和低失调之间的矛盾。为了克服这一问题后来大量学者提出了 NLMS 算法的可变步长(VSS)算法。可变步长算法主要分为两类。第一类是通过考虑到一些限制条件获得的,比如估计误差就是一种根据步长因子减小最快的。第二类被命名为“启发式”算法,它们基于一些直觉规则,例如增加步长增加它们的收敛速度,或者是减少步长来降低它们的稳态误差。许多算法采用 Sigmoid 函数改进自适应算法。

可变步长算法的目的是为了平衡收敛速度和稳态误差之间的平衡。其核心思想是在算法的初始阶段采用较大步长提高收敛速度,然而在算法结束收敛阶段时候采用较小的步长,这样可以保证一个较小的稳态误差。

可变步长 NLMS 算法用于平衡的收敛速度和稳态误差之间的折衷。可变步长 NLMS 算法迭代更新公式如下所示:

$$W(n+1) = W(n) + \mu(n) \frac{X(n)e(n)}{\varepsilon + X^T(n)X(n)} \quad (3-22)$$

其中 $\mu(n)$ 是具有可变正标量的可变步长可以控制滤波器系数更新。

为了考虑到自适应滤波的不同方面因素,如收敛速度,均方误差,和 NLMS 算法的计算成本,需要一个适当的步长因子去进行控制。在众多的可变步长 NLMS 算法里,许多学者用非参数和广义 Sigmoid 算法两种类型进行描述。

(1) 非参数可变步长 NLMS 算法

非参数可变步长 NLMS 算法的推导方程是通过每一时刻减少平方误差调整步长因子值来获得的,并且让背景噪音的数学期望等于后验估计误差的数学期望。先验估计误差表达式和后验估计误差表达式分别为公式 3-16 和公式 3-17 所示。

非参数可变 NLMS 算法的步长因子如下面公式所示:

$$\mu_{NPVSS} = \begin{cases} \beta(n), & \hat{\sigma}_e(n) > \sigma_b \\ 0, & \text{otherwise} \end{cases} \quad (3-23)$$

$$\text{其中 } \beta(n) = \frac{1}{\psi + X^T(n)X(n)} \left(1 - \frac{\sigma_b}{\varepsilon + \hat{\sigma}_e(n)}\right)$$

$\hat{\sigma}_e^2(n)$ 是误差信号的功率, $\sigma_b^2(n)$ 是噪声信号的功率。 $\hat{\sigma}_e(n)$ 迭代公式如下所示:

$$\hat{\sigma}_e(n) = \lambda \hat{\sigma}_e^2(n-1) + (1-\lambda)e^2(n) \quad (3-24)$$

其中 λ 为指数加权因子, $0 < \lambda < 1$ 。非参数可变步长总结算法如下表所示。

表 3-3 非参数可变步长 NLMS 算法总结

| | |
|--------|---|
| 参数: | $L, N, \varepsilon, \lambda, \psi$ |
| 初始化: | $W(0) = [w_0 \ w_1 \ \dots \ w_{L-1}]^T = 0_{L \times 1}$ $\hat{\sigma}_e^2(0) = 0$ $X(0) = 0_{L \times 1}$ |
| 输入: | $X(n) = [x(n)x(n-1) \dots x(n-L+1)]^T$ |
| 期望输出 | $d(n)$ |
| 滤波器输出: | $y(n)$ |
| 误差估计: | $e(n) = d(n) - y(n)$ |
| 迭代更新: | $\beta(n) = \frac{1}{\psi + X^T(n)X(n)} \left(1 - \frac{\sigma_b}{\varepsilon + \hat{\sigma}_e(n)}\right)$ $\mu_{NPVSS} = \begin{cases} \beta(n), & \hat{\sigma}_e(n) > \sigma_b \\ 0, & \text{otherwise} \end{cases}$ $\hat{\sigma}_e(n) = \lambda \hat{\sigma}_e^2(n-1) + (1-\lambda)e^2(n)$ $W(n+1) = W(n) + \mu_{NPVSS}X(n)e(n)$ |

(2) 广义 Sigmoid 可变步长 NLMS 算法

广义 Sigmoid 可变步长 NLMS 算法是基于自适应滤波算法调整 Sigmoid 函数。它的原理就是在初始阶段要有个较大步长获得一个好的跟踪速度。但是在收敛阶段时维持一个小的步长保证稳态误差。

Sigmoid 可变步长 NLMS 算法的步长方程式如下所示:

$$\mu_{SVSS} = B \left(\frac{1}{1 + \exp(-A|e(n)|)} - 0.5 \right) \quad (3-25)$$

其中 A 和 B 是可变步长函数的参数。

广义 Sigmoid 可变步长 NLMS 算法的步长方程式如下所示:

$$\mu_{GSVSS} = B \left(\frac{1}{1 + \exp(-A(\sigma_e(n) - \sigma_b)^m)} - 0.5 \right) \quad (3-26)$$

其中 m 是广义函数的参数, $0 < m < 5$, 广义 Sigmoid 可变步长 NLMS 算法总结如下表所示。

表 3-4 广义 Sigmoid 可变步长 NLMS 算法总结

| | |
|--------|---|
| 参数: | $L, N, \varepsilon, \lambda, A, B, m$ |
| 初始化: | $\mathbf{W}(0) = [w_0 \ w_1 \ \dots \ w_{L-1}]^T = \mathbf{0}_{L \times 1}$ $\mathbf{X}(0) = \mathbf{0}_{L \times 1}$ |
| 输入: | $\mathbf{X}(n) = [x(n)x(n-1) \ \dots \ x(n-L+1)]^T$ |
| 期望输出 | $d(n)$ |
| 滤波器输出: | $y(n)$ |
| 误差估计: | $e(n) = d(n) - y(n)$ |
| 迭代更新: | $\mu_{GSVSS}(n) = B\left(\frac{1}{1+\exp(-A(\sigma_e(n)-\sigma_b)^m)} - 0.5\right)$ $\mathbf{W}(n+1) = \mathbf{W}(n) + \mu_{GSVSS}(n) \frac{x(n)e(n)}{\varepsilon + \mathbf{X}^T(n)\mathbf{X}(n)}$ |

3.3 改进的可变步长子带自适应算法

首先将公式 (3-16) 和公式 (3-17) 的先验误差和后验误差进一步推导得出:

先验误差:

$$e(n) = d(n) - \mathbf{W}^T(n)\mathbf{X}(n) = \mathbf{x}^T(n)[\mathbf{w} - \hat{\mathbf{w}}(n-1)] + b(n) \quad (3-27)$$

后验误差:

$$\varepsilon(n) = d(n) - \mathbf{W}^T(n+1)\mathbf{X}(n) = \mathbf{x}^T(n)[\mathbf{w} - \hat{\mathbf{w}}(n)] + b(n) \quad (3-28)$$

其中 $\hat{\mathbf{w}}$ 是对 \mathbf{w} 的估计值。

针对非参数变步长 NSAF 算法 (NPVSS-NSAF) 在输入信号着色时收敛速度慢和失调率低的矛盾, 提出了一种改进的 NPVSS-NSAF 算法。该算法在低信噪比条件下, 将待更新的最优权值向量与过去估计的权值向量之差的平方欧氏范数之和最小化, 以获得可变步长, 避免了回波路径估计向量在低信噪比条件下不会产生大的波动。为了进一步提高 VSS-NSAF 算法的鲁棒性, 通过使子带信号噪声的功率等于子带后验误差的功率, 得到了无参数步长因子。可以建立以下优化问题

$$\sum_{L=0}^{L-1} \varphi^L \|\mathbf{W}_o - \mathbf{w}(k)\|^2 \quad (3-29)$$

其中, φ 是权重因子, $0 < \varphi < 1$ 。

然后有:

$$d_{(i,D)}(k) = \mathbf{x}_i^T(k)\mathbf{W}_o \quad (3-30)$$

利用拉格朗日乘子导出了一个新的公式:

$$J(k) = \sum_{L=0}^{L-1} \phi^L \|W_o - w(k)\|^2 \lambda_i [d_{(i,D)}(k) x_i^T(k) W_o] \quad (3-31)$$

其中, λ_i 是拉格朗日乘子, 对公式 (3-31) 中的 W_o 求偏导, 可以得到:

$$\frac{\partial J(k)}{\partial W_o} = 2 \sum_{L=0}^{L-1} \phi^L [W_o - w(k)] - \sum_{i=0}^{N-1} \lambda_i x_i^T(k) \quad (3-32)$$

令公式 (3-32) 结果为零, 可以推出,

$$W_o = (\sum_{L=0}^{L-1} \phi^L)^{-1} \sum_{L=0}^{L-1} \phi^L w(k) + \frac{1}{2} (\sum_{L=0}^{L-1} \phi^L)^{-1} \sum_{i=0}^{N-1} \lambda_i x_i^T(k) \quad (3-33)$$

将公式 (3-30) 矩阵化, 可以得出

$$d_D(k) = X^T(k) W_o \quad (3-34)$$

其中, $d_D(k) = [d_{0,D}(k), d_{1,D}(k), \dots, d_{N-1,D}(k)]$ 。

将公式 (3-33) 带入公式 (3-34):

$$d_D(k) = X^T(k) A \sum_{L=0}^{L-1} \phi^L w(k) + X^T(k) X(k) \frac{A}{2} \lambda \quad (3-35)$$

其中, $A = (\sum_{L=0}^{L-1} \phi^L)^{-1}$, $\lambda = [\lambda_0, \lambda_{(1)}, \dots, \lambda_{(N-1)}]^T$ 。

从公式 (3-35) 中可以推出:

$$\lambda = \frac{2}{A} [X^T(k) X(k)]^{(-1)} [d_D(k) - X^T(k) A \sum_{L=0}^{L-1} \phi^L w(k)] \quad (3-36)$$

重新定义 $\tilde{e}_{(i,D)}(k) = d_{(i,D)}(k) - X^T(k) A \sum_{L=0}^{L-1} \phi^L w(k)$, 然后可以得到:

$$W_o = A \sum_{L=0}^{L-1} \phi^L w(k) + X(k) [X^T(k) X(k)]^{-1} \tilde{e}_{(i,D)}(k) \quad (3-37)$$

使子带的后验误差信号等于子带系统噪声

$$\varepsilon_{(i,D)}(k) = d_{(i,D)}(k) x_i^T(k) w(k+1) = x_i^T(k) [W_o - w(k+1)] + v_{(i,D)}(k) \quad (3-38)$$

其中, $\varepsilon_{i,D}(k) = v_{i,D}(k)$, $[W_o - w(k+1)] = 0$, 从公式 (3-37) 可以得出:

$$w(k+1) = A \sum_{L=0}^{L-1} \phi^L w(k) + X(k) [X^T(k) X(k)]^{-1} \tilde{e}_{(i,D)}(k) \quad (3-39)$$

根据文献^[175], 可以得到:

$$E\{\varepsilon_{(i,D)}^2(k)\} = \sigma_{v_{i,D}}^2 \quad (3-40)$$

其中, $\sigma_{v_{i,D}}^2 = E\{v_{(i,D)}^2(k)\}$ 是子带系统噪声的功率。重新定义 $\sigma_{x_i}^2 = E\{x_{(i,D)}^2(k)\}$ 作为子带输入信号的功率。将公式 (3-27) 代入公式 (3-38), 然后用公式 (3-28) 消除 $w(k)$ 。

$$E\{\varepsilon_{(i,D)}^2(k)\} = [1 - \mu_{(i,D)}(k) \sigma_{x_i}^2]^2 \sigma_{\tilde{e}_{i,D}(k)}^2 = \sigma_{v_{i,D}}^2 \quad (3-41)$$

其中, $\sigma_{\tilde{e}_{i,D}}^2(k) = E\{\sigma_{\tilde{e}_{i,D}}^2(k)\}$ 是子带误差信号功率。重新排列公式 (3-41):

$$\mu_{(i,D)}^2(k) - \frac{2}{\sigma_{x_i}^2} \mu_{i,D}(k) + \frac{1}{(\sigma_{x_i}^2)^2} [1 - \frac{\sigma_{v_{i,D}}^2}{\sigma_{\tilde{e}_{i,D}(k)}^2}] = 0 \quad (3-42)$$

可以得到:

$$\mu_{i,D}(k) = [1 - \frac{\sigma_{v_{i,D}}^2}{\sigma_{\tilde{e}_{i,D}(k)}^2}] = 0 \quad (3-43)$$

可以清晰的看出 $\sigma_{\tilde{e}_{i,D}(k)}^2 > \sigma_{v_{i,D}}^2$ ，这也就意味着 $\mu_{i,D}(k) > 0$ 。实际中，第 i 个子带误差信号和第 i 个子带输入信号估计值为：

$$\sigma_{\tilde{e}_{i,D}}^2(k) = B\sigma_{\tilde{e}_{i,D}}^2(k-1) + (1-B)\sigma_{\tilde{e}_{i,D}}^2(k) \quad (3-44)$$

$$\sigma_{x_i}^2(k) = B\sigma_{x_i}^2(k-1) + (1-B)\sigma_{x_i}^2(kN) \quad (3-45)$$

其中，B 是一个指数窗口，所提出的算法可以总结如表 3-5 所示。

表 3-5 NPVSS NASF 算法总结

初始化： $W(0) = 0, \mu_i(0) = 1$

$$\sigma_{\tilde{e}_{i,D}}^2(0) = 0, \sigma_{v_{i,D}}^2 = 0$$

参数： φ 权重因子 $1 > \varphi > 0$

$$\sigma_{v_{i,D}}^2 \quad \text{噪声功率}$$

更新： $\sigma_{\tilde{e}_{i,D}}^2(k) = B\sigma_{\tilde{e}_{i,D}}^2(k-1) + (1-B)\sigma_{\tilde{e}_{i,D}}^2(k)$

$$\sigma_{x_i}^2(k) = B\sigma_{x_i}^2(k-1) + (1-B)\sigma_{x_i}^2(kN)$$

子带权重向量自适应：

$$w(k+1) = A \sum_{L=0}^{L-1} \varphi^L w(k) + \mu(n) \frac{X(k)\tilde{e}_{i,D}(k)}{[X^T(k)X(k)]}$$

$$\mu_{i,D}(k) = [1 - \frac{\sigma_{v_{i,D}(k)}}{\sigma_{\tilde{e}_{i,D}(k)}}]$$

3.4 实验结果与分析

3.4.1 实验设置

为了验证本文提出的改进的可变步长子带自适应算法的优异性能，在这一部分进行了声学回声消除的仿真实验。实验主要分为两个大组，第一组是用声学脉冲信号进行模拟，通过模拟实验进行实验结果评估。在第一组中采用高斯白噪声（WGN）作为语音输入信号。输入信噪比（SNR）是 30dB 和 10dB。子带系统中使用的 $M = 8$ 段。的全频带和子带滤波器的长度被选择以使得在 8kHz 的取样速率的 512 回波路径建模。在这个实验中，本文假设系统的噪声功率是已知的。第二组采用真实的语音信号作为语音输入信号，从 TIMIT 数据库中采集语音信号，采样频率为 8kHz，其它设置与第

一组实验设置保持一致。

为了让实验更有说服力，本章所述的改进算法与 NLMS 算法、VSS-NLMS 算法和 SM-NLMS 算法进行比较。通过对比实验说明本文提出的改进算法在收敛速度和稳态误差都优于其它传统算法。

3.4.2 评价标准

声学回声消除已经成为一个有吸引力的挑战，因为这样的事实，提高语音质量仍然是通信系统，如电话，免提和移动通信的最重要目标之一。自适应滤波算法常常被应用在回声消除的应用当中。针对回声消除应用的自适应算法评价标准有许多种。其中均方误差（MSE）、均方差（MSD）、失调（MIS）和回声返回损耗增量（ERLE）是最常用的四种评价标准。

使用不同的自适应滤波算法，应该选择不同的评价标准。比如说归一化子带自适应滤波，如果以 MSE 和 ERLE 作为评价标准，往往会忽略自适应滤波器的精度，然而 MIS 和 MSD 评价标准就会更加适合归一化子带自适应滤波算法。

这四种自适应评价标准定义如下：

- (1) $MSE(dB) = 10\log_{10}(E\{|e(n)|^2\})$
- (2) $MSD(dB) = 10\log_{10}[\|w_0 - w(n)\|_2^2 / \|w_0\|_2^2]$
- (3) $MIS(dB) = 10\log_{10}(\frac{\|w(n) - h\|^2}{\|h\|^2})$
- (4) $ERLE(dB) = 10\log_{10} \frac{e^2(k)}{d^2(k)}$

从表达式中可以看出 MSE 和 ERLE 两个评价标准主要以噪声功率作为标准，然而 MSD 和 MIS 两个评价标准主要以自适应滤波器识别精度作为评价标准。所以说如果自适应滤波器中含有大量随机噪声，MSE 和 ERLE 两个评价标准会受到噪声的影响，无法准备的评价收敛速度与稳态失调。在本文中为了让实验更有说服力，本文采用了 MIS 和 ERLE 两个评价标准进行实验验证。

3.4.3 结果与讨论

四种算法的实验结果如图 3-6 所示，它们都是在相同的信噪比（SNR=30dB）进行的。从图 3-6 中可以看出，在回声消除的初始阶段中，NLMS（ $\mu=1.0$ ）、SM-NLMS 和本文提出的改进算法 NPVSS-NASF 都具有较快的收敛速度，VSS-NLMS 算法的收敛速度要略低于以上三种算法。但是在收敛阶段，从图中可以发现 VSS-NLMS 算法的稳态误差低于 SM-NLMS 和 NLMS（ $\mu=1.0$ ）算法的稳态误差，而 NPVSS-NASF 的稳态误差要比 VSS-NLMS 算法稳态误差低 5dB。综上所述，在这四个算法里，本

章提出的改进算法在收敛速度和稳态误差中的表现都是最优的,也就说本章提出的改进算法的回声消除性能优于其它三种算法。

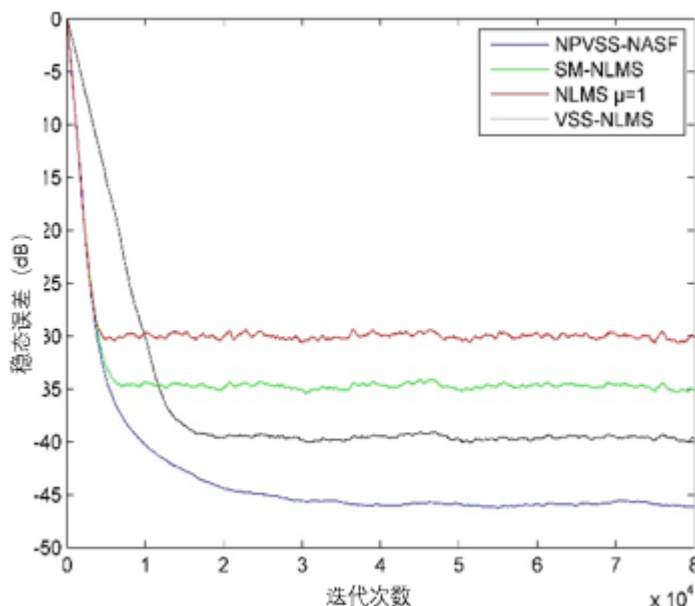


图 3-6 四种算法的稳态误差比较, SNR=30dB

为了仿真实验更有说服力,防止改进的算法在低信噪比情况下出现回声消除性能极度衰减的情况发生。在这一组实验中降低了信噪比,使用的输入信号是一个白高斯噪声,信噪比为 10dB。与上面的仿真实验相比,只降低了信噪比,其它参数保持不变。从图 3-7 可以看出,尽管降低信噪比 NPVSS-NASF 算法的回声消除性能有所降低,但是在初始阶段仍然有较快的收敛速度和较低的稳态误差。与其它三种算法相比, NPVSS-NASF 算法的最终稳态误差比其它三种算法低于超过 10dB。从以上两个实验可以看出, NPVSS-NASF 算法的性能无论在低信噪比情况下还是高信噪比情况下,它都具有更快的收敛速度和较低的稳态误差。

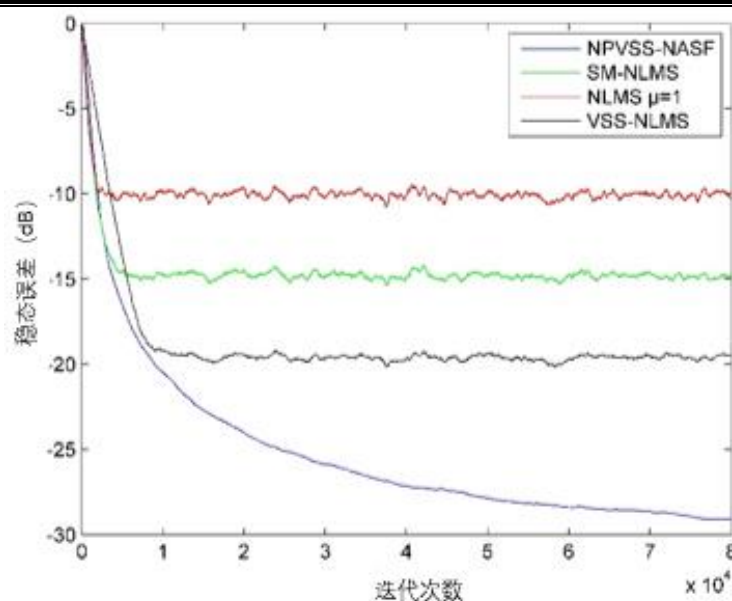


图 3-7 四种算法的稳态误差比较, SNR=10dB

由于单一的评价标准往往具有片面性,本章又对上述四种算法进行 ERLE 曲线评估仿真实验。在这个仿真实验中,仍然采用上述的 WGN 作为输入信号,仿真结果如图 3-8 所示。从图中可以看出,这四种算法的收敛速度从快到慢分别为 NPVSS-NASF、VSS-NLMS、NLMS、SM-NLMS,ERLE 的值从高到底分别为 NPVSS-NASF、SM-NLMS、VSS-NLMS、NLMS。当 ERLE 值越高,说明算法的回声消除性能越好。从回波损耗增量评价标准的角度出发,可以看出本章提出的 NPVSS-NASF 算法性能要优异其它三种传统回声消除算法。

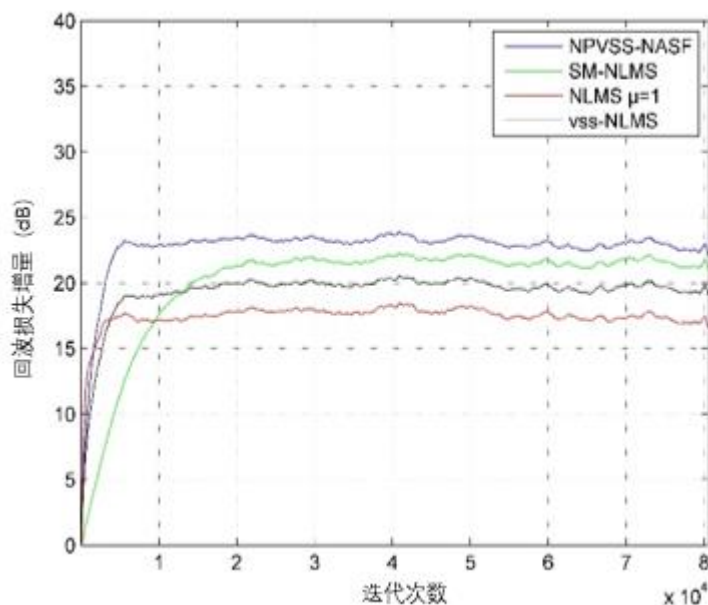


图 3-8 四种算法的 ERLE 比较, SNR=30dB

为了证明本章提出的改进算法在真实语音作为输入的情景下具有优异的性能,从

TIMIT 数据库中采集语音信号，采样频率为 8kHz。在本实验中，分别使用两种语音信号作为远端信号和近端信号。远端语音信号 $x(n)$ 由女性发音，发音的句子为：“She always ask an objective question”；近端语音信号 $s(n)$ 由男性发音，发音的句子为：“don’t let him eat too many strawberries”。

使用高斯白噪声作为噪声信号，然后将其添加到回波信号中，且选择不同的信噪比：30dB 和 10dB。采用三种传统的自适应算法 NLMS、VSS-NLMS 和 SM-NLMS 进行声学回声消除实验。将它们的实验结果与本章提出的 NPVSS-NASF 算法进行比较。为了使实验更有说服力，四种算法的参数保持一致（步长 $\mu=1$ ，总迭代次数 $n=80000$ ）。在这一组实验中，采样频率为 8kHz。声学回波路径的响应长度也是滤波器长度 $L=512$ 。

图 3-9 比较了 NPVSS-NASF 算法与其他三种算法（步长 $\mu=1.0$ 的 NLMS、VSS-NLMS 和 SM-NLMS）的失调。在本实验中，输入信号为非平稳环境下信噪比为 30dB 的语音信号。从图中的结果可以看出，在回声消除的初始阶段中，NLMS（ $\mu=1.0$ ）、SM-NLMS 和 NPVSS-NASF 都具有较快的收敛速度，VSS-NLMS 算法的收敛速度要略低于以上三种算法。但是在收敛阶段，从图中可以发现 VSS-NLMS 算法的稳态误差低于 SM-NLMS 和 NLMS（ $\mu=1.0$ ）算法的稳态误差，而 NPVSS-NASF 的稳态误差要比 VSS-NLMS 算法稳态误差低 5dB。综上所述，在非平稳环境中 NPVSS-NASF 算法在收敛速度和稳态误差中的表现都是最优的，NPVSS-NASF 算法的回声消除性能优异于其它三种算法。

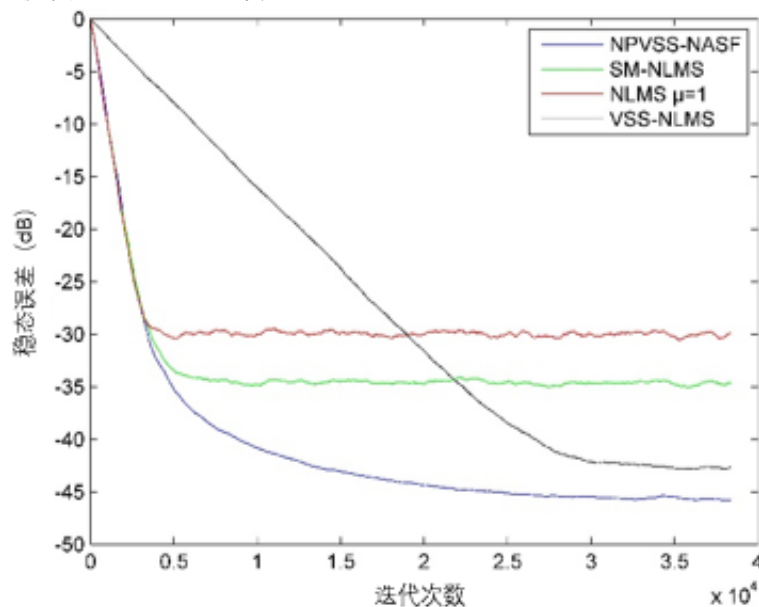


图 3-9 双音讲话稳态误差评估，SNR=30dB

为了消除信噪比的影响，在本组实验中仍然使用上述语音作为输入信号，但将信

噪比降低为 10dB。与上面的仿真实验相比，只降低了信噪比，其它参数保持不变。实验结果如图 3-10 所示，从图中可以看出，NPVSS-NASF 算法仍然具有最快的收敛速度，且最终稳态误差比其它三种算法低于超过 5dB。从以上两组实验可以看出，NPVSS-NASF 算法的性能无论在低信噪比情况下还是高信噪比情况下，在非平稳环境下它都具有更快的收敛速度和较低的稳态误差。

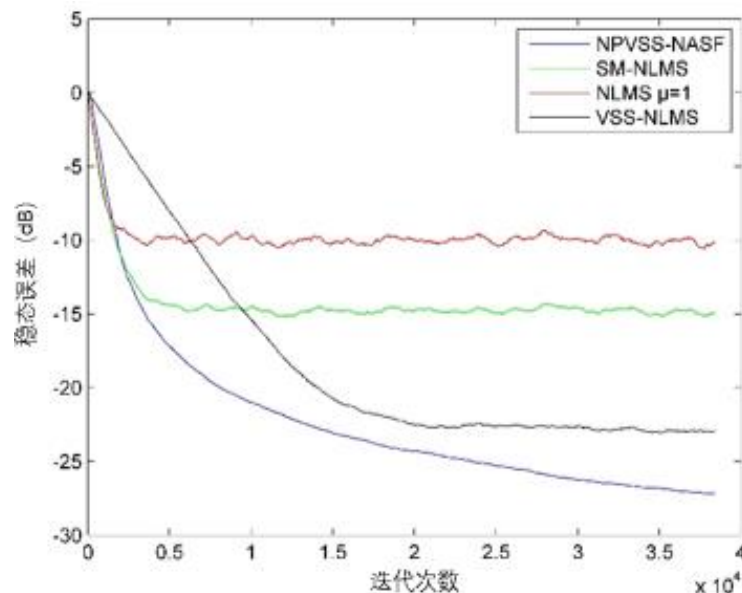


图 3-10 双音讲话稳态误差评估，SNR=10dB

在这一组实验中，实验参数与上面两组实验保持一致，从 ERLE 评价标准对上述四种算法进行评价，实验结果如图 3-11 所示。从图中可以看出，NPVSS-NASF 算法的 ELRE 值要比其它三种算法高出 5dB 以上，根据 ERLE 评价标准的特性可以说明 NPVSS-NASF 算法的回声消除能力优于 VSS-NLMS 算法、SM-NLMS 算法和 NLMS 算法。

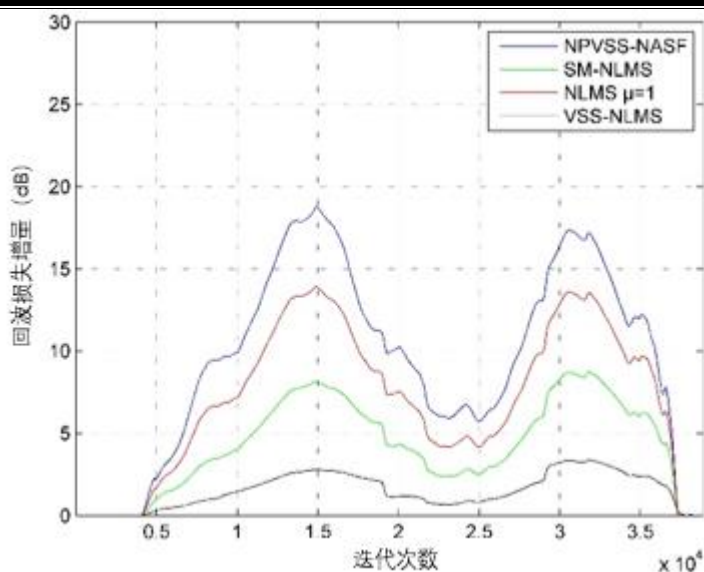


图 3-11 双音讲话的 ERLE 评价标准, SNR=30dB

声学回声消除自适应滤波器的设计要求是具有收敛速度快和稳定性高等特点。由于这个两个要求是相互矛盾的,许多传统算法是无法同时满足二者的需求。一些学者提出了许多算法解决上述矛盾,但这些算法在实际应用中并不令人满意。从上面的对比实验中可以看出本章提出的 NPVSS-NASF 算法与 VSS-NLMS 算法、SM-NLMS 算法和 NLMS 算法相比,在失调和 ERLE 两个评价标准方面都展现了更优异的性能。即使面对非平稳信号,声学回声消除的性能也比其它三种传统方法更加突出。

3.5 本章小结

本章针对回声消除问题,对回声信号进行了特征研究。并且对传统的自适应滤波回声消除算法进行了理论推导与分析,而且针对不同的自适应滤波算法对回声消除算法评价标准进行了探讨。最后根据自适应滤波算法的理论研究与分析,提出了改进的可变步长子带自适应滤波算法。用对比实验对该算法进行了分析与验证,实验结果表明改进的算法比 VSS-NLMS, SM-NLMS 和标准 NLMS 算法具有更快的收敛速度和最小的稳态误差。

第四章 文本匹配在语音控制中应用方法研究

第二章和第三章主要对多媒体播放机器人的语音信号进行了预处理从而提高语言识别的准确率,本章内容主要是基于语音识别进行文本语音匹配从而实现语音指令控制。本章对传统的文本匹配方法进行了研究,及对文本匹配遇到的挑战进行了简要的概述,并介绍了几种传统文本匹配学习模型。基于以上的研究提出了基于局部匹配的数字化文本信息检索的方法,并在本章最后对该方法进行了实验验证与分析。

4.1 引言

文本语音匹配技术近些年来发展迅速,它的成熟发展使人们与机器人的语音交互变成了可能。语音交互与传统的键盘鼠标交互相比,语音交互更加直接,更加自然、更加简洁。文本语音匹配技术是在语音识别基础上实现的,语音识别技术将语音转换成文本信息,最终目的是将获得的文本信息与参考文本进行匹配。它可以应用在模型训练,人机交互、多媒体的语音检索等一些领域。

文本语音匹配技术与语音识别技术最大区别在于,前者参考文本是已知的,后者参考文本是未知的。近年来,文本语音匹配技术逐渐成熟,多媒体语音信息检索应用该技术也越来越多。现在手机上越来越多的音乐 APP 不仅仅能播放歌曲,而且还可以根据歌曲片段寻找用户所需要的音乐。

随着数字化信息管理技术的飞速发展,现在越来越多的人已经开始使用语音信息在网络上进行搜索。然而网络信息资源是巨大的,它具有很大的自由度和随意性,并且它的分布和组合又缺乏组织结构。所以这增加了信息资源管理难度,这也是网络上信息检索、文本语音匹配的一大挑战^[176-178]。

在现有的分布式和复杂的网络环境下信息面前,传统的检索方法的局限性也变得更加清晰^[179]。一方面,信息过载带来的低召回率和精度低的问题,用户必须花费大量的时间和精力去筛选出自己想要的信息^[180]。另一方面,信息检索缺少智能化^[181]。针对这些问题,基于局部数字化的文本信息检索方法被提出,文字树和查询树构造生成祖先子孙的查询之间的关系,并将其映射到文本信息。根据局部匹配的检索方法,该矢量检索方法被用于计算所述数字化文本并提交查询之间的相似性。相似度由大至小排序,使文字树可以依次输出数字的文本信息^[182]。该方法是在信息检索更有效率并能获得全面,准确的搜索结果,可用于数字文本信息的快速检索。

4.2 文本匹配的方法研究与概述

4.2.1 文本匹配方法概述

文本匹配是自然语言理解中的一个重要基础的问题,大量的自然语言处理任务中都应用文本匹配技术。根据文本匹配的文本内容长度不同可以分为三大类:第一类主要是短文本和短文本的匹配,这种类型在工业界的应用场景出现较多。第二类主要是短文本和长文本的匹配,这种类型在工业界的应用场景出现频率也是非常高的。第三类主要是长文本和长文本的匹配,这种类型主要在个性化推荐的应用中出现广泛。这些不同的文本匹配类型都有着很大区别,即使众多学者提出了许多解决办法,针对不同的任务找到不同的匹配模型,但是仍然有一些挑战需要去解决。

(1) 词语匹配多元性

汉语博大精深,有时同一个语义可以用多个词语表示,比如人们常常吃的土豆,它的中文学名叫马铃薯,在不同的地方还有一些别称,荷兰薯、地蛋、薯仔、荷兰薯、番仔薯这些名称描述的都是同一种植物。随着网络的不断发展,网络词语的出现经常会让同一个词在不同的环境下具有不同的语义,比如种草既有狭义的一面也有广义的一面,它可以形容推荐某一个商品让他人具有购买的意愿,也可以形容向他人推荐自己喜欢的东西让他人也随之喜欢,然而种草最早的含义是指种植草这种植物。

(2) 短语匹配结构性

短语往往是由多个词语组成的,短语的语义不仅仅是由词语本身语义决定,而且还需要考虑短语结构框架。比如“机器学习”和“学习机器”,这两个短语仅仅由两个词语组成,但是不同的顺序短语语义是完全不相同的。机器学习是指一种多领域交叉学科,它使计算机具有人一样的思维,是人工智能的核心。然而学习机器就可以按照字面的含义理解为一种学习的机器。两种匹配的结构导致短语语义是完全不相同的。

(3) 文本匹配的层次性

一个长的文本都是由众多段落组成的,段落由句子组成,句子由短语组成,短语由词组成。这些段落是按照层次组成了一篇文章,不同的层次性应该考虑到不同的文本匹配信息。所以在文本匹配的过程中许多学者提出了根据上下文段落进行文本匹配的模型,因为只考虑本段内容忽略了上下文的内容是不足以完全理解本段的语义,匹配的过程往往也是局限性的。

人们为了解决文本匹配的问题,常常将文本匹配的问题转换成数学的形式展现出来,通过对数学问题的描述解决文本匹配。一般会给定两个文本 $T_1 = (w_1, w_2, \dots, w_m)$ 和 $T_2 = (v_1, v_2, \dots, v_m)$,以文本检索为例,这两个文本其中一个文本属于要查询的选

项, 另一个文本属于被查询的文档 $\text{mathc}(T_1, T_2)$ 是指两个文本的匹配程度, 匹配程度是由从大到小进行排序, 成功的匹配算法需要捕获匹配过程中的丰富交互结构。

4.2.2 传统文本匹配学习模型

文本匹配是一个多层次的问题, 它涉及到研究人员的一些数学模型的基础研究, 也涉及到一些应用系统上的应用研究等。由于文本匹配的难点在于自身语言的复杂性, 众多学者都从语言自身的复杂性去解决文本匹配困难的问题, 一些数学模型的产生对文本匹配的发展起到了促进的作用^[183]。文本匹配处理模型基于规则的方法可以分为分析模型、混合模型和统计语言模型三种类型。分析模型和混合模型是基于理性主义的方法, 从语法理论角度出发进行匹配, 并且涉及到了多种学科专业知识, 统计语言模型是本文介绍的重点。

统计数学模型可以很好地描述反应理论分析过程的本质, 能从真实的数据中寻找学习方法。一般来说, 统计数学模型方法是不需要先验知识的, 它是需要进行预先训练的, 当训练到一定程度后可以应用。人们一般都是根据不同的语音特征基选择不同的数学模型, 传统的文本匹配学习数学模型一般有 N-Gram 模型、向量空间模型、词袋模型和深度神经网络模型。

(1) N-Gram 模型

N-Gram 语音模型就是将 N 个连续的词作为一个单位进行处理, 它经常在大词汇的语言识别中使用, 它也被称作一阶马尔科夫链。该模型利用上下文信息中的相邻词语之间的关系得出最大概率的语句, 从而实现两段文本的匹配。例如: “I want to eat a cake.” 通过处理可以变成为[“I want”: 1, “want to”: 1, “to eat”: 1, “eat a”: 1, “a cake”: 1]。也就是说第 N 个词指与它前面的一个词相关, 不考虑其它的词, 从而一段文本的概率就是每个词语出现的概率乘积, 这些概率是通过统计 N 个词同时出现的次数算出来的。

(2) 向量空间模型

Salton^[113]等人提出了向量空间模型并且将其应用到文本检索系统当中, 获得了显著的效果。该算法把文本内容匹配简单地转换成空间向量运算, 从而以空间的相似度获取文本的相似度, 这种模型比较直观易懂。向量空间模型的主要思想就是将文档以向量的形式表现出来, 向量中的每个元素表示着词语。然后通过向量运算的方法表示文档之间的相似度, 常见的运算方法就是余弦相似度计算。

$$\text{sim}(d_1, d_2) = \frac{V(d_1) \cdot V(d_2)}{|V(d_1)| |V(d_2)|} \quad (4-1)$$

(3) 词袋模型

词袋模型是利用一个大的文本集去建立一个词库表,在这个词库表中会出现所有的词语,就犹如一个口袋里装着所有的词汇,它们每个个体是独立的,没有任何关联的。在词袋模型中可以忽略词语的顺序及语法,对一个词语在词袋中出现的次数进行排序,用向量形式表示成 $s=\{(\text{word}_1, \text{value}_1), \dots, (\text{word}_i, \text{value}_i), \dots, (\text{word}_n, \text{value}_n)\}$,其中 s 表示任意句子, word_i 表示词表的第 i 个词汇, value_i 表示第 i 个词汇出现的频率。但是词袋模型中每个词语的权重是一样的,在实际应用中是不符合实际的,在文本匹配中有些关键词语存在的重要性远远大于一些普通词语出现的频率。

为了解决词袋模型的权重问题,许多学者提出了 TF-IDF 统计模型,它的核心思想就是一个词语在某一段文本出现的频率较高,在其它的文本出现频率较低,则可以认为该词语具有一定的区分能力。TF 是 Term Frequency 的缩写,代表着词频它是指一个词语在文本中出现的频率。IDF 是 Inverse Document Frequency 的缩写,代表着逆向文件频率,它是指某个一词在该文本中的重要程度。 $\text{TF-IDF} = \text{TF} \times \text{IDF}$,其计算方法如下式:

$$\text{tfidf}_{ij} = \text{tf}_{ij} \times \text{idf}_j = \frac{n_{ij}}{\sum_k n_{kj}} \times \log\left(\frac{|D|}{|\{j: t_i \in d_j\}|}\right) \quad (4-2)$$

(4) 深度学习模型

由于传统文本匹配模型需要大量的人工的标注和特征提取^[184-186],如何配置一个优异的文本匹配模型是人们需要解决的核心问题,所以人们也逐渐地从传统文本匹配模型转移到了深度学习模型。许多深度学习模型是基于传统模型进行改进的,例如在自动问答领域中,传统的模型都是基于一个庞大的知识库检索^[187-189],也可以认为是对数据库中数据的结构检索,深度学习匹配的模型也是基于这个知识库构建的,主要思想是对文本的语义提取和句子模式识别^[190-191];在信息检索领域中,深度学习匹配模型也是利用文档模型和 LDA 模型的相结合,提取词的特征向量进行匹配^[192-193]。利用深度学习方法可以节省大量人工,而且传统文本匹配模型具有局限性,不能同时满足多种场景的文本匹配,而深度学习模型根据训练的数据不同可以适应各个文本匹配的任务当中,深度学习与词向量的相结合更好地解决了文本匹配的多元性、结构性及层次性^[194-199]。深度学习模型在文本匹配中与传统的文本匹配模型相比主要有以下优势,在深度学习模型中,可以用语义空间向量表示文本中的词语,利用向量之间运算可以更加准确的描述两个词语之间的关系;在深度学习模型中,由于自身具有结构性和序列性,所以它可以更准确地去描述文本中结构化和层次化;深度学习模型可以更好地去利用较大规模的数据库,由于计算机的软件和硬件技术的不断提升使深度学习模型更加快捷与方便^[200-202]。

针对深度学习模型,通常包括以下几个步骤,首先定义 $s_1 = \{x_i\}_{i=1}^n$, $s_2 = \{y_i\}_{i=1}^n$

表示两个文本样本 s_1 和 s_2 中的单词序列，其中 n 和 m 表示文本的长度， x_i ， y_i 表示句子中的单词。

- (1) 文本表达
- (2) 文本交互
- (3) 匹配空间模式提前
- (4) 匹配程度得分

在深度学习模型中，可以认为为对两个文本基于深度学习用特征向量进行表达，通过计算这两个向量之间的相似度输出两个文本的匹配度，图 4-1 为深度学习匹配模型框架。

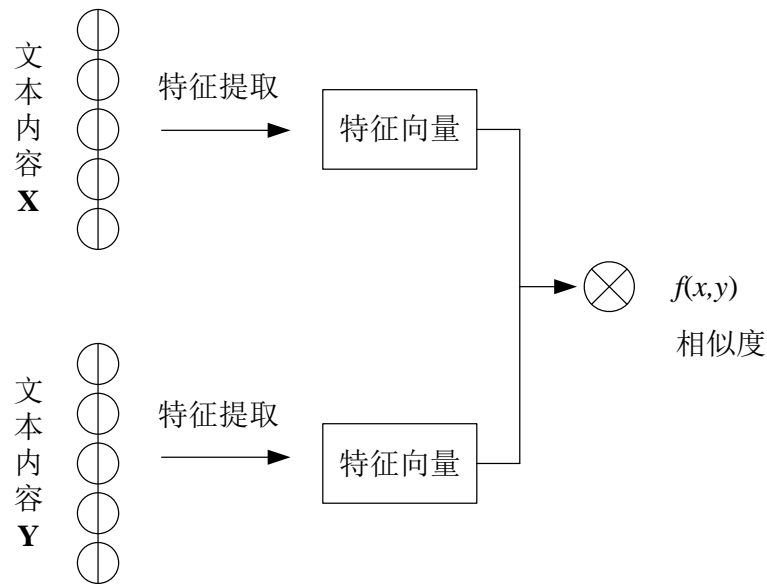


图 4-1 深度学习文本匹配框图

4.3 基于局部匹配的数字化文本信息检索方法

在本文中，将数字文本信息描述为一种名为“数字文本树”的树其描述公式表示如下。

$$d = (r, SN, TN, T, \prec_{snh}, \tau, \sigma) \quad (4-3)$$

其中， r 是树中的虚拟根节点，表示整个文本； SN 是树中结构节点的集合； TN 是树中文本节点的集合； T 是树中所有结构节点类型的集合； \prec_{snh} 描述树中结构节点之间的父子关系。

假设， $(sn_i, sn_j) \in SN$ 并且 sn_i 是 sn_j 子，则 $sn_i \prec_{snh} sn_j$ 。 τ 是 SN 到 T 的映射， σ 是 SN 到 $TN \cup \{NULL\}$ 的映射。

定义文本树中的路径 $p = (sn_1, sn_2, sn_3, \dots, sn_m)$ 表示结构节点 sn_1 和结构节点 sn_m 之

间的路径,描述 sn_1 和 sn_m 之间的祖先-后代关系。 $head(p)$ 和 $tail(p)$ 分别表示路径的起点和终点,节点之间的距离被定义为 $dist(x,y)=|p|-1$, 其中 $head(p)=x, tail(p)=y$ 和 $|p|$ 表示在所述路径中的节点的数目。

在文本树中,该组结构节点的后代节点的由公式 (4-4) 表示:

$$DESC(sn)=\{sn_i|\exists p((head(p)=sn)\wedge(tail(p)=sn_i))\} \quad (4-4)$$

设置结构节点的文本信息由公式 (4-5) 表示:

$$CONTENT(sn)=\left\{tn_j\left|\begin{array}{l}(sn_i\in DESC(sn)\cup\{sn\})\wedge(tn_j=\sigma(sn_i))\\ \wedge(\sigma(sn_i)\neq NULL)\end{array}\right.\right\} \quad (4-5)$$

将子文本被定义为 $sd_n=(n,SN_n,TN_n)$, 其中 $n\in SN$, $SN_n=DESC(N)$ 和 $TN_n=\{tn|(tn=\sigma(sn))\wedge(sn\in SN_n)\}$ 。也就是说,子文本 sd_n 是文本树中以节点 n 为根的子树,子文本的类型与节点 n 的类型相同 $TYPE(sd)=\tau(n)$, 子文本的内容是 $CONTENT(sd_n)=CONTENT(n)$ 。

与文本的描述类似,包含结构信息的提交查询被描述为查询树,查询树的每个节点都是子查询,提交的查询中必须有文本信息的描述。

在利用本文方法搜索文本信息的过程中,查询树上的每个子查询都是自下而上逐步处理的。同时,使用向量搜索计算子查询与子文本之间的相似度,最终得到文本与整个查询之间的相似度。根据相似度排序,输出文本信息的检索结果,详细搜索过程如下。

首先,子查询 q 的向量是 $q=(sq_1,sq_2,sq_3,\dots,sq_m)$ 并且 $sq_i\in DESC(q)$, $sq_i\in DESC(q)$,其中 $result(sq)$ 表示子查询 sq 的搜索应满足的匹配结果,如公式(4-6)所示。

$$result(sq)=\left\{n\left|\begin{array}{l}(n\in SN)\wedge((\tau(n)=sq)\vee(sq\in CONTENT(n)))\\ \wedge relD(n,sq)\end{array}\right.\right\} \quad (4-6)$$

其中, $relD(n,sq)$ 意味着结构特征并且应该满足下列关系,即:

$$relD(n,sq)=\exists n_d\exists sq_d\left\{\begin{array}{l}(n_d\in DESC(n))\wedge(sq_d\in DESC(sq))\\ \wedge(n_d\in result(sq_d))\end{array}\right\} \quad (4-7)$$

从 $result(sq)$ 中可以看出,检索到的结果文本树可能并不完全包含查询树,查询树是不完全匹配过程,即局部匹配过程。因此,只要在任何子查询中节点之间的关系满足祖先后代,就认为满足查询要求,从而实现了局部匹配的过程。

在局部匹配检索过程中,每种方法都需要按照文本树中根遍历的顺序构造相应的

结果子文本，并计算子查询与子文本之间的相似度。这里相似度函数被定义为 $SIM(sd, q)$ ，它表示子文本 SD 与其匹配子查询之间的相似度。为了计算相似函数，需要进一步确定向量中元素的值。

在执行局部匹配时，查询向量的每个元素的值应根据查询 Q 与其后代之间的距离来确定。本文的值是距离的倒数，即：

$$w_{qi} = QW(dist(q, sq_i)) = 1 / dist(q, sq_i) \quad (4-8)$$

查询之间的距离越大，与用户查询的关联度越低。

假设 $w_{result(sq_i)}$ 是子查询 sq_i 上的子文本向量的每个元素的值，子文本向量可以用下列公式来描述：

$$sd = (w_{result(sq_1)}, w_{result(sq_2)}, \dots, w_{result(sq_m)}) \quad (4-9)$$

采用传统的文本分析方法，找出子文本向量中各元素的值，根据文本的特殊性，采用公式（4-10）计算信息局部匹配过程中关键字子查询的权重值：

$$w_{result(qt)} = \frac{f_{qt}}{|CONTENT(sd_{sup})|} \cdot \log \frac{n_{sup(qt)}}{N_{sup}} \quad (4-10)$$

其中， sd_{sup} 是包含关键字 qt 的子文本并且满足前面的子查询； f_{qt} 是在文本信息集中的关键字 qt 的出现频率；设置 $CONTENT(sd_{sup})$ 中包含的子文本文字信息的总长度； $n_{sup(qt)}$ 是包含关键字 qt 和查询匹配子文本 q_{sup} 的数量； $n_{sup(qt)}$ 是为所有子文本匹配查询次数。

最后，根据余弦匹配系数法，获得的潜台词，并在本地匹配处理子查询之间的相似性：

$$SIM(sd, q) = \sum_{i=1}^n w_{qi} \cdot w_{result(qi)} / \sqrt{\sum_{i=1}^n w_{qi}^2 \cdot \sum_{i=1}^n w_{result(qi)}^2} \quad (4-11)$$

在自下而上处理子查询的过程中，利用上述方法得到数字文本与提交查询的相似度。根据相似度，文本树依次输出数字文本信息。

4.4 实验结果与分析

4.4.1 实验设置

为了验证本文新提出的方法的检索效果，本文是用数字农业文本信息来进行文本信息检索实验。实验主要分为两个大组，第一组是小规模的数据集（组 A）和第二组大规模数据组（组 B）进行。集合 A 是从已分类的农业网页随机选择的，它大概有 2000 个农业的文本信息，包括农业科技，农业新信息，农产品和农业发展四个方面，主要用于评估检索准确率和召回率。集合 B 是从由农业搜索引擎捕捉到的 120 万个

网页中提取的，主要用于评估模型的检索效率。

为了让实验更有说服力，本章所述方法与基于双语义空间和基于最大权重匹配计算两种方法进行比较。

（1）文本匹配准确率和召回率的实验设置

在组 A 的实验中，具体的实验设置如下，实验文本信息数量分别是 100, 200, 300, 400 和 500，然后用双语义空间、最大权重匹配和局部匹配三种计算方法进行准确率和召回率的比较实验，每个样本实验进行 5 次。为了保证实验更有说服力，验证算法的稳定性，将 10%，20% 和 30% 综合新闻信息添加到测试集合 A 进行干扰，然后在进行文本信息匹配。最后对上面的三个对比实验重新记录结果。

（2）文本匹配检索效率实验设置

在组 B 的实验中，对集合 B 中的 20 万件农业文本信息分别用基于双语义空间、最大权重匹配和局部匹配三种计算方法进行信息检索，然后记录三种方法的检索时间，将三组数据进行比较。

因为众所周知基于最大权重匹配计算农业信息检索方法比双语义空间方法检索需要更少的时间。为了进一步验证本文所提出的方法具有检索时间少的优点。本研究基于最大权重匹配计算农业信息检索方法与本文方法进行比较。具体的实验设置如下：实验分为 5 次，每次的实验网页数分别为 50,000、100,000、150,000、200,000、250,000，将这两种方法的耗时情况记录下来。

为了让实验更有说服力，本文将实验组 B 在三个阶段进行检索效率实验。第一阶段：40 万次农业网页检索；第二阶段：80 万次农业网页检索；第三阶段：120 万次农业的网页进行检索。

4.4.2 评价标准

对一个算法进行评价就需要一些算法的评价指标，在文本检索当中，准确率和召回率是两个基础的评价指标。召回率是指检索到文本内容集合占总文本总数的相关比例。准确率是指检索出相关的文本内容集合占检索出文本集合的比例。它们对文本检索非常重要，单一指标不能准确的评价检索算法的有效性，所以需要多个评价指标综合考虑算法的可靠性和优越性，文本检索结果分布可以用图 4-2 表示。

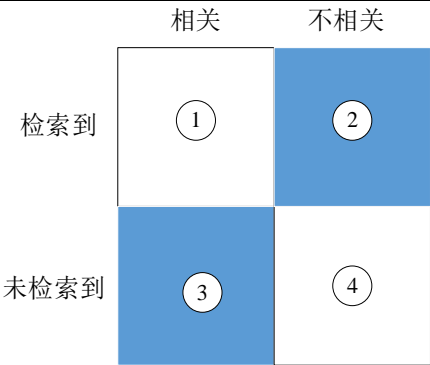


图 4-2 文本检索结果分布图

从图中可以看出 A 检索到相关文本内容集合,B 是检索到不相关文本内容集合, C 表示未检索到相关文本内容集合, D 表示未检索到不相关文本内容集合。那么本文可以用以下公式表达召回率和准确率。

$$\text{召回率} = \frac{A}{A+C} \times 100\% \tag{4-12}$$

召回率是代表检索系统找到所有相关文本信息集合的能力。

$$\text{准确率} = \frac{A}{A+B} \times 100\% \tag{4-13}$$

4. 4. 3 结果与讨论

三种方法的召回率实验结果,如图 4-1 所示。当实验组 A 的文本信息数目是 100, 200, 300, 400 和 500 的基础上,最大权重匹配计算的方法召回率分别为 96.8%、97.0%、96.8%、96.8%和 96.9%。其结果平均值为 96.8%;

在相同的情况下,通过双语义空间方法获得的召回率分别为 98.8%、98.8%、99.2%、97.6%和 99.2%。然而,当实验组 A 的文本信息数目是 400 时,召回率只有 97.6%,与其它文本信息数目相比召回率相对较低,这表明了该方法的稳定性不如局部匹配的方法稳定性。从图 4-3 中条形图的长度可以看出,基于局部匹配方法的召回率是这三种方法中最高的,并且平均值为 99%以上。以上数据表明,本文提出的方法的信息检索速度是最高的。

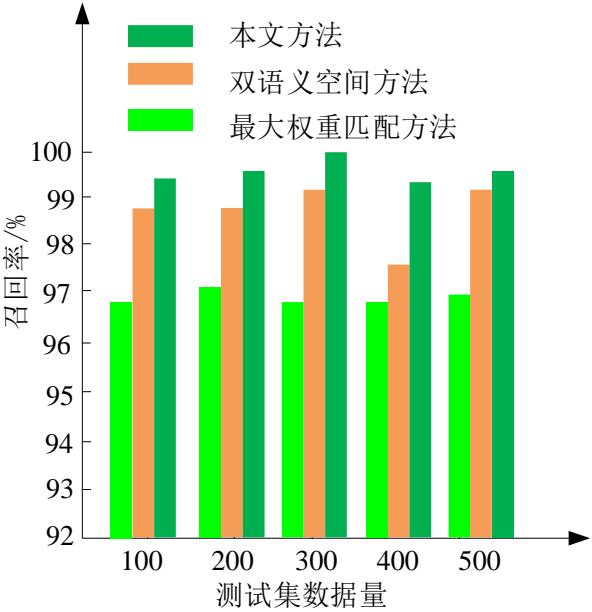


图 4-3 三种方法召回率对比实验

为了确保实验的准确性，本文对集合 A 进行干扰实验并且将 10%的干扰信息加入到集合 A 中，然后用三种方法进行信息检索对比实验，实验比较结果如图 4-4 所示。

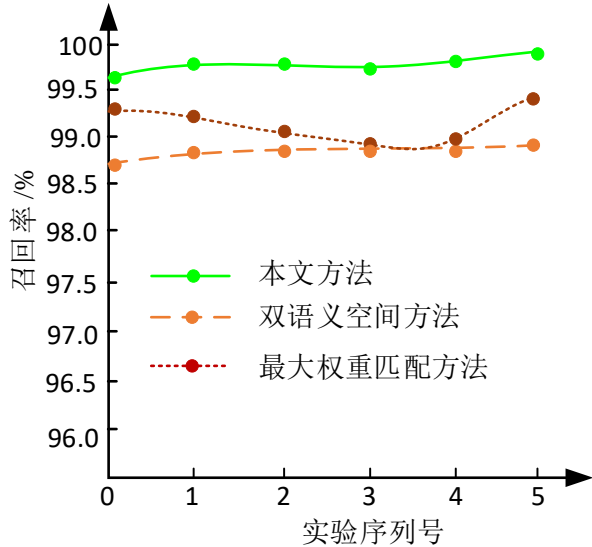


图 4-4 加入 10%干扰信息三种方法召回率对比实验

从图 4-4 实验比较结果中可以看出，基于局部匹配方法的曲线位于图中的最顶端，这表明了该方法具有最高的召回率。在这个实验中，基于局部匹配方法的召回率分别为 99.8%、99.8%、99.7%、99.8%和 99.9%。从基于局部匹配方法的召回率数据看出，在五组实验中召回率略有小幅度的波动，而且 5 组实验数据都在 99.5%以上。这表明，基于局部匹配方法的数字信息检索是全面优于其它两种方法，并且该方法稳定性也非常好。

为了确保实验的多样性，本文将 20%的干扰信息加入到集合 A 中，然后用三种方法进行信息检索对比实验，实验比较结果如图 4-5 所示。

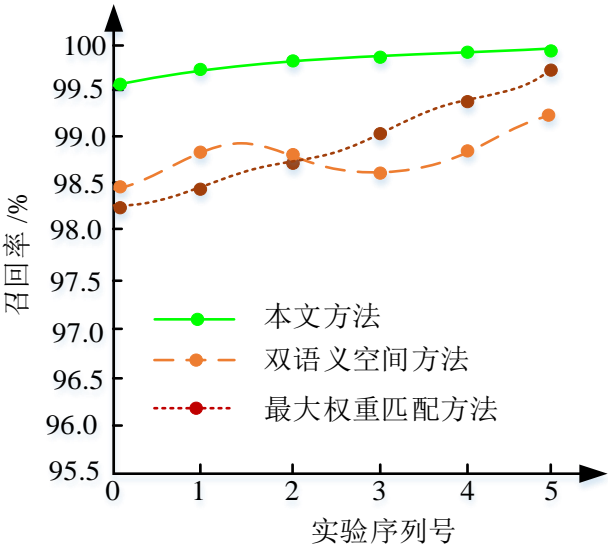


图 4-5 加入 20%干扰信息三种方法召回率对比实验

从图4-5实验比较结果中可以看出，基于局部匹配方法的曲线位于图中的最顶端，这表明了该方法具有最高的召回率。在该实验中，基于最大权重匹配计算的方法召回率与实验检文本信息数目基本成正比。该方法在 5 个实验中得到的召回率是 98.4%、98.6%、99.0%、99.3%和 99.6%时，该组的数据表明基于权重匹配技术方法具有很大的潜力，但不稳定。在这五组实验中基于局部匹配文本检索召回率分别是 99.7%、99.7%、99.8%、99.8%和 99.9%，这五组数据中可以看出，本文提出方法的召回率没有因干扰信息的增加而降低，进一步说明基于局部匹配文本检索方法具有良好的稳定性。

为了验证干扰信息对文本检索召回率的影响，本文将 30%的干扰信息加入到集合 A 中，然后用三种方法进行信息检索对比实验，实验比较结果如图 4-6 所示。

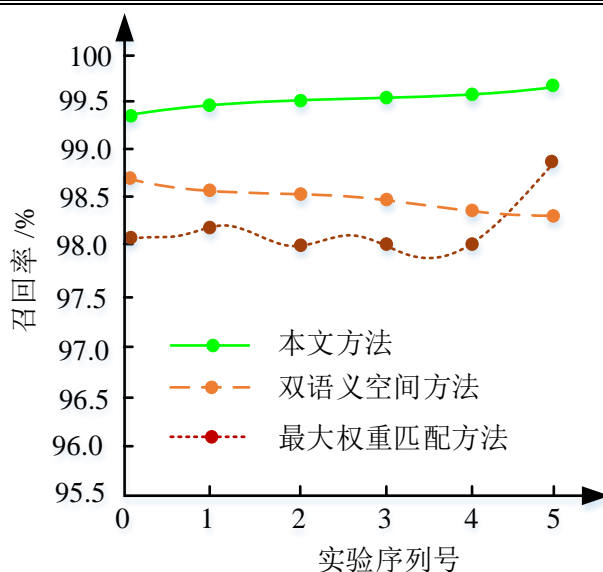


图 4-6 加入 30%干扰信息三种方法召回率对比实验

从图 4-6 实验比较结果中可以看出, 随着干扰信息在实验组 A 中比例不断提高。基于局部匹配文本检索方法的召回率也随之下降。但与其它两种方法相比, 下降幅度较小, 并且在 5 个实验获得的召回率分别是 99.5%、99.5%、99.6%、99.6%和 99.7%, 仍然具有较高水平。同时也说明, 三种方法文本检索召回率都是受干扰信息影响的, 只是基于局部匹配文本检索方法影响较小。

综上所述, 本文可以得出基于局部匹配文本检索方法与其它两种方法在检索数字农业文本信息的比较过程中, 具有较高的召回率和较好的鲁棒性。在验证文本信息检索准确率的实验中, 本文将上面不加干扰信息的五组实验中进行多样本化。然后对每组实验结果取平均值。

基于最大权重匹配计算信息检索方法的准确率实验结果如表 4-1 所示。

表 4-1 基于最大权重匹配计算信息检索方法的准确率

| 实验次数 | 第一组实验 | 第二组实验 | 第三组实验 | 第四组实验 | 第五组实验 |
|---------|-------|-------|-------|-------|-------|
| 50 | 94.4 | 94.5 | 93.4 | 94.5 | 94.6 |
| 100 | 94.5 | 94.3 | 93.4 | 94.3 | 94.2 |
| 150 | 94.6 | 93.2 | 94.5 | 93.2 | 94.5 |
| 200 | 94.2 | 93.9 | 94.2 | 93.9 | 94.8 |
| 250 | 94.5 | 93.7 | 93.5 | 93.5 | 93.8 |
| 300 | 94.8 | 93.8 | 93.5 | 94.8 | 94.8 |
| 350 | 94.8 | 94.8 | 94.8 | 93.8 | 94.9 |
| 400 | 94.8 | 94.9 | 93.8 | 94.1 | 93.4 |
| 450 | 93.5 | 93.9 | 94.1 | 93.8 | 93.4 |
| 500 | 94.9 | 94.7 | 93.8 | 93.8 | 94.5 |
| 准确率 (%) | 94.5 | 94.1 | 94.0 | 94.0 | 94.3 |

从表 4-1 中可以看出基于最大权重匹配计算农业信息检索方法的准确率约为 94.2%。

基于双语义空间的信息检索方法的准确率实验结果如表 4-2 所示。

表 4-2 基于双语义空间的信息检索方法的准确率

| 实验次数 | 第一组实验 | 第二组实验 | 第三组实验 | 第四组实验 | 第五组实验 |
|---------|-------|-------|-------|-------|-------|
| 50 | 95.8 | 95.4 | 96.8 | 96.5 | 96.5 |
| 100 | 96.2 | 95.6 | 97.1 | 96.8 | 96.6 |
| 150 | 96.8 | 96.8 | 97.2 | 96.7 | 95.6 |
| 200 | 96.7 | 96.5 | 96.2 | 95.9 | 95.8 |
| 250 | 95.9 | 96.6 | 96.4 | 95.8 | 96.5 |
| 300 | 95.8 | 95.6 | 97.1 | 95.6 | 95.8 |
| 350 | 96.8 | 95.8 | 96.5 | 95.8 | 96.7 |
| 400 | 96.5 | 96.2 | 95.6 | 97.1 | 95.6 |
| 450 | 95.8 | 96.8 | 95.8 | 97.2 | 96.8 |
| 500 | 96.7 | 95.1 | 96.5 | 96.2 | 96.5 |
| 准确率 (%) | 96.3 | 96.0 | 96.5 | 96.3 | 96.2 |

从表 4-2 中可以看出基于双语义空间的农业信息检索方法的准确率约为 96.3%。

基于本文方法的信息检索方法的准确率实验结果如表 4-3 所示

表 4-3 基于本文方法的信息检索方法的准确率

| 实验次数 | 第一组实验 | 第二组实验 | 第三组实验 | 第四组实验 | 第五组实验 |
|---------|-------|-------|-------|-------|-------|
| 50 | 99.5 | 99.2 | 99.3 | 99.3 | 100 |
| 100 | 99.5 | 99.8 | 99.6 | 99.4 | 99.5 |
| 150 | 99.1 | 99.3 | 99.7 | 100 | 99.1 |
| 200 | 99.2 | 99.4 | 99.4 | 99.6 | 99.2 |
| 250 | 99.5 | 99.6 | 99.8 | 99.8 | 99.5 |
| 300 | 99.6 | 99.6 | 99.4 | 99.9 | 99.8 |
| 350 | 99.8 | 99.5 | 99.6 | 99.5 | 99.9 |
| 400 | 99.9 | 99.8 | 99.6 | 99.1 | 99.5 |
| 450 | 99.7 | 99.7 | 100 | 99.2 | 99.1 |
| 500 | 100 | 99.6 | 99.4 | 100 | 99.2 |
| 准确率 (%) | 99.6 | 99.6 | 99.6 | 99.6 | 99.5 |

从表 4-3 中可以看出基于局部匹配的农业信息检索方法的准确率约为 99.6%。

为了清楚地表达的方法检索的准确度，将上述表 4-1、表 4-2 和表 4-3 准确度的平均值制作成图，如图 4-7 所示。

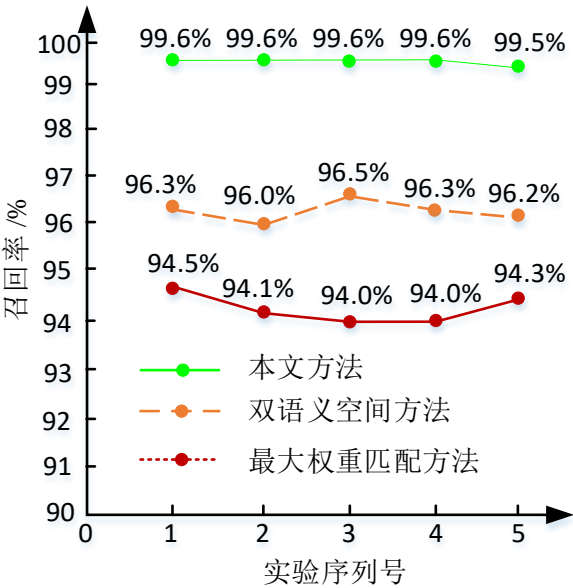


图 4-7 三种比较方法检索准确度

可以从图 4-7 中可以清楚地看到这三种方法的平均值的比较。本文提出基于局部匹配信息检索方法的准确度量比基于双语义空间方法的准确度量要高约 3% 的。比基于最大权重匹配计算的文本检索方法的准确度高约 5% 的。综上所述，基于局部匹配文本信息检索方法可以用于数字化的农业信息有效的检索。

为了验证基于局部匹配文本信息检索时间的优越性，本文用这三种方法做了比较

实验，实验比较结果如表 4-4 中所示。

表 4-4 三种方法文本信息检索时间

| 测试页数 | 本文方法 | 最大权重检索 | 双语义空间 |
|---------|------|--------|-------|
| 2 | 2.5 | 3.2 | 6.8 |
| 4 | 3.2 | 4.2 | 8.2 |
| 6 | 3.2 | 5.1 | 10.2 |
| 8 | 4.2 | 6.8 | 13.8 |
| 10 | 4.8 | 7.2 | 15.4 |
| 12 | 5.2 | 9.1 | 19.4 |
| 14 | 5.5 | 9.9 | 23.4 |
| 16 | 6.8 | 11.2 | 26.7 |
| 18 | 8.6 | 12.9 | 28.9 |
| 20 | 9.1 | 15.2 | 32.2 |
| 平均时间 /s | 5.3 | 8.5 | 18.5 |

从表 4-4 中可以看出，基于局部匹配文本检索方法检索时间分别为 2.5s、3.2s、3.2s、4.2s、4.8s、5.2s、5.5s、6.8s、8.6s 和 9.1s，平均时间为 5.3S。并且该方法文本检索时间随实验组的检索数量而的增加。而其它两种方法的平均检索时间分别为 8.5s 和 18.5s，和本文提出方法相比检索效率显著降低。

图4-8显示了本文提出基于局部匹配方法和基于最大权重文本匹配的所需要信息检索时间。

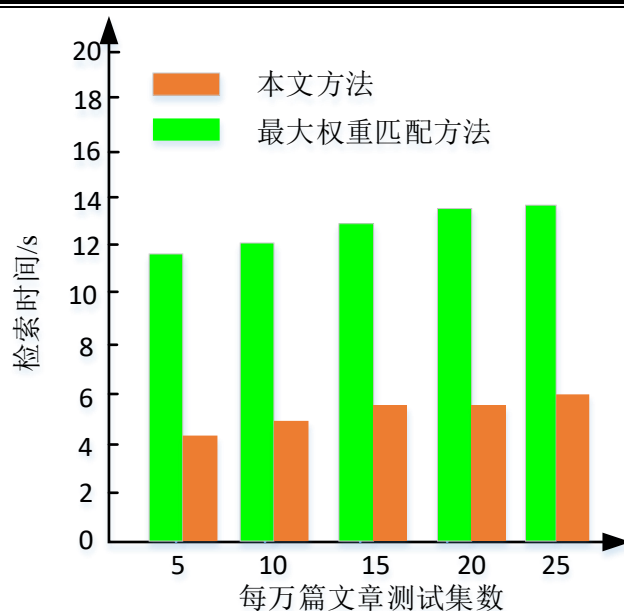


图 4-8 局部匹配方法和最大权重文本匹配的所需要信息检索时间

从图 4-8 中可以看出, 基于局部匹配的检索农业信息的方法的时间消耗分别是 4.1s、5.0s、5.5s、5.5s 和 6.0s。在测试的初始阶段, 测试页的数量是 50000 至少 4.3s, 在测试的最后阶段, 测试页的数量是 25 万, 最大时间仅为 6.0s。

这两种方法的条形图比较, 基于局部匹配方法的条形图明显优于基于最大重量约匹配计算农业信息检索方法。并且这两种方法的检索时间与实验组的检索数量的增加而增加。从表 4-4 和图 4-8 中的数据的结果可以看出, 本文提出基于局部匹配方法在较短的时间中检索数字农业信息具有高效率的优点。

在三个阶段的三种方法的检索效率, 参照分别描述图 4-9, 图 4-10 和图 4-11 中。

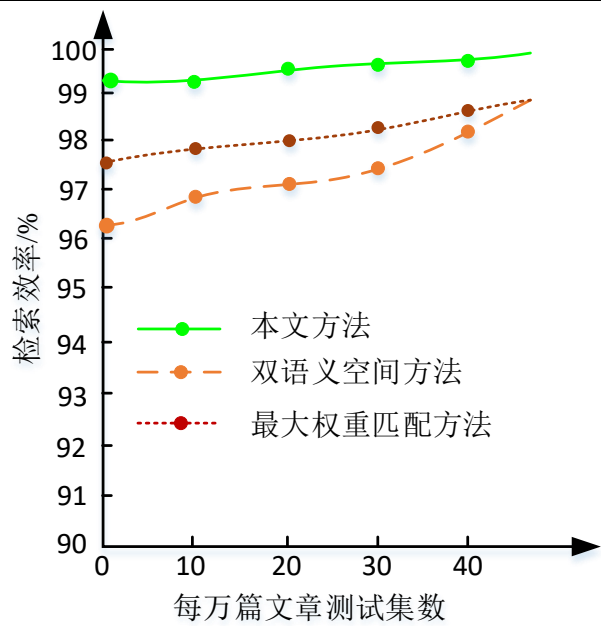


图 4-9 三种方法检索效率在第一阶段中的比较

从图 4-9 可以看到的是，在第一阶段中，三种方法的检索效率的测试组的数目的增大而增大，显示出更好的状态。通过该方法获得的检索效率是 99.2%、99.1%、99.4%、99.4%和 99.5%，它们位于在图的顶部，并有三种方法中的最高效率。

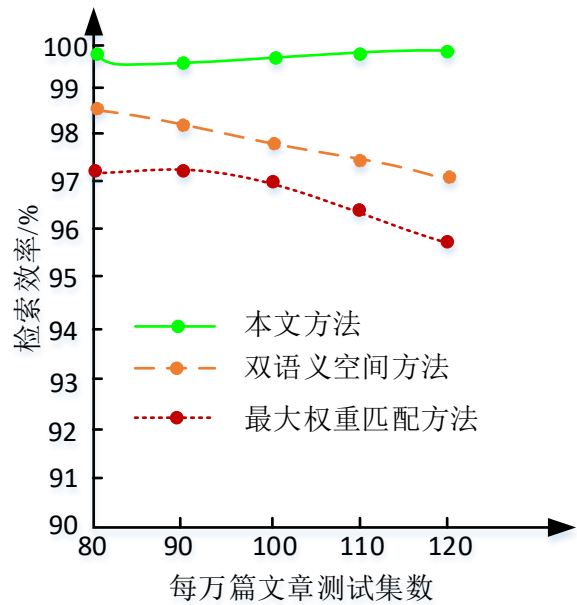


图 4-10 三种方法检索效率在第二阶段中的比较

图 4-10 的分析表明，在第二阶段中，所提出的方法的检索效率仍然 99.5% 以上。其他两种方法的检索效率降低，并用的试验组的数目的增加而大大地波动。一般来说，测试集的数量增加，其它两种方法的检索效率显著降低。这种方法的检索效率仍然是稳定且具有高效率的优势。

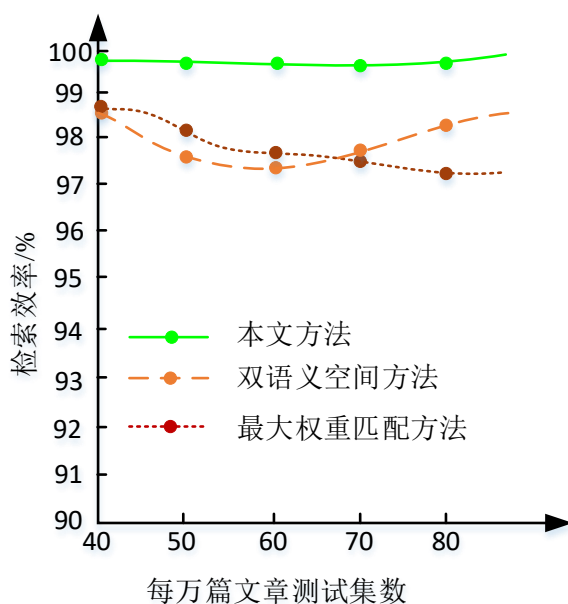


图 4-11 三种方法检索效率在第三阶段中的比较

从图 4-11 可看出的是,在第三阶段中,本文所提出的方法的检索效率仍然在 99.5% 和 99.8% 之间。实验结果表明,本文提出的基于局部匹配的数字农业信息检索的方法是具有时间短、效率高的特点,平均检索时间为 4s 和 6s 之间,平均检索效率是 99% 以上,它在数字农业信息快速检索的应用具有很好的效果。

现有的数字农业的文本信息,以提高农业数字化文本信息的检索结果,提高检索效率的检索方法的局限性,用于搜索基于局部配套数字农业的文本信息的方式提出。农业文字树和查询树构造生成祖先子孙的查询之间的关系,并将其映射到农业文字当中。根据局部匹配的检索方法,所述检索矢量方法被用来计算数字农业文本并提交查询之间的相似性,相似度由大至小排序,使农业文字树可以依次输出数字农业的文本信息。

实验证明,该方法能获得召回率和准确率较高,同时花费更少的时间和具有效率高的优点。本文提出的方法不仅弥补了传统方法的缺点,而且还提供了数字文本信息,它提供了信息处理时的参考、有效、科学的检索方法。

4.5 本章小结

本章对文本匹配的方法进行了概述,深入的研究了文本匹配的关键问题和目前存在的挑标,并对传统的文本匹配模型进行阐述,而且提出了基于局部匹配的数字农业信息检索方法,最后在仿真实验中与基于最大权重匹配计算方法和基于双语义空间匹配方法在召回率、准确率和检索效率进行比较。从实验结果中可以看出,基于局部匹

配的信息检索具有更高的检索效率，更高的的准确率及更高的召回率。

第五章 语音控制多媒体播放机器人搭建及实验研究

本章概述了语音控制多媒体播放机器人的功能与系统结构，并在最后的实验中对语音控制多媒体播放机器人原型系统进行验证。通过实验验证并解决了机器人环境噪声及回声的问题，并且准确的执行用户发出的语音指令。

5.1 引言

随着社会的发展进步，机器人已经逐渐地走入了人们的生产生活中，人们在商场，医院，银行里等一些服务场所里，语音控制服务机器人随处可见。它们可以为人们进行品牌介绍、场地引导、以及简单问题的回答，它们可以和人们进行简单的交流与互动。语音控制多媒体播放机器人应用范围非常广，比如说农业、医疗、教育等方面。语音控制多媒体播放机器人使其在农村应用、农业推广成为可能，对于促进我国农业物联网发展具有重要的理论和现实意义。语音控制多媒体播放机器人在医疗方面为人们带来了便捷，也减轻了医护人员的工作压力，增加了医院办事效率。

但是上面所述的语音控制多媒体播放机器人都会面临环境噪声和回声的干扰以及语音指令失效的难题。本章针对多媒体播放机器人语音降噪、回声消除和文本匹配在语音控制中存在的问题，开展基于语音控制的多媒体播放机器人研究，并研发了语音多媒体播放机器人原型系统，从多种场景下进行实验验证。

5.2 语音控制多媒体播放机器人介绍

语音控制多媒体播放机器人是本文研究的应用成果，如图 5-1 所示。

语音控制多媒体播放机器人基本功能：

- (1) 语音交互；它可以回答用户提出的问题，与用户进行简单的交流。也可以进行语音控制，通过语音指令执行播放视频、音频、图片及文档，语音指令可以控制进行行走。
- (2) 触屏交互；用户可以通过触摸屏幕对机器人进行操作，可以通过触摸屏幕发送控制指令。屏幕显示机器人的系统状态和交互信息。
- (3) 鼠标键盘交互；机器人具有外接 USB 端口，可以通过鼠标键盘进行操作与发送指令。
- (4) 外接存储设备；语音控制多媒体播放机器人可以外接存储设备，通过控制进行播放外接存储设备的内容。
- (5) 避障功能；通过雷达检测前方是否有障碍物，并且通过雷达返回信号进行

避障和避碰。

- (6) 行走功能：机器人具有轮式机构进行行走，易操作，行动灵活而且稳定可靠。
- (7) 无线网络交互功能：将语音控制多媒体播放机器人与计算机连接在同一局域网下，可以通过计算机对机器人进行状态监测与控制。
- (8) 蓝牙控制功能：通过蓝牙通讯，用手机控制机器人的行走。
- (9) 故障诊断功能：机器人可以定期自我巡检，随时检测机器人的软件系统BUG 和硬件系统故障。



图 5-1 语音控制多媒体播放机器人

5.3 系统结构

语音控制多媒体播放机器人主要结构为计算机设备、语音采集设备、播放设备、显示设备、行走机构、通信系统及故障诊断单元模块。

5.3.1 语音控制多媒体播放机器人硬件设备

计算机设备是语音控制多媒体播放机器人的核心设备，它也是机器人的大脑。它

主要负责信息处理、数据决策、指令发送、信息储存等。所以要求计算机主板消耗低、软件兼容性强、体积较小、USB 扩展口多、稳定性强等一些优点。本研究主要采用的计算机主板是迅为三星 Exynos 4412 豪华版四核 Cortex-A9ARM Android 开发板。它单核主频为 1.6G,内存 2GB,运算能力可以达到 16000DMIPS,DC 电源输入为 5V。同频下功耗远低于同类产品,它可开发性强,并且可扩展多个接口。工作温度可在 -20sheshidu 到 80℃,十分稳定。它可以在 60MHZ 频率下运行,采用 3 级流水线结构,处理能力大大超过 8/16 位的单片机,接近 PC 机,可以支持大屏幕的液晶显示,提供功能强大的图形用户界面,另外基于 ARM 的控制器可以内嵌实时操作系统,克服了传统基于单片机的控制系统的软件“前后台”形式造成的实时性差的缺点。开发板如图 5-2 所示。



图 5-2 语音控制多媒体播放机器人主板

本研究使用的语音采集设备是科大讯飞六麦环形麦克风阵列评估板,它体积小且可以准确识别语音指令。它可以进行 360 度声源定位,具有语音功能,而且它可以对唤醒人方向进行指定波速信号增强从而抑制其他方向的声音。它远场拾音距离可以达到 5 米。它的语音降噪和回声消除效果都非常好。电源输入为 5V。麦克风阵列评估板如图 5-3 所示。



图 5-3 语音控制多媒体播放机器人麦克风阵列

本研究采用的播放设备是 10 寸高清触摸液晶，它可以播放机器人的内容，并且也可以采用触摸屏幕的方式对机器人进行操作，开发板直接对其进行供电。



图 5-4 语音控制多媒体播放机器人显示设备

行走机构采用的是基于开源的 Arduino 开发板行走设备。该设备采用直流电机驱动，而且可进行相关的二次开发。电源输入 24V，通过安卓发送 16 位进制指令进行控制。操作简单方便，且稳定性强。



图 5-5 语音控制多媒体播放机器人行走系统

机器人软件系统利用开源的 Android 作为开发平台，以 framework 定制修改为依托，针对机器人工作环境设计、制作该平台，以创新的用户交互方式操作简便，降低使用者的要求，达到农业文化传播的目的。操作界面如图 5-6 所示。

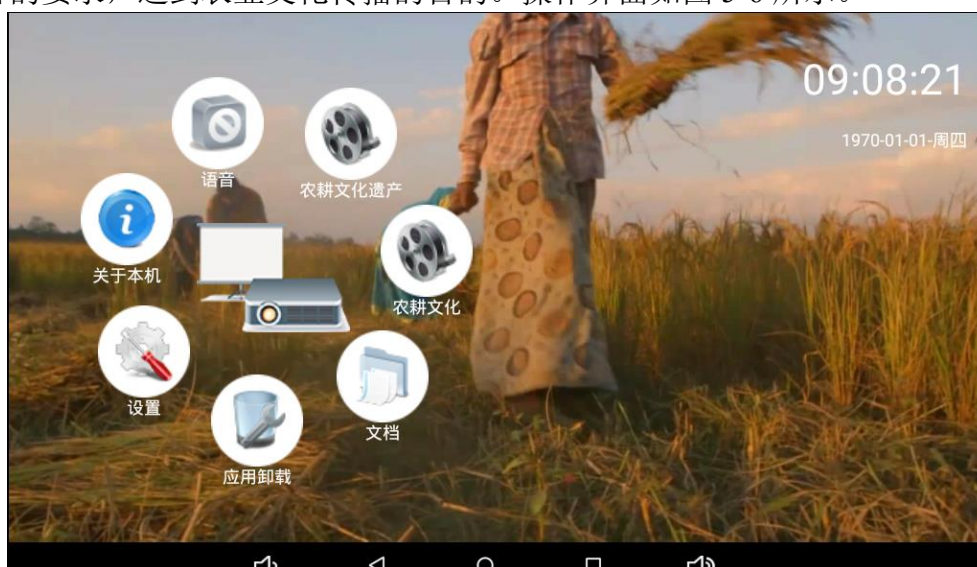


图 5-6 语音控制多媒体播放机器人系统操作界面

5.3.2 语音控制多媒体播放机器人通信系统

语音控制多媒体播放机器人采用 CAN 总线通讯方式向上位机发送数据, 采用 CAN 协议进行串行通信, CAN 总线数据传输具有可靠性高、传输速度快、体积小、远距离传输等特点, 总线图如图 5-7 所示。CAN 的工作位速率可达 1M 位/秒, 实际传输速率可能受 CAN 总线上所选择的传输数据物理层的限制, 所有数据发送和接收过滤的协议处理由 CAN 控制器独立完成, 这就使得用于 CAN 通信的 CPU 带宽最小。

CAN 总线具有以下特点:

- 1、完成对通信数据的成帧处理: CAN 总线通信接口中集成了 CAN 协议的物理层和数据链路层功能, 可完成对通信数据的成帧处理。
- 2、使网络内的节点个数在理论上不受限制: CAN 协议的一个最大特点是废除了传统的站地址编码, 而代之以对通信数据块进行编码。采用这种方法的优点可使网络内的节点个数在理论上不受限制, 这种按数据块编码的方式, 还可使不同的节点同时接收到相同的数据, 保证了通信的实时性。
- 3、可在各节点之间实现自由通信: CAN 总线上任意节点可在任意时刻主动地向网络上其它节点发送信息而不分主次, 因此可在各节点之间实现自由通信。
- 4、结构简单: 只有 2 根线与外部相连, 并且内部集成了错误探测和管理模块。
- 5、传输距离和速率: 通信距离最远可达 10KM(速率低于 5Kbps)速率可达到 1Mbps(通信距离小于 40M)。

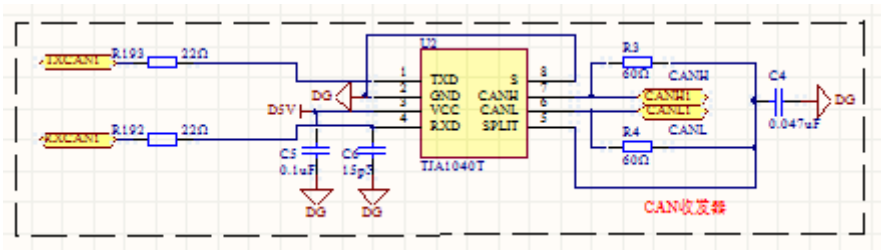


图 5-7 CAN 总线图

5.3.2 语音控制多媒体播放机器人故障诊断单元模块

在语音控制多媒体播放机器人的控制系统上增加一个故障诊断单元模块, 该单元模块主要对机器人功能层进行诊断处理。机器人故障可以分成软件系统 BUG 和硬件系统功能损坏, 软件系统 BUG 可以通过重启系统进行修复, 硬件系统功能损坏需要替换备用单元进行修复。

故障诊断单元模块定期对机器人的功能层进行定期巡检, 该单元模块向机器人功能层各个控制节点发送广播, 功能层控制节点工作正常则发生返回信号, 无返回信号

则说明控制节点发生故障。然后通过对机器人进行复位重启,如果机器人工作正常则说明是软件系统 BUG;如果机器人仍然不能正常工作则发生了硬件系统故障,需要替换备用单元进行修复。

5.4 总体实验验证

针对前面提出的以语音识别系统为核心进行语音控制,根据不同的对话人、不同应用场景进行模型训练。语音控制与语音识别最大差别在于语音控制系统中,参考文本内容是已经获取的,本研究只需进行文本匹配执行命令即可。语音控制部分程序框图如图 5-8 所示。本文通过以下的实验设置和实验评估对语音控制多媒体播放机器人原型系统进行验证。

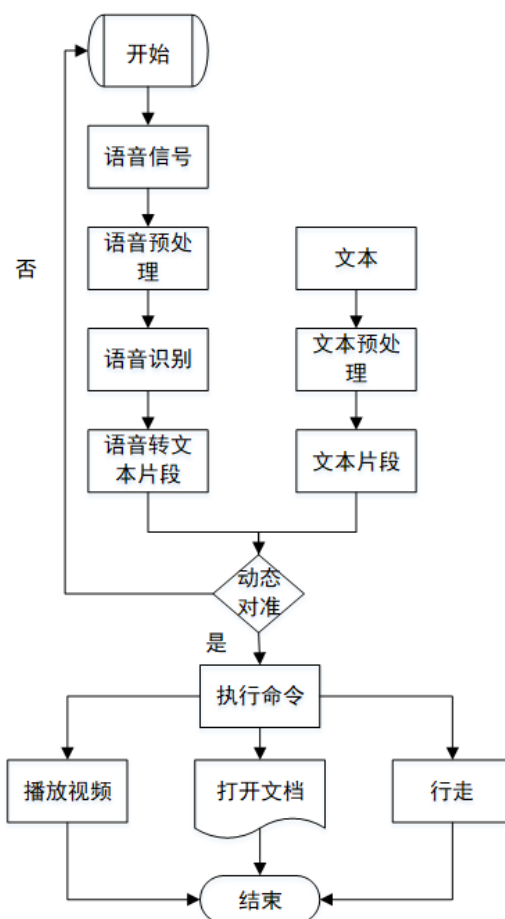


图 5-8 机器人语音控制程序框图

5.4.1 实验设置

本文的实验一共分为两大类:第一类实验是不同的用户在相同的现实环境下与语音控制多媒体播放机器人进行交互;第二类实验是相同的用户在不同的现实环境下与

机器人进行交流。本实验找到了 6 位志愿者（A、B、C、D、E、F）在室内与语音控制多媒体播放机器人进行语音交互。A、B、C 是男性，D、E、F 是女性，他们都是中国人且来自不同的地方。他们的年龄范围是 25 岁到 33 岁，平均年龄大约为 29.5 岁。他们所有人都讲普通话，但是 B 的湖北口音较重，F（汉族）稍微有一些新疆口音，其他人接近标准普通话。这些志愿者都是来自中国农业大学的学生。为了保证实验的可信度，本文使用多次实验来测评。

（1）个性化语音分析

在第这一组实验中，6 位志愿者都在同一安静的室内与机器人进行语音交流，并且每个志愿者距离机器人为 1m，每一个志愿者与机器人交谈 10 次，并且交谈的内容相同。6 位志愿者都会对机器人说“小农”，对其进行唤醒，然后说 10 次不同的语音指令。本研究测试一下不同性别的人与不同口音的人是否对语音控制多媒体播放机器人产生影响，其结果如表 5-1 所示。

表 5-1 不同人群对语音控制多媒体播放机器人语音控制的影响

| 名称 | 性别 | 年龄 | 家乡 | 语音质量 | 测试次数 | 成功次数 |
|----|----|----|-----|----------|------|------|
| A | 男 | 31 | 黑龙江 | 标准普通话无口音 | 10 | 10 |
| B | 男 | 25 | 湖北 | 普通话较重口音 | 10 | 10 |
| C | 男 | 32 | 山东 | 标准普通话无口音 | 10 | 10 |
| D | 女 | 30 | 山东 | 标准普通话无口音 | 10 | 10 |
| E | 女 | 26 | 安徽 | 标准普通话无口音 | 10 | 10 |
| F | 女 | 33 | 新疆 | 普通话较轻口音 | 10 | 10 |

从上表中可以看出，不同口音、不同性别和不同年龄的用户不会影响语音控制多媒体播放机器人语音拾取。这些个性化因素是不会影响其语音控制指令应用的效果。所以在后面的对比实验中，本研究可以忽略个性化因素的影响。

（2）语音播放识别距离分析

在日常的生活中，人们因自己习惯不同对机器人的使用方式会不同。有些人喜好远距离与人进行交流，有些人喜好近距离与人交流。因此本研究考虑到了语音识别距离是否对机器人的语音控制造成影响。在第二组实验中对语音识别距离进行了分析与论证。

第二组实验中仍然选择 6 位志愿者进行试验，测试次数为 10 次。6 位志愿者与

语音控制多媒体播放机器人的距离各不相同，A、B、C、D、E、F 与机器人的距离分别为，50cm、80cm、100cm、150cm、200cm 和 300cm，研究者对识别距离这个指标进行测试，其结果如表 5-2 所示。

表 5-2 不同识别距离对对语音控制多媒体播放机器人语音控制的影响

| 名称 | 性别 | 年龄 | 家乡 | 测试距离 | 测试次数 | 唤醒次数 | 指令执行次数 |
|----|----|----|-----|-------|------|------|--------|
| A | 男 | 31 | 黑龙江 | 50cm | 10 | 10 | 10 |
| B | 男 | 25 | 湖北 | 80cm | 10 | 10 | 10 |
| C | 男 | 32 | 山东 | 100cm | 10 | 10 | 10 |
| D | 女 | 30 | 山东 | 150cm | 10 | 10 | 9 |
| E | 女 | 26 | 安徽 | 200cm | 10 | 9 | 8 |
| F | 女 | 33 | 新疆 | 300cm | 10 | 8 | 6 |

从表 5-2 中的数据可以看出，不同测试距离是对机器人的语音控制是有影响的。志愿者与机器人距离在 1m 以内时，本实验测试的成功率为 100%。1m 的距离是人们实际应用场景中最为实用的距离。当志愿者与机器人的距离大于 1m 时，测试成功率是随着距离增加而降低。在上述表格中也可以发现，当志愿者与机器人距离为 150cm 时，唤醒成功率仍然为 100%，然而指令执行的成功率是 90%。从而说明此时麦克风阵列成功采集到了志愿者的语音唤醒信号，但是在获取语音指令的效果却没有获取唤醒词效果好。从志愿者 E 和志愿者 F 的两组实验中可以看出，麦克风阵列采集唤醒词的效果优于采集语音指令的效果。

（3）环境噪声分析

由于语音控制多媒体播放机器人的使用场景一般是在家庭或者单位的环境中，在这些环境经常会存在键盘敲打声、电视播放声音以及各种人为产生的环境噪声。环境噪声一直以来都是困扰语音识别的最大因素，现在许多语音识别系统在安静的环境下语音识别率非常的高，但是在噪声环境下识别率大幅度降低。在第三组实验中，测试不同分贝的环境噪声对语音控制多媒体播放机器人的性能影响。

第三组实验中，本文选定了 4 位志愿者（A、C、D、E），其中两位为男性，两位为女性。4 位志愿者与机器人的距离为 1m，他们分别在环境噪声为 40db、50db，60db 和 70db 与机器人进行语音交互。本研究对环境噪声进行语音测试，测试结果如表 5-3 所示。

表 5-3 不同噪音对语音控制多媒体播放机器人语音控制的影响

| 名称 | 性别 | 年龄 | 家乡 | 噪音分贝 | 测试次数 | 唤醒次数 | 指令执行次数 |
|----|----|----|-----|------|------|------|--------|
| A | 男 | 31 | 黑龙江 | 40db | 10 | 10 | 10 |
| C | 男 | 32 | 山东 | 50db | 10 | 10 | 10 |
| D | 女 | 30 | 山东 | 60db | 10 | 10 | 9 |
| E | 女 | 26 | 安徽 | 70db | 10 | 8 | 6 |

从表格 5-3 中可以看出，当环境噪声在 40db 和 50db 的情况下，语音控制多媒体播放机器人测试成功率为 100%。但是随着噪声增加，机器人的唤醒率和语音控制指令执行成功率都在降低。众所周知，人们在安静的环境下噪声大小在 20-40db，人与人正常交流的情况，噪声大小在 40-60db。当人们觉得周围环境有些吵闹，环境噪声大小在 60db-70db。当环境噪声大小超过 70db，人们就感觉周围环境已经是非常吵闹了，并且会对听力神经有损害。综上所述，语音控制多媒体播放机器人在人们正常生活的环境下，唤醒率和语音指令执行成功率都是很高的，可以满足用户在家庭与单位环境下使用。

（4） 回声路径分析

在前面文章提出，声学回声是多路径的，路径的不确定性导致了回声多样性。所以本文在此部分实验中考虑一下回声路径对机器人语音控制指令的影响。

在四组实验中，本文只选出一个志愿者 A 作为语音指令执行者。本文将语音控制多媒体播放机器人分别放在两个房间里。第一个房间（R1）长大概为 10 米，宽 7 米左右。第二个房间（R2）长大概为 7 米，宽大概为 4 米左右。在这组实验中，本文将实验环境设置成四个场景。第一个场景（S1）：在 R1 房间内，志愿者 A 与机器人距离为 1m，且双方都保持静止。第二个场景（S2）：在 R1 房间内，志愿者 A 与机器人距离为 1m，且双方都保持静止，但在机器人麦克风阵列 0.1m 处放置一个小盒子。第三个场景（S3）：在 R2 房间内，志愿者 A 与机器人距离为 1m，且双方都保持静止。第四个场景（S4）：在 R2 房间内，志愿者 A 与机器人距离为 1m，且双方都保持静止，但在机器人麦克风阵列 0.1m 处放置一个与 S2 一样的小盒子。在上述四个场景中，语音多媒体播放机器人都在工作中，测试结果如表 5-4 所示。

表 5-4 不同回声路径对语音控制多媒体播放机器人语音控制的影响

| 名称 | 性别 | 年龄 | 家乡 | 场景 | 测试次数 | 唤醒次数 | 指令执行次数 |
|----|----|----|-----|----|------|------|--------|
| A | 男 | 31 | 黑龙江 | S1 | 10 | 10 | 10 |
| A | 男 | 31 | 黑龙江 | S2 | 10 | 10 | 9 |
| A | 男 | 31 | 黑龙江 | S3 | 10 | 10 | 10 |
| A | 男 | 31 | 黑龙江 | S4 | 10 | 10 | 9 |

表格 5-4 中可以看出，S1 和 S2 两个场景是都是在 R1 房间内发生的，由于同一房间天花板、墙壁都是一样的，所以回声延迟是一样的。但是 S2 场景在麦克风前有阻碍物导致了回声路径突然的变化，所以 S2 指令执行次数成功率低于 S1。从 S3 和 S4 两个场景中，也可以得出一样的结论。从而可以判断出，语音控制多媒体播放机器人的语音追踪能力还需要有待提高。

众所周知，不同的房间会导致不同回声路径的不同。在 S1 和 S3 两个场景中，语音控制多媒体播放机器人语音指令执行成功率都 100%。从中可以发现，回声路径的变化没有影响到语音指令执行成功率。

(5) 机器人指令语义分析

在人们的日常生活中，一些语音指令都是具有口语化的往往与书面语言不相符，而且每个人的说话习惯不太相同造成语音指令具有多样性。所以在本组实验中，主要对同一语义不同语音的指令进行分析与验证。

在这一组实验中，本文只选出三个志愿者 A、C、E 作为语音指令执行者，他们三个人虽然来自不同地方，但是他们的口音都非常接近普通话，只是他们的说话习惯略有不同。首先将语音控制多媒体播放机器人放在一个空旷而且安静的房间里，给出三位志愿者三组文本信息，三位志愿者用他们自己的说话习惯对语音控制多媒体播放机器人进行语音指令发送。第一组文本信息为“前进”，A 志愿者用“走”描述第一组文本信息；C 志愿者用“向前”描述第一组文本信息；E 志愿者用“go”描述第一组文本信息。第二组文本信息为“播放土豆种植教学视频”，A 志愿者用“我想看土豆是咋种的教学视频描述第二组文本信息；C 志愿者用“打开马铃薯的种植教学录像”描述第二组文本信息；E 志愿者用“播放土豆视频”描述第二组文本信息。第三组文本信息为“60 岁的老人如何在生活中预防感冒的小妙招”，A 志愿者用“如何让 60 岁老人不感冒”描述第三组文本信息；C 志愿者用“对于 60 岁老人预防感冒都有哪些小妙招”描述第三组文本信息；E 志愿者用“如何在花甲之年预防感冒”描述第

三组文本信息。三位志愿者分别按照上述语音指令描述三组文本信息，测试结果如表 5-5 所示。

表 5-5 不同指令相同语义对语音控制多媒体播放机器人语音控制的影响

| 名称 | 测试次数 | 唤醒次数 | 第一组指令执行 | 第二组指令执行 | 第三组指令执行 |
|----|------|------|---------|---------|---------|
| A | 10 | 10 | 10 | 7 | 6 |
| C | 10 | 10 | 10 | 8 | 8 |
| E | 10 | 10 | 4 | 9 | 5 |

从表格 5-5 中可以看出，在第一组文本信息指令执行结果可以发现，中文短文本的语音指令成功率是 100%的，但是在执行英文指令上成功率仅有 40%，说明中文语音指令的成功率要远远大于英文语音指令。在第二组文本信息指令执行结果中可以发现，当语音指令加入少量方言会影响指令执行的成功率，并且关键词对语音指令执行起着至关重要的作用。在三组文本信息指令执行结果可以发现，同义词对语音指令的执行会有一定的影响，语序的变换对语音指令执行成功率的影响不大。

从三组的语义指令分析实验可以发现，当原始文本信息较短时，同语义的不同词语的语音指令执行成功率是较高的，方言与外语对语音指令执行的成功率是有较大的影响，但是语序结构的变换对语音指令执行的成功率影响较少。

(6) 机器人动态分析

在上述五组实验，本文的机器人与志愿者都是保持静态。但是在日常生活中，用户与机器人并不是始终保持静止状态。所以在第六组实验里，本文主要对机器人的动态进行分析。

在第五组实验中，本文只选出一个志愿者 A 作为语音指令执行者。A 与机器人距离为 1m。本文将本组实验分成三种情况：（1）志愿者 A 保持静止；（2）志愿者 A 快速走移动；（3）志愿者 A 小跑移动 。测试结果如表 5-6 所示。

从上述表格中可以看出，志愿者的移动是会影响语音控制多媒体播放机器人的指令执行成功率，且移动速度越快对机器人指令执行影响就越大。

从上面六组实验中，人们可以发现影响语音控制多媒体播放机器人的语音指令执行主要有语音识别距离因素、环境噪声因素和回声路径因素。然后在人们日常生活中，语音识别距离因素和环境噪声因素是最主要的两大因素。

表 5-6 用户动静状态对语音控制多媒体播放机器人语音控制的影响

| 名称 | 性别 | 年龄 | 家乡 | 场景 | 测试次数 | 唤醒次数 | 指令执行次数 |
|----|----|----|-----|----|------|------|--------|
| A | 男 | 31 | 黑龙江 | 静止 | 10 | 10 | 10 |
| A | 男 | 31 | 黑龙江 | 快走 | 10 | 10 | 10 |
| A | 男 | 31 | 黑龙江 | 小跑 | 10 | 10 | 8 |

5.4.2 实验评估

由于本文的设计目的是用户可以与语音控制多媒体播放机器人更好的进行语音交互。因此，本实验的评估准则是机器人的语音指令执行成功率和用户满意度。

当 6 位志愿者完成上述五组语音控制多媒体播放机器人交互实验后，本文统计了机器人语音指令执行成功率。并且让志愿者对机器人进行评价打分。分值在 0-100 分，从机器人外观、反应速度及主观感知三大方面进行打分，分值越高评价越高。

(1) 客观评估

图 5-9 显示了语音控制多媒体播放机器人的整体实验评估结果，从图中大家可以看出语音识别距离和环境噪音对机器人的影响最大。而且这两个因素也是人们日常生活和工作中经常遇到的问题。由于志愿者基本上讲的都是标准普通话，个性化语音不会对机器人语音控制产生干扰与影响。语音控制多媒体播放机器人的自身背景声音会对机器人语音控制产生一部分影响，本文通过语音唤醒功能对志愿者进行定位，从而对该方向进行语音增强，削弱其他方向的语音信号。

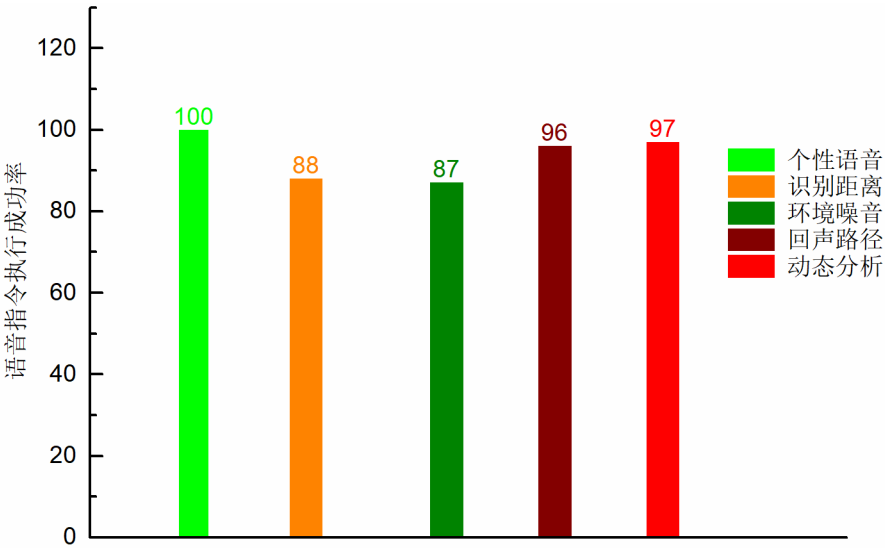


图 5-9 机器人原型系统客观评估

(2) 主观评估

在完成五组实验之后，本文对志愿者进行了用户体验调查。表 5-7 是 6 位志愿者的调查结果：

表 5-7 志愿者体验调查表

| 名称 | 性别 | 年龄 | 家乡 | 外观 | 反应速度 | 主观感知 |
|----|----|----|-----|----|------|------|
| A | 男 | 31 | 黑龙江 | 90 | 90 | 90 |
| B | 男 | 25 | 湖北 | 75 | 80 | 70 |
| C | 男 | 32 | 山东 | 80 | 80 | 85 |
| D | 女 | 30 | 山东 | 85 | 80 | 80 |
| E | 女 | 26 | 安徽 | 80 | 80 | 85 |
| F | 女 | 33 | 新疆 | 75 | 80 | 60 |

因为主观评估是志愿者对机器人的个人评价，每个人都有自己不同意见与观点。表 5-7 中显示出 6 位志愿者对机器人打分评估，从表中可以看出所有志愿者对反应速度是最为满意的，整体主观感知是最差的。因为样本数较小，本文在所有评价中去掉一个最低分数和一个最高分数然后取平均值，其结果如图 5-10 所示。从图中可以看出，三个评价分值是一样的都为 80 分。

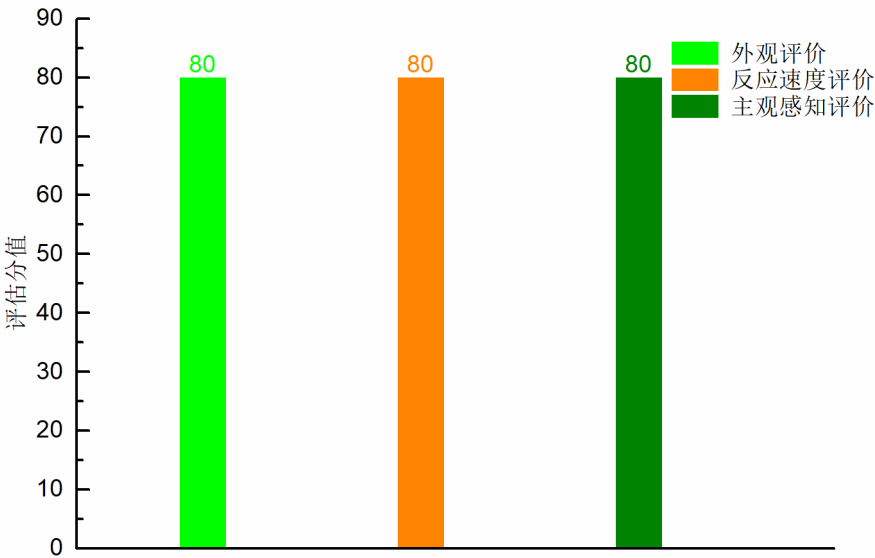


图 5-10 机器人原型系统主观评估

5.5 本章小结

本章内容主要对语音控制多媒体播放机器人的功能进行了介绍。也将机器人的主板、语音采集系统、显示设备、行走机构和通讯协议进行了详细的说明。并对机器人的原型系统进行了实验验证。实验分别从个性化语音、语音识别距离、环境噪声、回声路径和机器人动态进行了论证和分析。最后通过六组实验给出了客观评估和主观评估，验证了语音控制多媒体播放机器人原型系统的稳定性和优异性。

第六章 结论与展望

6.1 结论

语音控制技术随着社会发展已经进入人们生活中的各个角落。人与机器人的语音交互逐渐代替了传统的交互手段，它是一种更直接、更简便和更容易操作的方式。虽然近些年语音控制技术发展迅速，相关衍生产品也不停更新换代。但是环境噪声影响、回声干扰和语音文本匹配等一些因素阻碍了语音控制技术更广泛的应用在生活场景中。本文设计了语音控制多媒体播放机器人原型系统，并对它进行整体实验验证，主要结论如下：

(1) 针对环境噪声的影响，提出了改进的语音降噪算法。对语音信号特性和噪声信号种类进行特征分析，对传统语音降噪算法进行概述与总结。基于硬阈值函数提出了一个改进的阈值函数小波包语音降噪算法，对该语音降噪算法进行了语音仿真实验，并基于三个评价标准进行对比试验。从实验结果中可以看出本文提出的改进算法比传统语音降噪算法在语音降噪方面性能更加优异。

(2) 克服语音控制多媒体播放机器人的回声消除问题。针对回声干扰难题，本文对回声特性进行讨论与概述，并重点对声学回声消除进行深入研究。对传统的自适应滤波回声消除算法进行了推导与总结，提出了改进的可变步长子带自适应滤波算法，并对改进的算法进行了实验仿真。在对比仿真实验中，本文提出的改进算法在稳态误差和 ERLE 两个评价标准方面与传统的 NLMS 算法、可变步长算法和 SM-NLMS 算法进行了对比实验。实验结果表明本文提出的改进算法具有更快的收敛速度和更低的稳态误差，且回声消除效果更加优异。

(3) 基于语音文本匹配技术实现语音指令执行。由于文本匹配的多元性、结构性和层次性给文本匹配技术给文本匹配带来巨大挑战。本文对传统文本匹配模型的关键技术和文本匹配的评价指标进行了研究与概述。通过对传统文本匹配学习模型的了解，构建了基于局部匹配的数字化文本信息检索模型，弥补了传统方法的缺点。最后通过对比实验表明，本文提出的方法在检索数字农业文本信息的过程中，具有较高的召回率和检索率，实现了文本检索花费时间少和效率高的优点。

(4) 语音控制多媒体播放机器人的原型系统搭建与实验研究。针对环境噪声、回声消除和语音文本匹配关键技术问题，进行了原型系统的搭建。实现了语音控制多媒体播放机器人的语音交互、传统交互、多媒体播放和行走等基本功能。根据功能的实现，设计机器人软件系统和硬件结构。并根据上述关键问题，对语音控制多媒体播放机器人进行总体实验验证，并根据志愿者主观体验感知和实验结果对机器人进行了

主观与客观的评价。

6.2 创新点

本文提出的创新点主要有：

(1) 提出了改进的阈值函数小波包语音降噪算法，解决了传统阈值函数的缺陷，将硬阈值函数与软阈值函数进行了折衷，同时保持两种阈值函数的优点。

(2) 提出了改进的可变步长子带自适应滤波算法，该算法解决了固定步长因子不能同时满足快速收敛和低稳态误差的问题。该算法在实际应用中不受任何实际参数的影响，并且计算复杂度较低。

(3) 提出了局部匹配的数字文本信息检索方法，构造文本树和查询树生成查询中的祖先后代关系，并映射到文本中。采用矢量检索方法计算文本查询之间的相似度，相似度从大到小进行排序输出文本信息。

6.3 展望

论文对语音降噪、回声消除和文本匹配的理论 and 关键技术做了一些探索性的研究，并在最后完成了语音控制多媒体播放机器人的原型系统。但是整个原型系统还有许多不足之处，从志愿者给机器人原型系统的主观评价中可以看出，机器人原型系统仍然需要大量改进。由于本人知识水平和时间的关系，只能从后续工作对该系统进行改进，后续工作将以下几大方面展开。

(1) 在语音降噪系统中，没有考虑到更为复杂的工作环境，大部分研究只在一个相对平稳的生活环境中。

(2) 在回声消除实验中，本文是假设已知噪声环境，噪声估计方差问题没有得到更好的解决。

(3) 在语义分析实验中，面对英文单词和长句子时，语音指令执行成功率较低，达不到理想的结果。

(4) 语音控制多媒体播放机器人原型系统的仍然有许多结构需要进行改进，机器人在行走时语音控制指令执行成功率较低，大大降低了用户体验感。并且整个系统结构相对简单，各个模块之间的通信稳定性有待提高。

参考文献

- [1] Dautrich B, Rabiner L R, Martin T, et al. On the effects of varying filter bank parameters on isolated word recognition[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1983, 31(4): 793-807.
- [2] 孟欣,马建芬,张雪英.改进的参数自适应的维纳滤波语音增强算法[J].计算机工程与设计,2017,38(3):714-718.
- [3] Yang Y, Bao C. RS-CAE-Based AR-Wiener Filtering and Harmonic Recovery for Speech Enhancement[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2019, 27(11): 1752-1762.
- [4] Rahali H, Hajaiej Z, Ellouze N, et al. De-noising by Gammachirp and Wiener filter-based methods for speech enhancement[J]. International Journal of Information and Communication Technology, 2018, 13(1): 55-67.
- [5] 赵彦平.基于稀疏表示的语音增强方法研究[D]. 吉林大学, 2014.
- [6] Katzman J L, Shaham U, Cloninger A, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network[J]. BMC Medical Research Methodology, 2018, 18(1):24-24.
- [7] 卢志强.基于谱估计统计模型的语音增强算法研究[D]. 湖南大学, 2013.
- [8] Surendran S, Kumar T K. Oblique Projection and Cepstral Subtraction in Signal Subspace Speech Enhancement for Colored Noise Reduction[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2018, 26(12):2328-2340.
- [9] Id S C L *, Wang J F, Chen M H. Threshold-Based Noise Detection and Reduction for Automatic Speech Recognition System in Human-Robot Interactions[J]. Sensors, 2018, 18(7):2068-2077.
- [10] Huanhuan Z, Sheng L, Xianbo W, et al. Non-air conducting speech enhancement based on multi-band spectral subtraction[J]. Wireless Internet Technology, 2018, 15(09):10-11.
- [11] 孙杨,原猛,冯海泓.一种基于统计模型的改进谱减降噪算法[J].声学技术,2018,32(2):115-118.
- [12] 闵姝君,田岚.一种自适应变阶谱减降噪算法在电子耳蜗中的应用[J].声学学报,2011(3):82-88.
- [13] 丁武,章小兵,吴鹏.谱减法和掩蔽后处理法结合的语音增强[J].电声技术,2018,42(3):52-54.
- [14] 刘丹亭,田斌,易克初,薄涵睿.一种用于噪声环境下语音识别的多模型谱减法[C].中国电子学会信号处理分会、中国仪器仪表学会信号处理分会、中国航空学会信号与信息处理分会.第九届全国信号处理学术年会(CCSP-99)论文集.中国电子学会信号处理分会、中国仪

- 器仪表学会信号处理分会、中国航空学会信号与信息处理分会:中国电子学会信号处理分会,1999:194-197.
- [15] 金薛冬,李东新. 基于谱减法的语音信号降噪改进算法[J]. 国外电子测量技术, 2018,37(5):63-67.
- [16] 郑成诗,胡笑浒,周翊. 基于噪声谱结构特性的谱减法[J]. 声学学报, 2010, 35(2):215-222.
- [17] Lim, Jae S. Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise[J]. IEEE Trans on Acoustic Speech & Audio Processing, 1978, 26(5):471 - 472..
- [18] Khan M S, Naqvi S M, Chambers J A, et al. A new cascaded spectral subtraction approach for binaural speech dereverberation and its application in source separation[C]. international conference on acoustics, speech, and signal processing, 2013: 6566-6570.
- [19] Udrea R M, Vizireanu N D, Ciochina S. An improved spectral subtraction method for speech enhancement using a perceptual weighting filter[J]. Digital Signal Processing, 2018, 18(4):581-587.
- [20] 孙琦. 基于子空间的低计算复杂度语音增强算法研究[D], 吉林大学, 2018.
- [21] 奚吉,梁瑞宇,王国伟,等. 多通道助听器语音降噪算法研究[J]. 计算机工程与应用, 2014, 50(11):237-240.
- [22] 靳立燕,陈莉,樊泰亨,等. 基于奇异谱分析和维纳滤波的语音去噪算法[J]. 计算机应用, 2015(8):230-234.
- [23] 彭健新,唐云飞. 谱减法与维纳滤波法相结合的睡眠鼾声降噪处理[J]. 华南理工大学学报:自然科学版, 2018, 46(3):103-107.
- [24] Ngo K, Moonen M, Jensen S H, et al. A flexible Speech Distortion Weighted Multi-channel Wiener Filter for noise reduction in hearing aids[C]. international conference on acoustics, speech, and signal processing, 2018: 2528-2531.
- [25] Cornelis B, Moonen M, Wouters J. Reduced-bandwidth Multi-channel Wiener Filter based binaural noise reduction and localization cue preservation in binaural hearing aids[J]. signal processing, 2014, 99(99):1-16.
- [26] M. Dendrinos, S. Bakamidis, G. Carayannis. Speech enhancement from noise: A regenerative approach[J]. Speech Communication, 1991, 10(1):45-57.
- [27] 吴冠宁,佟国香. 基于组合算法的语音降噪研究[J]. 信息技术, 2015(10):209-213.
- [28] 肖毅,杨绍清,章新华,等. 基于滑动窗口子空间的混沌信号降噪算法[J]. 应用声学, 2007(1):23-28.
- [29] 裴俊华,贾海蓉. 基于子空间域的自适应小字典的语音增强[J]. 现代电子技术, 2019, 42(1):54-58.

- [30] Ephraim Y, Van Trees H L. A signal subspace approach for speech enhancement[J]. IEEE Transactions on Speech and Audio Processing, 1995, 3(4): 251-266.
- [31] Jabloun F, Benoît Champagne. Incorporating the human hearing properties in the signal subspace approach for speech enhancement[J]. IEEE Transactions on Speech & Audio Processing, 2018, 11(6):700-708.
- [32] Hermus K, Wambacq P, Hamme H V. A review of signal speech enhancement and its application to noise robust speech recognition[J]. EURASIP Journal on Advances in Signal Processing, 2007, 2007(1):045821.
- [33] Yu-Jing T, Li-Jian Z, Hong-Wei Z. Simulation of Wavelet Package Speech Enhancement Based on Tracing Formant[J]. Journal of Yantai University (Natural ence and Engineering Edition), 2019, 24(4):292-297.
- [34] 任永梅, 张雪英, 贾海蓉. 一种新阈值函数的小波包语音增强算法[J]. 计算机应用研究, 2013(1):120-122+143.
- [35] Chang S, Kwon Y, Yang S, et al. Speech enhancement for non-stationary noise environment by adaptive wavelet packet[C]. international conference on acoustics, speech, and signal processing, 2018: 561-564.
- [36] Sanam T F, Shahnaz C. Noisy speech enhancement based on an adaptive threshold and a modified hard thresholding function in wavelet packet domain[J]. Digital Signal Processing, 2018, 23(3):941-951.
- [37] Messaoud M A B, Bouzid A. Speech Enhancement Based on Wavelet Transform and Improved Subspace Decomposition[J]. Journal of the Audio Engineering Society, 2018, 63(12):990-1000.
- [38] 郑鲲, 孙光民, 廖晟宇,等. 一种基于 RBF 神经网络的语音去噪方法[J]. 中国科技论文, 2018, 8(4):339-341.
- [39] Hongtu Z, Yun L. Speech Denoising Based on the Wavelet Neural Network of Improved Particle Swarm Algorithm[J]. Computer Measurement & Control, 2018, 21(10):2799-2800+2807.
- [40] Shi-Xue W, Lei S, Jun D U. Application of Progressive Learning Speech Enhancement Approach to Speech Recognition[J]. Journal of Chinese Computer Systems, 2018, 39(1):1-6.
- [41] Kolbcek M, Tan Z, Jensen J, et al. Monaural Speech Enhancement Using Deep Neural Networks by Maximizing a Short-Time Objective Intelligibility Measure[C]. international conference on acoustics, speech, and signal processing, 2018: 5059-5063.
- [42] Jing H, Mengkai H U, Ziwei W. An Improved Knowledge-aided Space-time Adaptive Signal Processing Algorithm for MIMO Radar[J]. Journal of Electronics & Information Technology, 2019, 41(4):795-800.
- [43] Sen M. Kuo, WoonSeng Gan, KongAik Lee. Subband Adaptive Filtering: Theory and Implementation[M]. Wiley Publishing, 2019.

- [44] Lu L, Zhao H, He Z, et al. A novel sign adaptation scheme for convex combination of two adaptive filters[J]. Aeu-international Journal of Electronics and Communications, 2015, 69(11): 1590-1598.
- [45] Djendi M, Henni R, Djebbari M. A new adaptive solution based on joint acoustic noise and echo cancellation for hands-free systems[J]. International journal of speech technology, 2019, 22(2):407-420.
- [46] Yu Y, Zhao H. Novel sign subband adaptive filter algorithms with individual weighting factors[J]. Signal Processing, 2018, 122(5):14-23.
- [47] Yu Y, Zhao H. Proportionate NSAF algorithms with sparseness-measured for acoustic echo cancellation[J]. AEU - International Journal of Electronics and Communications, 2017, 75(1):53-62.
- [48] Widrow B, Hoff M E. Adaptive switching circuits[J]. Neurocomputing, 1960, 4(1):126-134.
- [49] 赵欣波, 杨苹. 基于 LMS 算法的自适应滤波及在回声消除中的应用[J]. 信息技术, 2006(8):28-30+33.
- [50] 王正腾, 谢维波. 改进的 NLMS 算法在回声消除系统中的应用[J]. 计算机工程与应用, 2017 53(10):107-111.
- [51] Zhang S, Zheng W X, Zhang J, et al. A family of robust m-shaped error weighted least mean square algorithms: performance analysis and echo cancellation application [J]. IEEE Access, 2017, PP (99):1-1.
- [52] Ciochin S, Paleologu C, Benesty J. An optimized NLMS algorithm for system identification [J]. Signal Processing, 2016, 118(1):115-121.
- [53] Duttweiler, D. L. Proportionate normalized least-mean-squares adaptation in echo cancelers[J]. IEEE Transactions on Speech & Audio Processing, 2018, 8(5):508-518.
- [54] 李跃明, 侯楚林. 变步长比例仿射投影算法及在回声消除中的应用[J]. 计算机工程与应用, 2012, 48(35):126-130.
- [55] Shin H C, Sayed A H, Fellow, et al. Variable Step-Size NLMS and Affine Projection Algorithms [J]. Signal Processing Letters, IEEE, 2014, 11(2):132-135.
- [56] Mayyas K, Abuseba H A. A new variable length NLMS adaptive algorithm [J]. Digital Signal Processing, 2018, 34(1):82-91.
- [57] Yu Y, Zhao H. A Novel Variable Step Size NLMS Algorithm Based on the Power Estimate of the System Noise[J]. computer science, 2015, 11(2):605-609.
- [58] Benesty J, Rey H, Vega L R, et al. A Nonparametric VSS NLMS Algorithm [J]. IEEE Signal Processing Letters, 2016, 13(10):581-584.
- [59] K. A. Lee, W. S. Gan. Improving convergence of the NLMS algorithm using constrained subband updates. IEEE Signal Process.Lett, 2014, 11(9), 736-739.
- [60] Shen Z, Yu Y, Huang T, et al. Variable step-size version of improved normalized subband

- adaptive filter[C]. chinese control conference, 2016: 4972-4975.
- [61] Shin, Kong, NamWoong, et al. Normalised subband adaptive filter with variable step size (Circuits and systems) [J]. Electronics Letters, 2018, 48(4):204-206.
- [62] Jeong J J, Koo K, Choi G T, et al. A Variable Step Size for Normalized Subband Adaptive Filters [J]. IEEE Signal Processing Letters, 2018, 19(12):906-909.
- [63] Abadi M S, Shafiee M S. A new variable step-size normalized subband adaptive filter algorithm employing dynamic selection of subband filters[C]. iranian conference on electrical engineering, 2013: 1-5
- [64] Ohno S, Sakai H. On delayless subband adaptive filtering by subband/fullband transforms[J]. IEEE Signal Processing Letters, 1999, 6(9): 236-239.
- [65] Shin J W, Yoo J W, Park P G. Adaptive regularisation for normalised subband adaptive filter: mean-square performance analysis approach [J]. IET Signal Processing, 2018, 12(9):1146-1153.
- [66] Choi Y S. Noise-robust normalised subband adaptive filtering[J]. Electronics Letters, 2012, 48(8):432-434.
- [67] Ni J. Improved normalised subband adaptive filter[J]. Electronics Letters, 2012, 48(6):320-321.
- [68] Abadi M S, Shafiee M S. A new variable step-size normalized subband adaptive filter algorithm employing dynamic selection of subband filters[C]. iranian conference on electrical engineering, 2013: 1-5.
- [69] Ni J, Li F. A Variable Step-Size Matrix Normalized Subband Adaptive Filter[J]. Audio Speech & Language Processing IEEE Transactions on, 2018, 18(6): P.1290-1299.
- [70] Yu Y, Zhao H, Chen B. Set-membership improved normalized subband adaptive filter algorithms for acoustic echo cancellation[J]. iet signal processing, 2015, 12(1):42-50.
- [71] Ni J, Li F. Adaptive combination of subband adaptive filters for acoustic echo cancellation[J]. Consumer Electronics, IEEE Transactions on, 2010, 56(3): 1549-1555.
- [72] Azpicueta-Ruiz L A, Zeller M, Figueiras-Vidal A R, et al. Adaptive Combination of Volterra Kernels and Its Application to Nonlinear Acoustic Echo Cancellation[J]. IEEE Transactions on Audio Speech & Language Processing, 2017, 19(1):97-110.
- [73] Arevalo L, Apolinario J A, De Campos M L, et al. Convex Combination of Three Affine Projections Adaptive Filters[C]. international symposium on wireless communication systems, 2013: 1-5.
- [74] Arenas-Garcia J, Figueiras-Vidal A R, Sayed A H. Mean-square performance of a convex combination of two adaptive filters[J]. IEEE Transactions on Signal Processing, 2006, 54(3):1078-1090.
- [75] Arenas-Garcia, Jeronimo, Azpicueta-Ruiz, Luis A, Silva, Magno T.M. Combinations of Adaptive Filters: Performance and convergence properties[J]. IEEE Signal Processing Magazine,

2016, 33(1):120-140.

[76] Das B K, Chakraborty M. Sparse Adaptive Filtering by an Adaptive Convex Combination of the LMS and the ZA-LMS Algorithms[J]. Circuits and Systems I: Regular Papers, IEEE Transactions on, 2014, 61(5):1499-1507.

[77] Seo, Hyeji, Lee, et al. Integrated acoustic echo and background noise suppression based on stacked deep neural networks[J]. APPLIED ACOUSTICS, 2018, 1(1): 194-201.

[78] Muller M, Jansky J, Bohac M et al. Linear acoustic echo cancellation using deep neural networks and convex reconstruction of incomplete transfer function[C]. IEEE International Workshop of Electronics. IEEE, 2017, 1-5.

[79] Abadi M S, Mesgarani H, Khademiyan S M, et al. The wavelet transform-domain LMS adaptive filter employing dynamic selection of subband-coefficients[J]. Digital Signal Processing, 2017, 61(5): 94-105.

[80] 姚力, 刘强. VoIP 中一种基于 WebRTC 的回声消除改进算法[J]. 计算机科学, 2017(S1):319-321+328.

[81] 刘德全. 基于麦克风阵列的声学回声消除算法研究及软件仿真[D]. 电子科技大学, 2012.

[82] 冯江浩. 基于自适应滤波的声学回声消除算法研究[D]. 太原理工大学, 2018.

[83] Picone J, Goudie-Marshall K M, Doddington G R, et al. Automatic text alignment for speech system evaluation[J]. IEEE Transactions on Acoustics Speech & Signal Processing, 1986, 34(4):780-784.

[84] Toledano D T, Gomez L A, Grande L V. Automatic phonetic segmentation[J]. IEEE Transactions on Speech & Audio Processing, 2003, 11(6): p.617-625.

[85] 陶冶. 文本语音匹配的研究和应用[D]. 山东大学, 2019.

[86] Li H, Xu J. Semantic Matching in Search[J]. Foundations and Trends in Information Retrieval, 2017, 7(5):343-469.

[87] Brown P F, Pietra V J, Pietra S A, et al. The mathematics of statistical machine translation: parameter estimation[J]. Computational Linguistics, 1993, 19(2): 263-311.

[88] Xue X, Jeon J, Croft W B, et al. Retrieval models for question and answer archives[C]. international acm sigir conference on research and development in information retrieval, 2008: 475-482.

[89] Ekaterina Pronoza, Elena Yagunova, Anton Pronoza. Construction of a Russian Paraphrase Corpus: Unsupervised Paraphrase Extraction[J]. IEEE Transactions on Speech & Audio Processing, 2016, 19(2): 263-311.

[90] 李宏广. 基于深度神经网络的文本匹配算法研究[D]. 中国科学技术大学, 2019.

[91] 彭晓娅, 周栋. 跨语言词向量研究综述[J]. 中文信息学报, 2020, 34(2):1-15+26.

[92] 严小龙, 庞晓琼, 任孟琦. 支持动态更新的多关键词密文排序检索[J]. 计算机工程与设

- 计, 2018, 39(4):901-906.
- [93] 邓晓衡, 杨子荣, 关培源. 一种基于词义和词频的向量空间模型改进方法[J]. 计算机应用研究, 2019, 036(5):1390-1395.
- [94] Yun W, An W X, Jindan Z, et al. Combining Vector Space Features and Convolution Neural Network for Text Sentiment Analysis[C]. complex, intelligent and software intensive systems, 2018: 780-790.
- [95] Takama Y, Ishibashi T. M2VSM: Extension of vector space model by introducing Meta keyword[C]. world automation congress, 2008: 1-6.
- [96] 翟亚利. 非负矩阵分解及其在中文文本挖掘中的应用[D]. 国防科学技术大学, 2007.
- [97] 景永霞, 王治和, 苟和平. 基于矩阵奇异值分解的文本分类算法研究[J]. 西北师范大学学报(自然科学版), 2018, 054(3):51-56.
- [98] 姚伟娜. 基于深度哈希算法的图像-文本跨模态检索研究[D]. 北京交通大学, 2019.
- [99] Yu B, Xu Z B, Li C H. Latent semantic analysis for text categorization using neural network[J]. Knowledge Based Systems, 2018, 21(8):900-904.
- [100] Qiang W, Xiaolong W, Yi G, et al. A study of semi-discrete matrix decomposition for LSI in automated text categorization[C]. international joint conference on natural language processing, 2004: 606-615.
- [101] 张浩, 钟敏. 基于 Sentence-LDA 主题模型的短文本分类[J]. 计算机与现代化, 2019, 283(3):106-110.
- [102] 袁扬, 李晓, 杨雅婷. 基于 LDA 主题模型的维吾尔语无监督词义消歧[J]. 厦门大学学报:自然科学版, 2020(2):198-205.
- [103] 戚后林, 顾磊. 概率潜在语义分析的 KNN 文本分类算法[J]. 计算机技术与发展, 2017, 27(7):57-61.
- [104] 张倩. 用于网络评论文本挖掘的主题模型研究[D]. 北京交通大学, 2014.
- [105] 王云云, 张云华. 基于双词语义增强的 BTM 主题模型研究[J]. 软件工程, 2020(4):1-6.
- [106] 郑潮阳. 基于深度学习的文本多标签预测及问答匹配方法研究[D]. 辽宁大学, 2019.
- [107] 李雨亭. 基于深度学习的垃圾邮件文本分类方法[D]. 中北大学, 2018.
- [108] 任海平, 李伟忠, 姚诚伟. 基于深度学习的智能写稿辅助服务设计[C]. 新华通讯社重庆分社、重庆日报报业集团、重庆广播电视集团(总台)、中国新闻技术工作者联合会. 中国新闻技术工作者联合会 2017 年学术年会论文集(学术论文篇). 新华通讯社重庆分社、重庆日报报业集团、重庆广播电视集团(总台)、中国新闻技术工作者联合会:中国新闻技术工作者联合会, 2017:40-47.
- [109] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436.
- [110] T. Mikolov. Distributed Representations of Words and Phrases and their

- Compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26(1):3111-3119.
- [111] Mikolov T, Chen K, Corrado G S, et al. Efficient Estimation of Word Representations in Vector Space[C]. international conference on learning representations, 2013.
- [112] Huang P, He X, Gao J, et al. Learning deep structured semantic models for web search using clickthrough data[C]. conference on information and knowledge management, 2013: 2333-2338.
- [113] SALTON, G. A Vector space model for automatic indexing[J]. Communications of the Acm, 1975, 18(11):613-620.
- [114] 马玉玲. 一种新的基于向量空间的 XML 文档相似性度量方法及搜索技术[D]. 山东大学, 2007.
- [115] 李海蓉. 基于概念向量空间的文档语义分类模型研究 [J]. 图书情报工作, 2017(24):28+108-113.
- [116] 张佩云, 陈传明, 黄波. 基于子树匹配的文本相似度算法[J]. 模式识别与人工智能, 2014(3):36-44.
- [117] Sneiders E. Text Retrieval by Term Co-occurrences in a Query-based Vector Space[C]. international conference on computational linguistics, 2016: 2356-2365.
- [118] Watters C R. Logic framework for information retrieval[J]. Journal of the American Society for Information Science, 1989, 40(5):311-324.
- [119] Soyer H, Topic G, Stenetorp P, et al. CroVeWA: Crosslingual Vector-Based Writing Assistance[C]. north american chapter of the association for computational linguistics, 2015: 91-95.
- [120] Di A. Multiple source location--A matrix decomposition approach[J]. IEEE Transactions on Acoustics Speech and Signal Processing, 1985, 33(5):1086-1091.
- [121] Hofmann T. Probabilistic latent semantic analysis[C]. uncertainty in artificial intelligence, 1999: 289-296.
- [122] Brants T, Chen F R, Tsochantaridis I, et al. Topic-based document segmentation with probabilistic latent semantic analysis[C]. conference on information and knowledge management, 2002: 211-218.
- [123] Kong S, Lee L. Improved Spoken Document Summarization Using Probabilistic Latent Semantic Analysis (PLSA)[C]. international conference on acoustics, speech, and signal processing, 2006: 941-944.
- [124] Mikolov T, Chen K, Corrado G S, et al. Efficient Estimation of Word Representations in Vector Space[C]. international conference on learning representations, 2013.
- [125] Levy O, Goldberg Y. Neural word embedding as implicit matrix factorization[J]. Advances in neural information processing systems, 2014, 3(1):2177-2185.
- [126] Mikolov T. Distributed Representations of Words and Phrases and their Compositionality[J]. Advances in Neural Information Processing Systems, 2018, 26(1):3111-3119.

- [127] Conneau A, Kiela D, Schwenk H, et al. Supervised Learning of Universal Sentence Representations from Natural Language Inference Data[J], 2017, 3(1):2177-2185.
- [128] Subramanian S, Trischler A, Bengio Y, et al. Learning General Purpose Distributed Sentence Representations via Large Scale Multi-task Learning[J]. arXiv: Computation and Language, 2018, 3(1):2177-2185.
- [129] Mccann B, Bradbury J, Xiong C, et al. Learned in translation: contextualized word vectors[C]. neural information processing systems, 2017: 6297-6308.
- [130] Li T, Xu Z, Tang J, et al. Model-Free Control for Distributed Stream Data Processing using Deep Reinforcement Learning[J]. Proceedings of the VLDB Endowment, 2018, 11(6): 705-718.
- [131] Devlin J, Chang M, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. arXiv: Computation and Language, 2018, 3(1):2177-2185.
- [132] Zhang T, Kishore V, Wu F, et al. BERTScore: Evaluating Text Generation with BERT[J]. arXiv: Computation and Language, 2019, 3(1):2177-2185.
- [133] Wu X, Lv S, Zang L, et al. Conditional BERT Contextual Augmentation[C]. international conference on computational science, 2018: 84-95.
- [134] Cui B, Li Y, Chen M, et al. Fine-tune BERT with Sparse Self-Attention Mechanism[C]. international joint conference on natural language processing, 2019: 3546-3551.
- [135] Xenouelas S, Malakasiotis P, Apidianaki M, et al. SumQE: a BERT-based Summary Quality Estimation Model[J]. arXiv: Computation and Language, 2019, 3(1):2177-2185.
- [136] Wei R, Huang, Gao Y. Sharing Pre-trained BERT Decoder for a Hybrid Summarization[M] Chinese Computational Linguistics. 2019, 3(1):2177-2185.
- [137] Chen Y, Gan Z, Cheng Y, et al. Distilling the Knowledge of BERT for Text Generation[J]. arXiv: Computation and Language, 2019, 4(2):1177-1185.
- [138] 王玥. 说话人识别中语音特征参数提取方法的研究[D]. 吉林大学, 2009.
- [139] 郑君. 基于嵌入式系统超声波探伤的研究[D]. 北京交通大学, 2008.
- [140] 王娜. 基于小波变换与约束方差噪声谱估计的语音增强算法研究[D]. 燕山大学, 2011.
- [141] 王涛. 语音降噪处理技术的研究[D]. 兰州交通大学, 2018.
- [142] Huang P S, Kim M, Hasegawa-Johnson M, et al. Joint Optimization of Masks and Deep Recurrent Neural Networks for, Monaural Source Separation[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2015, 23(12):2136-2147.
- [143] Williamson D S, Wang Y, Wang D L. Reconstruction techniques for improving the perceptual quality of binary masked speech[J]. The Journal of the Acoustical Society of America, 2014, 136(2):892-902.
- [144] Benesty J, Jensen J R, Christensen M G, et al. Speech Enhancement: A Signal Subspace Perspective[J]. Speech Communication, 2018, 7(1):41-53.
- [145] Dendrinis M, Bakamidis S, Carayannis G. Speech enhancement from noise: A regenerative

- approach[J]. Speech Communication, 1991, 10(1):45-57.
- [146] Ephraim Y, Van Trees H L. A signal subspace approach for speech enhancement[J]. IEEE Transactions on Speech and Audio Processing, 1995, 3(4): 251-266.
- [147] Hu Y, Loizou P C. A generalized subspace approach for enhancing speech corrupted by colored noise[J]. IEEE Transactions on Speech and Audio Processing, 2003, 11(4):334-341.
- [148] Sun J, Zhang J, Small M. Extension of the local subspace method to enhancement of speech with colored noise[J]. Signal process, 2018, 88(7): 1881-1888.
- [149] Hinton, G, E, etc. Reducing the Dimensionality of Data with Neural Networks.[J]. Science, 2018, 313(5786): 504-507.
- [150] Ciresan D C, Giusti A, Gambardella L M, et al. Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks[C]. medical image computing and computer assisted intervention, 2018: 411-418.
- [151] Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning[C]. international conference on machine learning, 2018: 160-167.
- [152] Hinton G E, Deng L, Yu D, et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups[J]. IEEE Signal Processing Magazine, 2018, 29(6): 82-97.
- [153] Mobahi H, Collobert R, Weston J, et al. Deep learning from temporal coherence in video[C]. international conference on machine learning, 2019: 737-744.
- [154] Xu Y, Du J, Dai L, et al. A regression approach to speech enhancement based on deep neural networks[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2015, 23(1): 7-19.
- [155] Han K, Wang Y, Wang D, et al. Learning spectral mapping for speech dereverberation and denoising[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2015, 23(6): 982-992.
- [156] Sun L, Du J, Dai L R, et al. Multiple-target deep learning for LSTM-RNN based speech enhancement[C] 2017 Hands-free Speech Communications and Microphone Arrays (HSCMA). IEEE, 2017.
- [157] Wang Y, Narayanan A, Wang D, et al. On training targets for supervised speech separation[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2018, 22(12): 1849-1858.
- [158] Zhang X, Wang D. A deep ensemble learning method for monaural speech separation[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2016, 24(5): 967-977.
- [159] Delfarah M, Wang D. Features for Masking-Based Monaural Speech Separation in Reverberant Conditions[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2017, 25(5): 1085-1094.

- [160] 吴小珊. 变速器声品质室内评价方法研究[D]. 重庆理工大学, 2015.
- [161] 刘宁宁. 车内噪声综合烦躁度主观评价方法研究[D]. 上海工程技术大学, 2014.
- [162] 王娜, 陈克安, 黄凰. 水下噪声听觉属性的主观评价与分析[J]. 物理学报, 2009(10):660-668.
- [163] 毛东兴, 俞悟周, 王佐民. 声品质成对比较主观评价的数据检验及判据[J]. 声学学报, 2005(5):86-90.
- [164] De Courville M, Duhamel P. Adaptive filtering in subbands using a weighted criterion[J]. *IEEE Transactions on Signal Processing*, 2018, 46(9): 2359-2371.
- [165] Nguyen V D, Huynh V N, Sriboonchitta S. Integrating Community Context Information Into a Reliably Weighted Collaborative Filtering System Using Soft Ratings[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, 50(4):1318-1330.
- [166] Lee K A, Gan W S. Improving convergence of the NLMS algorithm using constrained subband updates[J]. *IEEE Signal Processing Letters*, 2004, 11(9):736-739.
- [167] Aali S R, Besmi M R, Kazemi M H. Smart VRP-NLMS algorithm for estimation of power system frequency[J]. *Compel*, 2019, 38(1):362-381.
- [168] Sandeep Pradham S, Reddy V U. A new approach to subband adaptive filtering[J]. *Signal Processing IEEE Transactions on*, 1999, 47(3):655-664.
- [169] Meng Y, Zha J, Liu Y. Intensifying the SNR of BOTDA using adaptive constrained least squares filtering[J]. *Optics Communications*, 2019, 437(1):219-225.
- [170] Abadi M S E, Shafiee M S, Zalaghi M. A low computational complexity normalized subband adaptive filter algorithm employing signed regressor of input signal[J]. *Eurasip Journal on Advances in Signal Processing*, 2018, 2018(1):21.
- [171] Mayyas K, Aboulnasr T. A fast exact weighted subband adaptive algorithm and its application to mono and stereo acoustic echo cancellation[J]. *Journal of the Franklin Institute*, 2005, 342(3):235-253.
- [172] Hirayama N, Sakai H, Miyagi S. Delayless subband adaptive filtering using the Hadamard transform[J]. *Signal Processing IEEE Transactions on*, 2019, 47(6):1731-1734.
- [173] Morgan D R, Thi J C. A delayless subband adaptive filter architecture[J]. *Signal Processing IEEE Transactions on*, 2017, 43(8):1819-1830.
- [174] Ohno S, Sakai H. On delayless subband adaptive filtering by subband/fullband transforms[J]. *IEEE Signal Processing Letters*, 2019, 6(9):236-239. S.
- [175] Jiang T, Liang R Y, Wang Q Y, Zou C R, Li C. An improved practical state-space FDAF with fast recovery of abrupt echo-path changes. *IEEE Access*, 2019; 7(1): 61353-61362.
- [176] Wu Z, Gao K, Wang Z L, et al. Linear information retrieval method in X-ray grating-based phase contrast imaging and its interchangeability with tomographic reconstruction[J]. *Journal of Applied Physics*, 2017, 121(21):213102.1-213102.8.

- [177] Yang Y G, Sun S J, Wang Y. Quantum Oblivious Transfer Based on a Quantum Symmetrically Private Information Retrieval Protocol[J]. International Journal of Theoretical Physics, 2015, 54(3):910-916.
- [178] Vladimir, Rocha, Fabio, et al. A hybrid cloud-P2P architecture for multimedia information retrieval on VoD services[J]. Computing, 2016, 98(1): 73-92.
- [179] Junnila V, Laihonon T. Information Retrieval With Varying Number of Input Clues[J]. IEEE Transactions on Information Theory, 2015, 62(2):1-1.
- [180] Cui L, Shi J, Wang Y, et al. Retrieval of contaminated information using random lasers[J]. Applied Physics Letters, 2015, 106(20):201101.1-201101.4.
- [181] M Besson, M KUTAS, C VANPETTEN. EFFECT OF SEMANTIC EXPECTANCY UPON INFORMATION-RETRIEVAL[J]. Psychophysiology, 2016, 23(4):425-426.
- [182] Khennak I, Drias H. An accelerated PSO for query expansion in web information retrieval: application to medical dataset[J]. Applied Intelligence, 2017, 47(3): 793-808.
- [183] 郭涛, 曲宝胜, 郭勇. 自然语言处理中的模型[J]. 电脑学习, 2011, 1(2):113-116.
- [184] Das D, Smith N A. Paraphrase Identification as Probabilistic Quasi-Synchronous Recognition[C]. international joint conference on natural language processing, 2019: 468-476.
- [185] Surdeanu M, Ciaramita M, Zaragoza H. Learning to Rank Answers to Non-Factoid Questions from Web Collections[J]. Computational Lingus, 2017, 37(2): 351-383.
- [186] Robertson S, Zaragoza H. The Probabilistic Relevance Framework: BM25 and Beyond[J]. Foundations & Trends in Information Retrieval, 2019, 3(4):333-389.
- [187] Shi Z. Overview of Question-Answering[J]. Journal of Chinese Information Processing, 2018, 16(6):47-53.
- [188] Bo X. Research on Question Answering & Evaluation: A Survey[J]. Journal of Chinese information processing, 2017, 19(3):2-14.
- [189] Xiang H, Gu J. Research on Question Answering System Based on Bi-LSTM and Self-attention Mechanism[C] 2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA). IEEE, 2020, 37(2): 351-383.
- [190] Peijie H, Qiang H, Xiupeng W U, et al. Question Understanding by Combining Grammar and Semantic for Chinese Dialogue System[J]. Journal of Chinese Information Processing, 2014, 28(6):70-78.
- [191] Yu K, Chen L, Chen B, et al. Cognitive Technology in Task-Oriented Dialogue Systems: Concepts, Advances and Future[J]. Chinese Journal of Computers, 2017, 38(12): 2333-2348.
- [192] Zhi-Qiong B U, Bo-Jin Z, Computer S O. Ad hoc information retrieval method based on LDA[J]. application research of computers, 2017, 32(5): 1369-1372.
- [193] Ganguly D, Roy D, Mitra M, et al. Word Embedding based Generalized Language Model for Information Retrieval[C]. international acm sigir conference on research and development in

- information retrieval, 2017: 795-798.
- [194] Hao X, Zhang G, Ma S. Deep Learning[J]. International Journal of Semantic Computing, 2018, 10(3):417-439.
- [195] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[J]. Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2018, 1(1):1-9.
- [196] Mikolov T, Chen K, Corrado G S, et al. Efficient Estimation of Word Representations in Vector Space[C]. international conference on learning representations, 2013, 795-798.
- [197] Wijerathna K A S A N , Sepalitha R , Indika T , et al. Recognition and translation of Ancient Brahmi Letters using deep learning and NLP[C]. 2019 International Conference on Advancements in Computing (ICAC). IEEE, 2020, 95-98.
- [198] Wang X, Kou L, Sugumaran V, et al. Emotion Correlation Mining Through Deep Learning Models on Natural Language Text[J]. IEEE Transactions on Cybernetics, 2020, 1(99):1-14.
- [199] Zhang Y, Tiryaki F, Jiang M, et al. Parsing clinical text using the state-of-the-art deep learning based parsers: a systematic comparison[J]. BMC Medical Informatics and Decision Making, 2019, 19(3): 51-58.
- [200] Zhang Y, Tiryaki F, Jiang M, et al. Parsing clinical text using the state-of-the-art deep learning based parsers: a systematic comparison[J]. BMC Medical Informatics and Decision Making, 2019, 19(3): 51-58.
- [201] Lee K, Kim B, Choi Y, et al. Deep learning of mutation-gene-drug relations from the literature[J]. BMC Bioinformatics, 2018, 19(1):21.
- [202] Trivedi H, Panahiazar M, Liang A S, et al. Large Scale Semi-Automated Labeling of Routine Free-Text Clinical Records for Deep Learning[J]. Journal of Digital Imaging, 2019, 32(1): 30-37.

致 谢

博士的时光很快就过去了，在这些年里痛苦过也开心过，痛苦的时候想放弃，开心的时候畅想未来。到了这个时候，突然就语塞了，自己的学业生涯马上就要结束了，没有老师与学长们的庇护，开始面临社会的大考验了。

首先我要感谢我的导师高万林教授。在过去的几年里，他向来是严格要求我，从整个博士期间的选题、研究路线、开题报告及毕业论文都是无微不至的进行指导。他不仅仅是在学术对我进行指导，在平时的生活细节也对我严格要求。在高教授的教导下，在整个博士期间基本上改掉我马虎的坏毛病。高教授还总是提醒我做人要稳扎稳打，戒骄戒躁。

感谢杨邦杰研究员，是他给我了一个走到中国农业大学的机会，谢谢他的耐心指导与关怀。

感谢实验室的王敏娟老师、陶莎老师、刘新亮老师、任延昭老师和赵龙老师，谢谢你们对我论文的指导以及生活上的关心。

感谢我们实验室的同级同学白小宁博士、何志辉博士，是你们的鼓舞与支持，我才有动力走到今天。

感谢我实验室的其他同学们，我们一直互帮互助，像是一家人共同前行，共同解决困难问题。

感谢我的家人，在我读博士期间，你们是我坚强的后盾与支柱。无论我遇到任何困难时，你们都会给我鼓励。

感谢论文评阅专家，对我的论文提出修改意见和建议。

感谢每一个帮助我的人，谢谢你们！

宋越

2020 年 04 月 15

作者简介

基本介绍（姓名、性别、出生日期、籍贯等）：

宋越 性别男 出生于 1988 年 03 月 30 日 籍贯 山东文登

教育经历：

2007-2011 就读于北京信息科技大学 学士学位

2012-2015 就读于北京信息科技大学 硕士学位

2015-至今 就读于中国农业大学 博士研究生

研究生期间发表的学术论文：

Yue Song, Yang Bangjie, Gao Wanlin, Yu Lina. Microphone array speech enhancement method based on Wavelet Packet Threshold De-noising, IAEJ 2017 Vol.26 (4): 300-306 (EI)

Yue Song, Sha Tao, Yanzhao Ren, Xinliang Liu, Wanlin Gao. An agricultural audio retrieval method based on inverted index of silenceword[C]. IOP Conference Series: Materials Science and Engineering, 2020 (EI)

Yue Song, Yanzhao Ren, Xinliang Liu, Wanlin Gao, Sha Tao, Lin Guo. A nonparametric variable step-size subband adaptive filtering algorithm for acoustic echo cancellation[J]. International Journal of Agricultural and Biological Engineering, 2020, 13(3):168-173. (SCI)

Yue Song, Wanlin Gao, Minjuan Wang. Method for retrieving digital agricultural text information based on local matching [J]. Symmetry, 2020, 12(7), 1103. (SCI)

研究生期间主持/参与的科研项目：

农村重点人群营养健康服务智能技术与装置研发及应用 课题编号：
2012BAJ18B07

文化资源建模技术研究及特色文化资源数字化 项目编号：2015BAK04B01

移动便携式农业实用技术视频放映成套装置推广与实用 项目编号：
2015GA600002

研究生期间获得的奖励和荣誉：

获得 2015-2016 博士等奖学金

获得 2018-2019 博士二等奖学金

其他成果：

专利：

高万林，**宋越**，张莉，于丽娜，陶莎，张港红. 基于模糊控制的亮度自适应方法、装置及视频放映系统，中国，发明专利，专利号：201610162652.4，已公示

高万林，**宋越**，张莉，于丽娜，陶莎，张港红. 基于模糊控制的声音自适应方法、装置及影音播放系统，中国，发明专利，专利号：201610161556.8，已公示

高万林，冯慧，**宋越**，刘云玲，张莉，李佳璇，于丽娜. 一种多媒体文件的标识方法及系统，中国，发明专利，（已授权）

高万林，**宋越**，张莉，于丽娜，陶莎，张港红. 基于模糊控制的亮度自适应装置及视频放映系统，中国，实用新型专利，专利号：201620217959.5，已授权

高万林，**宋越**，张莉，于丽娜，陶莎，张港红. 基于模糊控制的声音自适应装置及影音播放系统，中国，实用新型专利，专利号：201620217762.1，已授权