

江南大学

硕士学位论文

基于听觉掩蔽效应的数字助听器算法研究及DSP的实现

姓名： 阚仁根

申请学位级别： 硕士

专业： 检测技术与自动化装置

指导教师： 于凤芹

20080701

摘要

耳聋已成为一个世界性的问题，解决耳聋问题的最直接手段是给耳聋患者戴上助听器，提高人耳对语音感知。数字助听器最大的优点是能够植入各种语音处理算法，改变语音的分布，提高人耳感知语音的舒适度和分辨率。数字助听器的诞生给耳聋患者带来了希望。

本文重点研究了数字助听器的核心算法。将人耳的听觉掩蔽效应应用到数字助听器的核心算法中去，直接从人耳的听觉生理上出发，改善人耳对声音的听觉舒适度，并且把此算法在 DSP 上实现。

本文首先分析了数字助听器三个核心算法：语音增强，宽动态压缩和移频压缩。传统的数字助听器算法忽略了人耳的生理特性，语音听觉舒适度没提高，针对算法的缺点，提出用基于听觉掩蔽效应的语音增强算法取代原语音增强算法。这样利用人耳的生理特性降低噪声，提高语音感知的舒适度。

针对听觉掩蔽效应的语音增强算法，本文首先对噪声的估算运用了实时噪声估计，提高了噪声估计的准确性，降低了语音失真。还对原有算法中的听觉掩蔽阈值计算不准确，从而引起了谱减系数计算误差，导致语音失真过大的问题，提出了一种改进的计算方法。在原有的谱减系数计算公式中加入了一个修改参数 θ 来降低语音失真。实验仿真结果表明，本算法不但提高了语音信噪比，还改善了语音音质。

其次就传统的宽动态压缩算法，在实现频谱增益时忽略了语音共振峰，频谱增益后引起了共振峰的非线性变化，改变了共振峰的频域分布，针对此问题提出了一种基于共振峰估计的宽动态压缩算法，减少了共振峰的非线性失真，改善了听觉舒适度。

最后对数字助听器的三个核心算法进行了实验验证。利用合众达的 SEED—DEC5416 开发板，把数字助听器的三个核心算法植入其中，在此开发板上实现了三种语音处理功能。在算法的实现过程中，对其中出现的问题进行了分析研究。实践表明，该模块满足了语音的实时性并且达到了数字助听器语音处理的效果。

关键词：语音增强；听觉掩蔽；数字助听器；

Abstract

Deafness has become a global problem, the best solution for deafness means to wear hearing aids, voice through hearing aids for reduce noise, improve human ear on the voice level of comfort. Traditional analog hearing aids, only to strengthen the human ear to the voice of perception, this simple amplification can help the patients of conduction hearing loss. But the accounting for the majority of deaf patients with SNHL in patients with virtually no benefits, it will increase the ear to feel much bad. Solving the nerve deafness in patients with hearing loss, the most critical issue is to change the distribution of audio, let patients with limited hearing to feel the voice, and the core processors of digital hearing aids can add voice algorithm, and change the distribution of voice, the birth of digital hearing aids to deaf patients is a good news.

This paper focuses on the digital hearing aids of auditory masking. We use ear to the physiological characteristics-auditory masking effect of the digital hearing aids to the core algorithm, directly from the ear to the hearing of the physical to have an improvement of the human ear to the voice of comfort.

This paper firstly introduces the three core algorithm for digital hearing aids, on their strengths and weaknesses. We use the speech enhancement algorithms of masking effect to the digital hearing aids. This has the advantage of directly from the people to the characteristics of the ears, ears of the people voice perception that the maximum comfort.

Auditory masking effect for voice enhancement algorithms, the paper of the original algorithm in the auditory masking threshold value is not accurate, resulting spectrum by coefficient calculation error; voice distortion caused too great a problem, an improved method of calculation. In the original spectrum by joining the formula for calculating the coefficient of revised parameters, the algorithm can reduce the voice distortion. Simulation results show that the algorithm can improve the voice signal to noise ratio also improved voice quality.

Then, the traditional wide dynamic compression algorithm, in the realization of the gain spectrum ignoring the voice formant, after the spectrum gain resonance from the peak of the non-linear changes, change the resonance frequency of the peak. For this issue, we bring forward a new WDRC algorithm which is based on the estimated formant, reduce the resonance peak of nonlinear distortion and improved the auditory comfort.

Finally, the three core algorithm of the digital hearing aids were verified. The using of the SEED-DEC5416 development board has heard several figures for the algorithm to make a prototype of the digital hearing aids. Practice shows that the module to meet the real-time voice and reach of digital hearing aids voice processing results.

Keywords: speech enhancement; auditory masking; digital hearing aids

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含本人为获得江南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谢意。

签 名： 康仁根 日 期： 7月4日

关于论文使用授权的说明

本学位论文作者完全了解江南大学有关保留、使用学位论文的规定：江南大学有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅，可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文，并且本人电子文档的内容和纸质论文的内容相一致。

保密的学位论文在解密后也遵守此规定。

签 名： 康仁根 导师签名： 于凤芹
日 期： 2008-7-8

第一章 绪论

1.1 课题研究的背景及意义

语音是人类交流信息最自然,最有效,最方便的手段,也是人类进行思维交流的一种依托。然而随着社会的不断发展,人们对信息需求越来越高,耳聋阻碍着人类正常的信息交流,助听器越来越被人们重视。据英国医学研究理事会听力研究学会研究统计,到 2005 年全球听力损失程度在 25 分贝以上的人数就超过 5.6 亿,而中国的听力患者则有 1.2 亿之多。世界上每 10 个人就有 1 人听力损失能够通过助听器获益。

听力损失通常可分为两类:即传导型听力损失和感觉神经型听力损失(Sensorineural Hearing Loss,SNHL)。当通过患者外耳或中耳的声音传送异常时会发生传导型听力损失,而 SNHL 则发生在耳蜗中的感觉细胞或听觉系统中更高级的神经机理出现故障的场合。传导型听力损失患者可以通过传统的模拟助听器放大声音音量来改善听力,但是感觉神经型听力损失患者则无法通过简单声音放大来改善听力,加大音量只会加重患者的耳朵轰鸣感,而据统计这种纯粹的传导型听力损失患者只占听力患者总数的 5%。数字助听器的出现给感觉神经型听力损失患者带来了福音。

数字助听器能改善患者听力的核心是语音增强,移频压缩和宽动态压缩算法。通过语音处理算法提高语音的信噪比,改变语音的频谱分布,使语音能让患者感知到;自动适应环境,有效降低噪声,提高言语的清晰度;模仿人类正常耳蜗功能,使患者提高了对不同声音环境的适应性,不论声音大小都能听到,患者却没有不适感;再现自然声音,优化患者对自己声音的感受,提高声音的自然性、真实性和舒适性;自动消除反馈声,应用数字反馈抑制技术,使助听器不会出现令人烦恼的反馈声(啸叫),这对提高患者的语言分辨能力有重要的作用;引入计算机平台的概念,助听器的功能可以随助听器选配软件的升级而增强,这可以不更换助听器而保持其功能的先进性。

DSP 在医疗设备中已得到飞速的发展,大到医疗监护系统,小到数字助听器都有广泛的应用。DSP 是针对数字信号处理而设计的嵌入式微处理器,其特殊的系统结构和指令可以快速的实现各种数字信号处理算法,具有速度快、体积小、功耗低的优点。在数字助听器中以 DSP 作为核心,可以实时完成数字助听器的各种算法,减小产品体积和延长电池的续航能力。DSP 是数字助听器的最佳选择。

1.2 国内外研究的进展

1.2.1 听觉掩蔽效应的语音增强

人类听觉系统具有掩蔽效应。当两个响度不同的声音作用于人耳时,响度较高的频率成分的存在会影响到对响度较低的频率成分的感知,使其变得不易被人耳察觉,这种现象叫做掩蔽效应。听觉掩蔽模型已成功地应用于音频信号的压缩编码。

现在许多学者把听觉掩蔽效应应用于语音增强技术中来,噪声信号的存在之所以影响到人耳对纯语音信号的接收,从本质上来讲就是因为噪声信号对纯语音信号产生了掩蔽^[1]。因此,语音增强就是要减少,甚至消除噪声信号对于纯语音信号的掩蔽效应。

目前语音增强方法中谱减法是最简单,最容易实现的。其运算量不大,给语音增强系统的实时性提供了有力的保障。但是其处理后的语音信号残留了较多的音乐噪声,给人的听觉带来了疲劳^[2]。为了减少音乐噪声,美国人Berouti在传统的谱减法基础上增加了调整噪声功率谱大小的系数^[3],提高了谱减法的性能,但由于修正系数是根据经验得到的,适应性较差。P.Lockwood&J.Boudy又提出了非线性谱减法^[4],根据语音的信噪比自适应调整语音的增强系数,提高了信噪比。但是语音的信噪比高不等于语音的听觉质量就高,在1999年Nathlie Virag将人耳的听觉掩蔽特性应用到非线性谱减法中^[5],很好的解决了残留音乐噪声的问题。现在,国外很多学者针对残留音乐噪声作了很多研究,其中方法用的比较多是应用听觉掩蔽,Ching-Ta Lu把听觉掩蔽特性应用到去除有色噪声残留的音乐噪声中^[6],Zhong Lin&Rafik A.Goubra把子空间谱跟踪和听觉掩蔽特性结合起来应用到语音增强中^[7],在这些算法中音乐噪声去除的效果不错。日本学者Yoshinfumi Chisaki利用人耳的听觉特性实现了高质量的语音增强^[8],并很好的实现了处理信号的实时性。Kotta Manohar利用噪声特性在非平稳噪声环境下^[9],实现了语音增强。苏州大学的陶智和赵鹤鸣把听觉掩蔽特性和Bark子波变换结合^{[10][11]},信息工程大学的徐望王把信号子空间和听觉掩蔽效应结合^[12],这些方法把听觉掩蔽的优点和其他的去噪方法结合,改善了语音增强的质量。上海交通大学的王丽利用听觉掩蔽特性设计了感知滤波器^[13],改善了语音质量。东南大学刘海滨把听觉掩蔽效应应用到非平稳环境下的语音增强^[14],扩展了听觉掩蔽特性的应用范围。华南理工大学的谢志文对低频容易掩蔽高频,而高频较难掩蔽低频提出了新的理论解释^[15],为听觉掩蔽模型的建立开辟了新的路径。西北工业大学的同鸣又把听觉掩蔽模型应用到数字水印音频技术上^[16],听觉掩蔽特性越来越受到人们的注意。

听觉掩蔽特性应用到语音增强技术中,直接从人耳的听觉感知入手,为语音增强开辟了一个新的发展途径。

利用听觉掩蔽域值来掩蔽掉音乐噪声,这种方法可以很大程度上保证语音信号的完整性,降低了语音失真。

1.2.2 数字助听器

助听器历经电子管时代,集成电路时代,如今已到了数字时代,近几年来,数字助听器得到了实质性的发展,从常规的模拟机到编程机,进而发展到数字式。其中 DSP 技术在助听器的研究中有着举足轻重的作用。硅片技术降低了计算机芯片的功耗;用计算机分析研究外耳道和鼓膜对助听器放大的作用比传统的更加优越,这使 DSP 助听器越来越受重视。在 1996 年 Oitcon 和 Widex 就在美国市场推出了真正 DSP 助听器,目前几乎所有的生产厂商都推出了各自的全数字助听器,但各自的产品在其性能方面各有特色,各成体系,数字助听器代替模拟助听器成为了趋势。

我国的助听器生产和选配工作开始于20世纪60年代,近年来质量、种类、数量有较大快速的发展。90年代,各大国际名牌企业开始纷纷到中国投资建厂。但国内的助听事业的发展与国外相比存在一定差距,特别是在助听器的生产能力方面,国内停留在较低价位的盒式和中小功率的耳背式助听器,都属于模拟式的,产品与国际先进水平相比,在技术上存在着较大的差距,特别是在信号处理技术方面。国内厂家的产品仍采用模拟电路设计,对数字信号处理技术没有太多的研究,到目前为止没有国内自主知识产权的数字助听器产品面市,这些情况带来的直接结果就是国外的助听器产品几乎垄断了国内市场。因此,我国需继续投入人力、物力开发新型的国产助听器,特别是数字助听器。

在国外有不少公司利用TI的5000系列的DSP来开发语音系统,且应用的也很广泛,从军事通信到民用家电都有应用。用DSP研制的数字助听器已经有了很好的前景,以TMS320C54系列DSP为开发平台,开发语音系统是最高效最有性价比的,可以说5000系列的DSP是做语音处理系统的最佳选择^[17-22]。利用DSP开发的语音系统应用到数字助听器中有着广阔的市场前景。

本文所用的数字助听器语音处理算法,从人耳的听觉特性出发,利用人耳的听觉掩蔽特性,力求使人耳对经过语音处理后的语音感知最舒适,最自然。由此可见,从人类生理特性出发,听觉掩蔽效应的数字助听器研究具有深远的理论意义和广阔的市场应用前景。

1.3 本文研究的内容

本文主要研究了数字助听器的三个核心算法——语音增强算法、宽动态压缩算法和移频压缩算法。重点研究了语音增强算法,本算法从人耳的生理特性出发,利用人耳的听觉掩蔽特性,提高人耳对增强后语音的舒适感。先分析了语音和人耳的特征,针对听觉掩蔽效应,提出了一种改进的语音算法和一种噪声实时估计的语音增强算法。针对原宽动态压缩算法在实现频谱增益时引起了共振峰的非线性失真,提出了基于共振峰估计的宽动态压缩算法。在算法的实现上,本文利用了合众达的SEED-DEC5416硬件平台,编写数字助听器语音算法程序,并进行了实验验证。本文的主要工作内容如下:

(1) 对数字助听器的三个核心算法进行了研究,并分析了算法的优缺点。语音增强,宽动态压缩和移频压缩这三个算法是目前数字助听器的常用算法。语音增强主要是提高语音信噪比,改善人耳的听觉分辨率;宽动态压缩是改善语音不同频带的声强分布,使语音的声强与人耳的感知最佳匹配;移频压缩主要是对听力神经细胞损失严重的患者而设计的,它主要是改变声音的频域分布,把声音移到患者能感知的频带内。针对语音增强算法使用的自适应去噪不能提高语音听觉感知的缺点,本文提出了利用人耳的听觉感知来去除噪声,提高了语音的听觉质量。

(2) 分析了人耳听觉掩蔽的原理和特征,对原有的听觉掩蔽算法进行了改进。第一是在噪声估计阶段,本文使用了对有声段和无声段都进行噪声估计的算法,提高了噪声估计的准确性;第二是对原有的听觉掩蔽算法中计算谱减系数时出现的误差提出了一种

改进的算法，在原有的计算公式中加入了一个修改参数，提高了系数计算的准确性。

(3) 研究了共振峰对人耳听觉感知的影响，在实现宽动态压缩算法的过程中考虑到了共振峰的存在，在同一个共振峰内选用相同的频谱增益，降低了共振峰的非线性失真。在共振峰提取时选用了 LPC 法提取共振峰。实验表明此宽动态压缩算法降低了语音的失真，改善了听觉舒适度。

(4) 对数字助听器的三个核心算法进行了验证。利用合众达的 SEED-DEC5416 开发板，把数字助听器的三个语音处理算法移植到 DSP 中，实现了数字助听器语音处理的功能，并对其中出现的问题进行了分析。对输出语音进行测听，系统的实时性和功能满足数字助听器的需要。

1.4 论文结构

论文结构安排：

第一章主要介绍课题研究的背景、意义以及本文研究的主要内容。

第二章介绍了数字助听器的三个核心处理算法。主要介绍了语音增强，宽动态压缩和移频压缩这三个核心算法，并分析了他们的优缺点，对他们进行了仿真实验。

第三章研究了基于听觉掩蔽效应的语音增强算法。首先分析了噪声和语音的特征，重点研究了人耳的听觉感知中的听觉掩蔽效应，对这种听觉掩蔽特征进行了分析研究。在此基础上提出了两种改进的听觉掩蔽语音增强算法，第一是从噪声实时估计出发，第二是从阈值计算误差出发，对原有的算法进行了改进。第一种噪声实时估计，主要是降低了噪声估计误差，改善了语音的信噪比。第二是针对听觉掩蔽阈值计算不准确导致谱减系数计算不准确的问题提出了一种改进的算法，在原有的计算公式中加入了一个修改参数，减少了谱减系数计算的误差。仿真结果表明，改进的算法提高了语音的信噪比，降低了语音的失真。

第四章研究了基于共振峰估计的宽动态压缩算法。针对原始的宽动态压缩算法中没有考虑语音共振峰的存在，在实现频谱增益时引起了共振峰的非线性失真。就此问题对原有的算法提出了改进算法，在同一个共振峰内选用相同的频谱增益，降低了共振峰的非线性失真。

第五章对数字助听器的三个核心算法进行了验证。在 DSP 开发板上实现了数字助听器的语音处理算法。实践表明，在开发板上实现的语音处理算法，符合数字助听器的语音处理效果。

第六章对全文工作做了总结，提出进一步研究的方向。

第二章 数字助听器的算法

助听器可以看作一个语音回放系统。传统的模拟助听器只是把声音采集进来，通过放大器把声音放大，再通过耳麦回放出来，模拟助听器对声音的处理只是简单的放大，没有其他的声音处理方法，这种单一的处理不能满足耳病患者的需要。而数字助听器的最大优点就是能在其芯片内嵌入语音算法，经过处理后的语音能更好的适应于不同的患者。语音算法的优越性直接影响到患者的听觉感受，语音算法是数字助听器的核心^[25]。

数字助听器算法的最终目的是让人听觉最舒适和语言分辨率最高。目前针对于数字助听器的语音处理算法主要从声音的响度和频率出发，通过算法改变声音的响度和频率分布来适应患者的听觉感知。现在常用的算法是语音增强，宽动态压缩，移频压缩等^[26]，本章着重介绍以上三种语音处理算法的原理，并对三种算法分别进行了实验仿真。

2.1 语音增强算法

2.1.1 自适应语音增强

自适应信号处理是近年发展起来的信号处理领域的一个重要分支。随着其理论和技术的不断发展，它应用到越来越多的领域。自适应滤波不需要先验知识的初始条件，凭借测量信息不断实时统计来自动调节系统参数完成最佳滤波，它是通过自学来适应外部随机环境。这样自适应技术可以用来检查平稳和非平稳信号。用这项技术来检测出含噪语音中的噪声，达到降噪的效果。

自适应信号处理技术应用到噪声消除上面来了，给语音增强带来了新的发展和活力。自适应噪声消除^[27]是指从信号被噪声干扰所淹没的环境中检测和提取有用的信号，达到语音增强的目的。

利用自适应信号处理技术把含噪语音中的噪声检测出来，用含噪语音减去检测出来的噪声，达到语音增强。只要噪声检测的够准确那么相减后的增强语音效果就会更加优越。图2-1所示的是自适应噪声消除的原理图。

在图2-1所示的自适应噪声消除系统中，参考输入被一个可以通过最小均方误差算法自动调节系统冲击响应。利用误差信号经过自适应算法来调节系统的输出，使这个输出与 n_0 的相关分量相减来获得误差的最小化。

这里 s 为输入信号， n_0 为不相关噪声， n_1 为第二个传感器接收的不相关噪声， y 为输出信号，在假设 s 和 n_0 ， n_1 不相关，仅 n_0 和 n_1 相关。

输出误差为

$$\varepsilon = s + n_0 - y \quad (2.1)$$

将上式两边平方后，得到

$$\varepsilon^2 = s^2 + (n_0 - y)^2 + 2s(n_0 - y) \quad (2.2)$$

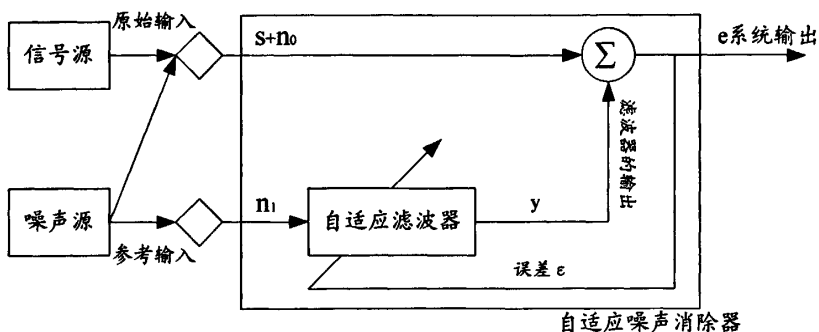


图2-1 自适应噪声消除器

Fig.2-1 Adaptive noise cancellation

两边取数学期望值，因为 s 和 n_0 ， y 无关，所以有

$$E[\varepsilon^2] = E[s^2] + E[(n_0 - y)^2] + 2E[s(n_0 - y)] = E[s^2] + E[(n_0 - y)^2] \quad (2.3)$$

调节参数使 $E[\varepsilon^2]$ 最小时，不希望信号能量 $E[s^2]$ 会受到影响，从而使最小输出能量为

$$E_{\min}[\varepsilon^2] = E[s^2] + E_{\min}[(n_0 - y)^2] \quad (2.4)$$

当改变参数使 $E[\varepsilon^2]$ 减小时， $E[(n_0 - y)^2]$ 也在同样的减小。滤波器的输出 y 是原始噪声 n_0 的最小方差估计，这也就是说当 $E[(n_0 - y)^2]$ 最小的时候， $E[(\varepsilon - s)^2]$ 也是最小的，由式2.1可以得到

$$(\varepsilon - s) = (n_0 - y) \quad (2.5)$$

对于一定的自适应滤波器调节滤波器参数，使其输出能量最小化，这等价于使输出误差 ε 为信号 s 的最小方差估计，输出误差中主要包含信号和一些噪声，由公式2.1可知，输出噪声为 $(n_0 - y)$ 。既然 $E[\varepsilon^2]$ 和 $E[(n_0 - y)^2]$ 已经最小化了，输出噪声的功率也为最小化，而且输出中的信号是不变的，所以输出功率最小化相当于信噪比为最大的情况。

以上是自适应噪声消除的基本原理。噪声消除简单的说就是从含噪语音中减去噪声得到纯净的语音，也就是语音增强里面用的最多的谱减法。由于噪声和语音的不确定性，他们混杂在一起，关键问题就是精确的得到噪声。用多通道技术可以很好的解决这个问题，实际中用的最多的就是双声道技术一个用来采集语音，另一个用来采集噪声。图2-2是一个双声道采集系统的噪声对消法的框图。

图中所示的两个采集通道，一个采集噪声，另一个采集语音。含噪语音序列为 $S(n)$ 和噪声序列 $d(n)$ 经过傅里叶变换后得到频谱分量 $S_k(\omega)$ 和 $D_k(\omega)$ ，噪声分量 $D_k(\omega)$ 经过滤波后含噪语音相减，再加上含噪语音的相位，傅里叶反变换后恢复了时域信号。采用这种噪声对消法，两个采集声道要保持一定的距离，防止两个声道采集到含噪语音。而且两路信号存在着时间差，采集到的信号所包含的噪声并不是同时的，回声及其他的衰减特性也会发生相应的变化，所以必须对噪声通道中的信号进行滤波，以便达到和含噪语音中噪声的同时性。

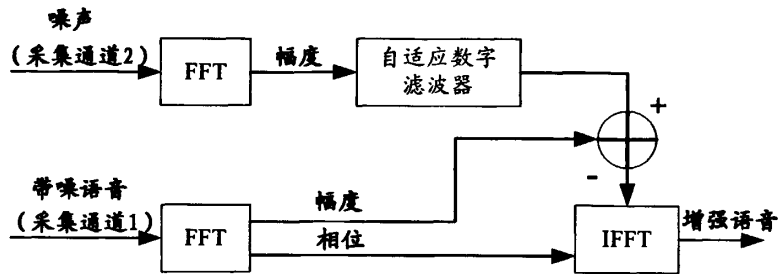


图2-2 双声道采集系统的噪声对消法

Fig.2-2 Dual-channel acquisition system noise cancellation method

在很多的实际情况下，用多个话筒来采集语音并不现实。在使用单通道采集系统时，可以利用无话期间采集到的噪声进行估计。使用这种方法时必须保证噪声是平稳的，否则会影响到噪声对消法的语音增强效果。大量的实验结果表明，在强背景噪声下，噪声对消法可以得到很好的消噪效果。如果采集的噪声足够充分，也可以直接在时域上相减。噪声对消法的一个主要的缺点就是增强后的语音会有“音乐噪声”，这种噪声是频谱相减的残留产物，由于有一定的节奏起伏性听上去有类似于音乐。由于这种特性，给这种算法在数字助听器里面的应用带来了限制。

2.1.2 听觉掩蔽效应语音增强

本文所用的语音增强算法是从人耳的生理特性出发，深入研究人耳的听觉掩蔽效应，对原有语音增强算法中残留下来的音乐噪声进行了分析，对算法的去噪性能进行了改进，提出了听觉掩蔽效应语音增强的改进算法。在本算法中计算谱减系数时加入了修改参数，提高了系统的鲁棒性，增强语音去噪的效果，很好的抑制了烦人的音乐噪声。在噪声估计方面，选用了噪声实时估计算法，一改已往用无声段噪声代替的方法。这种实时噪声估计对噪声的估计更加准确，为后面的语音去噪打下了优越的基础。

实验表明本文所用的语音增强算法对音乐噪声有很好的抑制，信噪比提高，语音音质优越，本算法在第三章中会有详细的介绍和说明。

2.2 宽动态压缩算法

在助听器算法中，听力补偿算法是其中一种重要的语音算法。听力补偿主要是对声音进行压缩放大，把正常人能听到的声音压缩放大到患者的听力动态范围里面，而且要保证声音的清晰度和听觉的舒适度，让患者能在自己的听力动态范围里面听清语音。这种听力补偿算法就是宽动态压缩算法^[28] (Wide Dynamic Range Compression, WDRC)。

动态范围指的是患者听觉范围，是从纯音听阈到不舒适阈。动态范围因人而异，两个听力图相同的患者，他们的响度增长函数是不同的，对响声的敏感程度也是不同的，对于比较严重的重振患者，他的动态范围比正常人要窄。WDRC 线路能最有效的解决他们的问题。对于声压级小声音，WDRC 将其放大；对于声压级大声音，放大量响应减少。

这样，患者听声压级小的声音没问题，听声压级大的声音又不至于太吵。

WDRC 线路将正常的声音压缩到患者的听觉动态范围内，对任何输入级均进行压缩，采用低压缩阈和低压缩比。对柔和声音和强声的放大量不同。选择 WDRC 线路要根据患者的实际情况而定，不仅要避免高强度声音失真又要对输入语音进行单位增益。

2.2.1 宽动态压缩算法的原理

宽动态压缩算法是数字助听器的主要语音信号处理方法，动态范围是指人耳对一频率声音从刚刚听见到难以忍受的声强范围。与正常的人相比较，耳病患者的听力损伤主要体现在对声音感受的范围大大的降低了，图 2-3 所示的就是听力正常人和听力不正常人的听力指标比较图， THR 为听阈， UCL 为不适域。耳病患者对可见声音动态范围缩小，最直接的体现就是在这个频率范围里面，正常人可以听到较高和较低的声音，而耳病患者则无法听到。为了让患者能听到正常人能听到的声音，同时为了满足耳病患者对声音的动态接受的需要，助听器研究人员提出了压缩放大的这个概念。

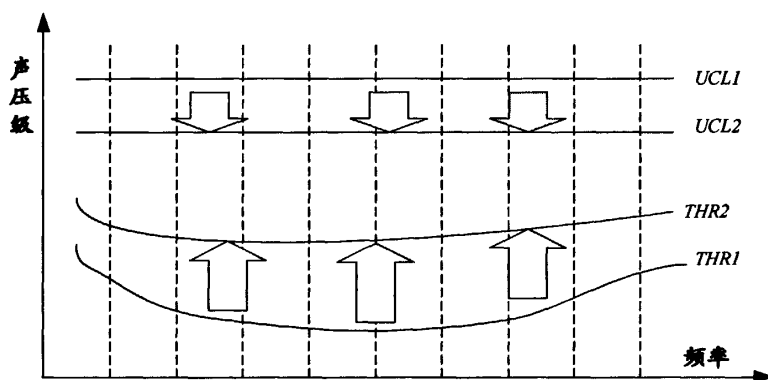


图 2-3 听力正常人和听力不正常人的听力指标比较图

(下标为 1 的是正常人的听力指标，下标为 2 的是耳病患者的听力指标)

Fig.2-3 Normal hearing and hearing no normal hearing indicators compare

(1 is the normal hearing indicators, 2 is for those with hearing ears indicators)

压缩放大主要是用于区别于线性放大，线性放大输入输出关系呈线性的，线性放大技术通常应用在补偿传导性听力损失中，如果声音本身就大，再经过线性放大，那么患者的会因为声音过大而会有轰鸣感。压缩放大为非线性放大，对于平稳信号，其输入输出函数曲线的斜率小于 1，助听器的增益随着输入信号强度的不同而自动的有所变化。感觉神经性听力损失因为神经细胞的损伤，其能感觉到声音的动态范围变窄，要求助听器将强度范围为 30dB 的言语声压缩至听力损失者残余听力的动态范围内。这就要求对小声给予足够的放大，对于大声放大量要减小，这样才能使输入信号较宽动态范围压缩到输出信号较窄的范围之中，让患者能对声音有舒适感，这就是“压缩”真正内涵。正是由于数字信号处理技术的发展，不断提高了压缩的准确性，增加了多样性。尽管人们已知人耳听觉是基于声音频率的敏感，然而对听觉和听力损失的具体生理及听觉心理尚缺乏明确的解释，因此目前通用的补偿算法在很大程度上仍然属于经验算法。

2.2.2 传统的压缩算法

传统的多通道压缩放大方法利用滤波器组将信号分频成 N 个段处理后再综合。利用多通道线路，特定的滤波器滤取特定频率的信号，声音信号被分成 N 个独立的频率区域信号，这些频率区域就被称之为通道。在各个通道内，然后根据患者听力损伤的情况对于不同频段加以不同的放大处理，对不同频率成分使用不同的压缩比，最后再将不同频率的声音段合成在一起，形成完整的声音。多通道不同的压缩比率使听力补偿更吻合患者听力损失情况，不同频率段用不同的放大比，使患者可以很舒适的感受到声音。因为其直接在时域里面对声音进行不同频率的滤波，无须将信号在时域和频域间转换，所以计算量相对较低。目前几乎所有的数字式助听器都采用了这种多通道频响补偿的方案。

多通道压缩系统的结构如图 2-4 所示，每个通道都有一个窄带滤波器，提取本通道的信号，然后各个通道的信号根据不同患者的听力曲线来确定不同的压缩放大比，再用一个特征相同的窄带滤波器进行滤波处理，最后把各个通道的语音信号叠加起来。

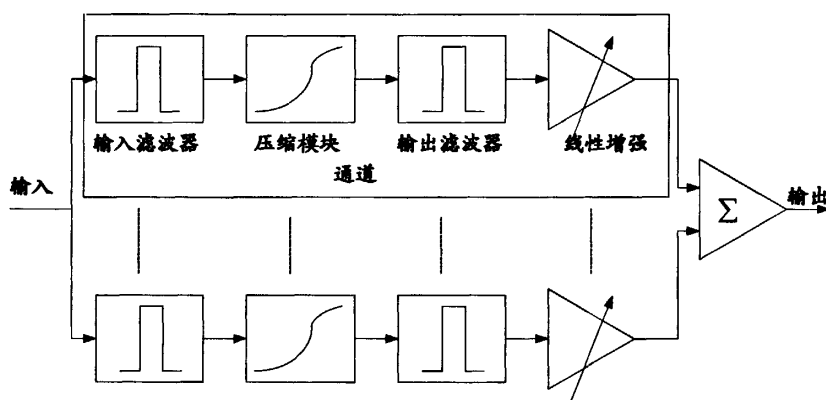


图 2-4 多通道压缩放大系统结构示意图

Fig.2-4 Multi-channel structure of compression amplification system

然而用这种多通道滤波器进行特定频率信号的选取带来了一些问题，就是在频率交接的地方，信号的处理带来了挑战。第一，若共振峰正好位于频段交叠处，会被“劈”成两部分，分别被不同频段给予不同程度的放大。这有可能造成共振峰移动，变形，甚至消失，从听觉生理学上来说，共振峰是人耳区别声音的重要因素，是语音分辨率的重要指标。共振峰在不同通道里面的变化对信号获取和分析能力均弱于常人的耳聋患者来说是一个重大损失；第二，当输入信号通过滤波器组进行放大时，由于助听器放大原理和叠加效应，当信号频率远离交叠区时，此信号获得的放大倍数小于信号频率位于两频段交叠区的放大倍数。如果根据半增益原理估计，单位功率信号增益的差值将高达 3dB，且放大比率越大，增益差值越大。这将会使原有声音频谱波谷消失或增加出一个波峰，从而破坏声音频谱结构。频段分的越多，这样的变形就越多，语音失真就越严重。随着人们对助听器的要求越来越高，上面的问题就越显的严重了。学者们为了解决上面的问题，从如下几个方面入手：

- (1) 不用多通道滤波，回到原来的单通道模式，先找到人耳的听觉特征曲线，在压

缩放大的过程通过听觉特征曲线来调整单频增益。

(2) 使用动态滤波分组，尽量的把一个共振峰划分在同一个滤波器的滤波频率内，但这种分法的计算量比较大

2.2.3 宽动态压缩算法

目前助听器压缩算法上使用的最多的是宽动态压缩算法，其是使用的单通道模式，可以解决上面多通道破坏语音波峰和波谷的问题。WDRC 算法是回到了研究信号频域^[29]的方法上了，先把计算出输入信号各个频域点上的增益，再进行各点频域的压缩放大得到能被耳病患者感知的语音频谱。算法的流程图如图 2-5 所示。

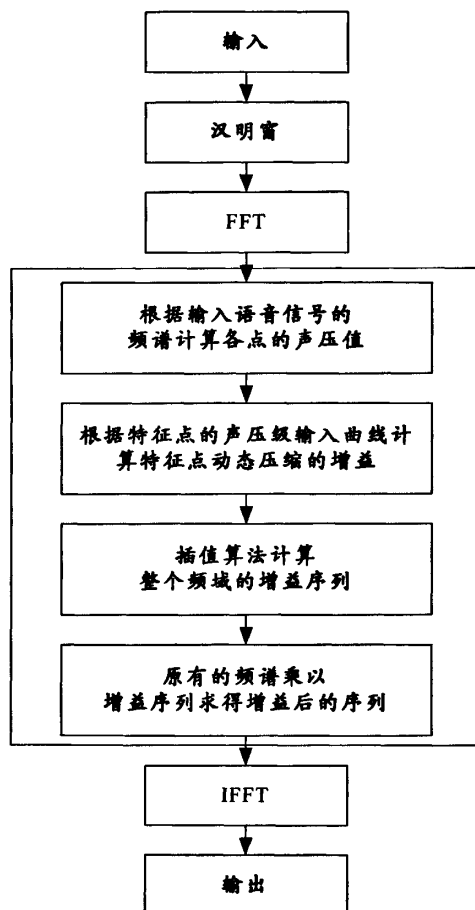


图 2-5 宽动态压缩算法流程图

Fig.2-5 WDRC algorithm flowchart

输入信号经过汉明窗加窗处理后得到一帧信号，再进行 FFT 变换得到频域信号。通常临床上进行听力测听时，主要测听几个特征频率点上的听阈，再把所有的听阈用线相连绘出患者的听力曲线。通常纯音听阈测定选用的特征频率点为 250、500、1000、2000、4000、8000Hz。因此首先根据耳病患者测听数据确定的输入输出函数，再求出特征频率点上的增益。宽动态压缩输入输出曲线见图 2-6 所示，图中横轴表示输入声压级，纵轴表示输出声压级，*THR* 表示听阈，*MCL* 表示最适级，*UCL* 表示不适阈，正常人耳的动

态范围 DRn 完全被映射在患者的动态范围 DRu ，两者并不是满足线性对应的关系，为了保证声强能够映射到患者的最适级，I/O 曲线上就增加了一个拐点。拐点把 I/O 曲线分成两段，拐点下段的压缩比为 CRa ，拐点以上曲线 b 段的压缩比为 CRb ，可得下式：

$$CRa = \frac{MCLn - THRn}{MCLu - THRu} \quad (2.6)$$

$$CRb = \frac{UCLn - MCLn}{UCLu - MCLu} \quad (2.7)$$

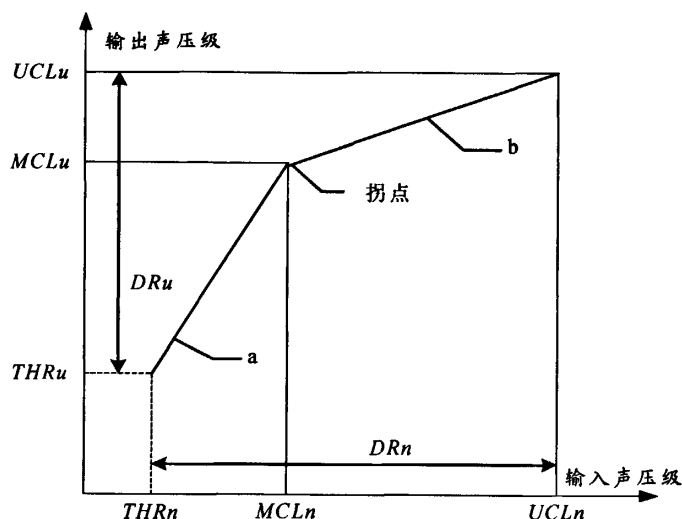


图 2-6 宽动态压缩输入输出曲线

Fig.2-6 I/O curve of WDRC

从图中可以看出 CRa 与 CRb 不一定相等。声压级 (Sound Pressure Level, SPL), $SPLin$ 和 $SPLout$ 分别为输入和输出声压级，结合图 2-6，式 2.6 和式 2.7 可以得到如下的输入输出表达式：

- (1) 当 $SPLin < THRn$ 时， $SPLout = 0$ ，算法不启用压缩，人耳感觉不到声音；
- (2) 当 $THRn \leq SPLin < MCLn$ 时， $SPLout = THRu + \frac{SPLin - THRn}{CRa}$ ，算法启用 a 段压缩，人耳可以感觉到很小的声音；
- (3) 当 $MCLn \leq SPLin < UCLn$ 时， $SPLout = MCLu + \frac{SPLin - MCLn}{CRb}$ ，算法启用 b 段压缩，人耳可以感觉到较大的声音；
- (4) 当 $SPLin = MCLn$ 时， $SPLout = MCLu$ ，人耳感觉最舒适；
- (5) 当 $UCLn \leq SPLin$ 时， $SPLout = UCLn$ ，这主要是为了防止声音增强到人耳的痛阈，算法要消减声压，保持算法的稳定输出。

同过上述的宽动态压缩算法，可以将语音的动态范围压缩到患者的听力动态范围内，让耳病患者能感知到声音。

2.2.4 实验仿真和结果分析

对语音进行宽动态压缩，首先要求得各个特征点的增益值，然后通过插值算法求得其他点的增益，求得增益后乘以各点的值就得到了宽动态压缩后的频谱补偿序列。本文下面所用的实验数据是利用一个中耳炎患者的听力曲线值，根据听力曲线值计算得到了每一点频率所对应的增益，再对一段语音进行处理。本次实验所用的语音数据是语音标准库中女声发音“one”，下面图 2-7 和图 2-8 所示的分别是一帧信号输入输出频谱比较和语音信号输入输出的时域波形。

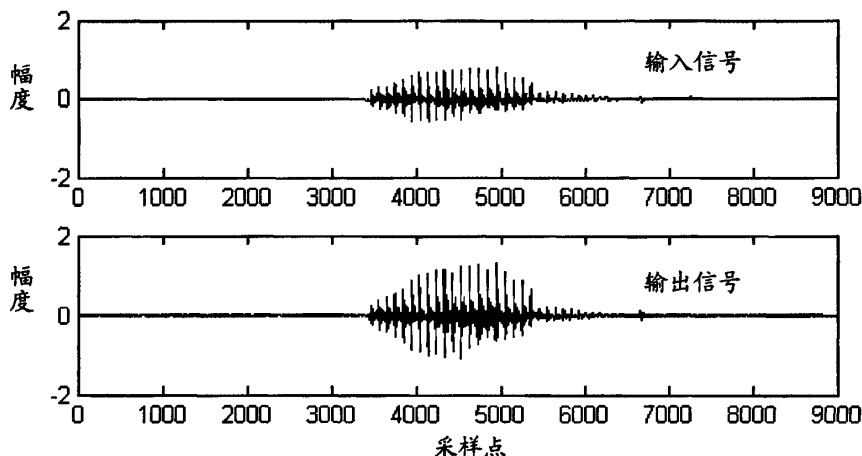


图 2-7 输入输出时域比较

Fig. 2-7 Input-output time-domain comparison

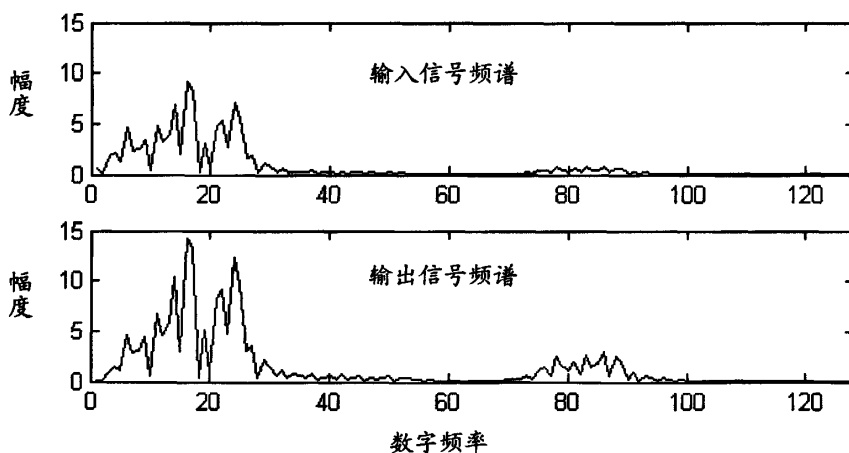


图 2-8 一帧信号输入输出频谱比较

Fig.2-8 A signal input and output spectrum comparison

从图中可以看出，语音经过处理后声音的整体幅度有所增加。图 2-8 语音的频谱图中发现高频部分的增益要大于低频部分的增益，这主要是为了补偿患者的高频听力损失严重的问题。通过宽动态压缩后语音变换到患者听力最适宜的范围里面，提高患者对语音的感知度。

2.3 移频压缩算法

传统助听器的最终目标是提高耳病患者的听力能力和改善听觉舒适感。传统的助听器为了提高听力能力,通常只是改变信号的振幅,信号各个频率点的振幅变大,这种简单的放大可以满足传导型听力损失的患者,而对于高频听力损失的患者这种简单的放大根本解决不了问题,尤其是高频听力损失达到 70dB 时,高频听觉细胞已经完全失去了功能,对任何高频都无法感知,声强的提高对语言的理解没有任何的效果,反而有了负面效果,原因有如下三点:

(1) 受输入输出设备性能的限制,对高频部分的信号放大有很大的限制,特别是 2kHz 以上的信号,那么仅通过压缩放大是不能满足高频听力损失达 60dB 的耳病患者。

(2) 生理学上的听觉掩蔽效应,如果助听器高频部分信号的增益不足话,低频部分的信号就会对高频信号产生掩蔽,但患者的听力损失达到 60dB 时,传统助听器的放大降低语音的分辨能力。

(3) 感觉神经性听力损失的患者要求语音的信噪比是普通人的 4 倍以上,所以对于这种患者要提高语言分辨能力,需要进行去噪处理,而声压的提高不会改善患者的听力。

解决上述的问题,移频^[30]是一个很好的解决方法。感觉神经型听力损失患者其残余的听力大部分集中在低频部分(100Hz~1000Hz),他们只对这个频域段的信号比较的敏感。学者们通过把高频部分语音信号的信息搬移到患者能听到的范围区域,这种技术也就是说“将输入语言信号的带宽,匹配到患耳残余的还有敏感性的有限频带上,而不是企图使受损伤的高频感受区来反映”,听觉生理学上研究表明人耳对声音的辨别能力是靠大脑对频率的比率而不是绝对频率。

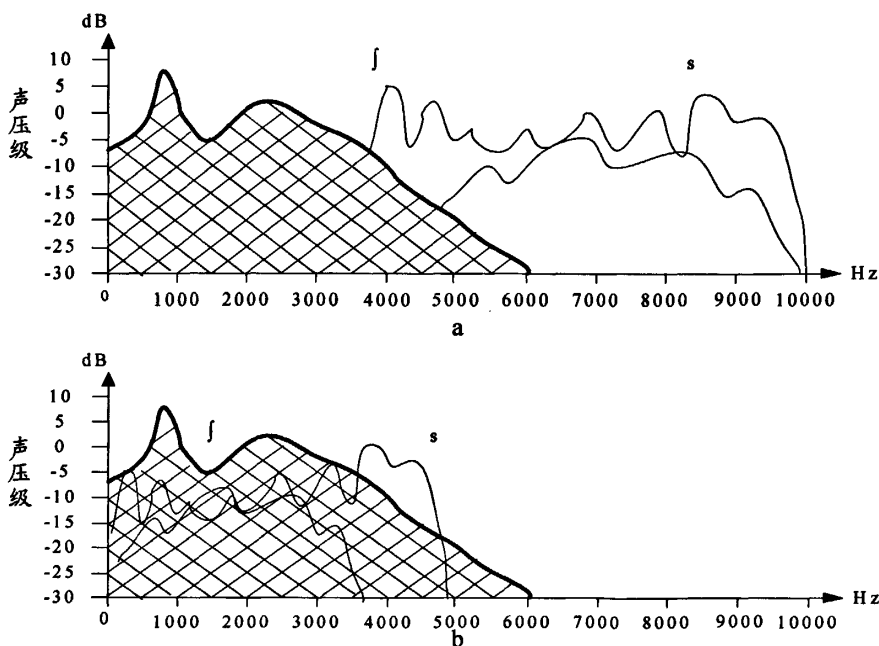


图 2-9 移频技术比率压缩工作原理示意图

Fig.2-9 Frequency shift schematic diagram of technical work

移频技术其实是一种按比例频率压缩。图 2-9 是移频技术的工作原理示意图,图 2-9a 中两条曲线分别代表 /j/ 和 /s/ 的频谱曲线包络图,网格区域表示患者的残余听力频率区域。图 2-9b 中助听器移频算法把 /j/ 和 /s/ 的频率按比例压缩到患者的残余听力频段内,由于是按比率压缩频率的,所以声音的能量峰的关系不变,那么患者可以听到高频部分的辅音,可以对识别其差异。

利用低频段的听觉细胞来听取语音需要注意信号信息流需要经过压缩,因为能感知的神经减少了,如果不加以压缩的话,则人耳无法感知。还需要注意要使语音声波能振动耳朵基底膜的低频段。

移频算法很简单,就是把原有的语音频谱压缩到患者的听觉范围内,根据不同患者的听觉域有不同的变换范围。 $X(f)$ 表示信号的频谱函数,需要移频的原始频段为

$f_{il} \sim f_{ih}$, 患者的听觉域为 $f_{ol} \sim f_{oh}$, 则频域压缩比为 $p = \frac{f_{ih} - f_{il}}{f_{oh} - f_{ol}}$, 压缩的函数表达式

如下:

$$\begin{cases} X'(f) = 0 & f < f_{ol} \\ X'(f) = X[p \cdot (f - f_{ol}) + f_{il}] & f_{ol} \leq f \leq f_{oh} \\ X'(f) = 0 & f_{oh} < f \end{cases} \quad (2.8)$$

其中 $X'(f)$ 为移频压缩后的函数。由式 2.8 可知信号的频谱完全按比率 p 移频压缩到耳病患者的听力范围内,而在听力范围以外的全为 0。图 2-10 所示的是移频算法的算法流程图。

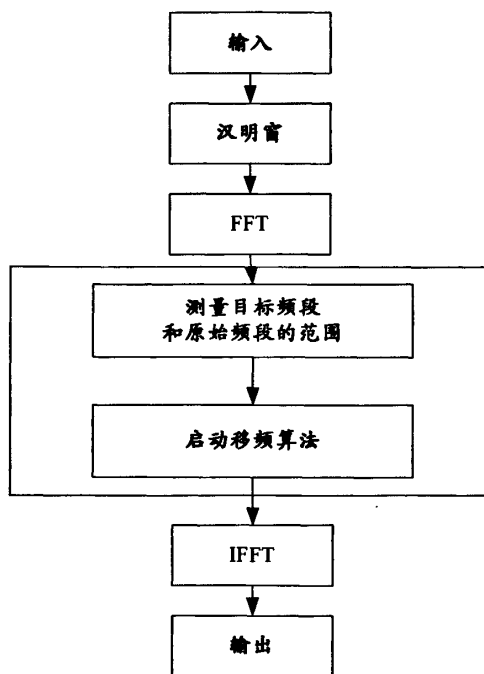


图 2-10 移频算法的流程图

Fig.2-10 Frequency shift algorithm flowchart

对于那些高频部分听力完全损失的耳病患者，如果仅对高频部分语音进行放大的话，患者还是无法听到高频部分的声音，这样对语音的理解会降低。所以对于这部分患者就要使用移频技术，把高频部分的语音压缩到患者的听力范围里面。由于语音是按比例再频域里面压缩，声音的能量峰关系保持不变，根据听力生理学研究表明患者可以分辨高频部分的语音。

下面对语音标准库中一段女声发音进行频谱搬移压缩，将原有的频率段 0-16kHz 变换到目标频段 0-8kHz 中去，压缩比率为 2，下图 2-11 和图 2-12 分别是处理前后的语音时域频域比较图。

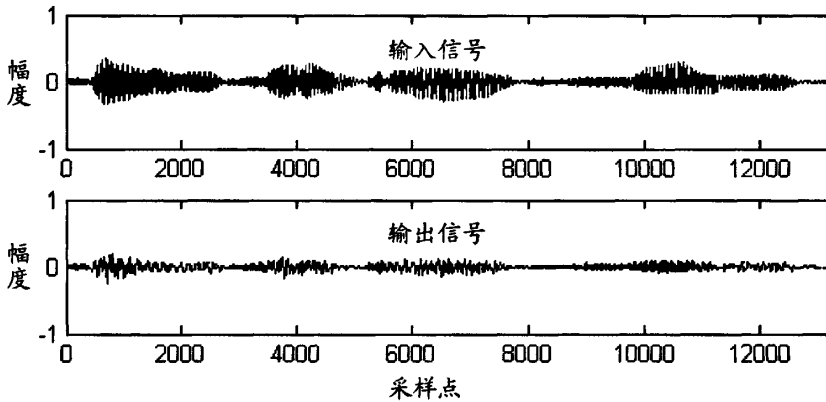


图 2-11 时域波形比较

Fig.2-11 Time-domain waveform comparison

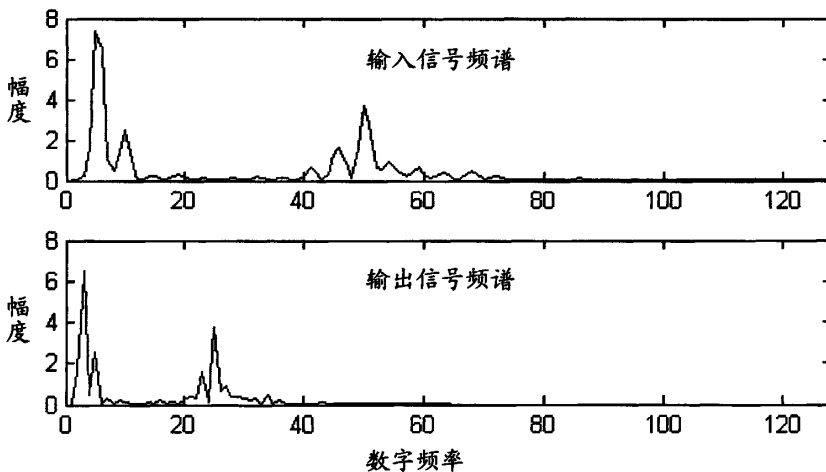


图 2-12 频域比较图

Fig.2-12 Frequency domain compares

从一帧信号的频域图中可以看出，信号按比例压缩到低频，高频部分的能量降低为零。由于其按比例压缩，保留了其包络的基本形状，也就是保留了其时域的基本特征。在图 2-11 中可以看到处理后的时域波形变稀疏了，但还保留了时域信号的基本特征，耳病患者可以通过他低频部分的感知来理解语音信号，达到助听的效果。

2.4 本章小结

在本章节中主要介绍了数字助听器的三个重要算法：语音增强算法、宽动态压缩算法、移频算法。本章对上述的几个算法从原理，算法结构上分析，研究了不同算法间的优点和不足。数字助听器的语音增强算法中常用的是自适应去噪算法，此算法的优点是能很好的提高信噪比，但是语音的音质没有提高。所以本文在不降低信噪比的情况下，提高语音音质，用听觉掩蔽效应的语音增强算法来代替自适应算法。其他算法本章也对他们进行了仿真试验，实验表明这些算法能对耳病患者的听力有所提高。从其长处可以看出这些算法很好的改善感觉神经型听力损失患者的听力，让他们能重新感知到原来不能感知的声音。由于算法都要进行频域的处理，所以计算量比较大。但是数字助听器的核心用的是 DSP，其高速的 FFT 变换，为以上的算法打下了坚实的基础。用听觉掩蔽效应语音增强算法来代替原自适应语音增强算法，这样可以提高语音音质，改善人耳的听觉感知，下一章就介绍了基于听觉掩蔽效应的语音增强算法。

第三章 基于听觉掩蔽效应的语音增强算法

针对数字助听器核心算法之一的语音增强算法,本章主要介绍了基于听觉掩蔽效应的语音增强算法。听觉掩蔽效应的语音增强算法是利用人耳的生理特性来有效的去除噪声,通过人耳的生理特性来研究语音增强是最近几年来研究的一个热点问题。利用人耳的听觉掩蔽效应来实现语音增强可以在不降低信噪比的前提下,改善语音的音质,提高听觉感知舒适度。此类算法加载到数字助听器中去,可以提高患者的听力水平和听觉的舒适度。

3.1 噪声的特性及分类

噪声的定义就是信号中无用的信号部分。噪声无处不在,噪声的种类很多,特性也不一样。要消除不同的噪声信号,根据其特性采用不同的方法。

最常见的噪声是白噪声和其他的有色噪声^[31]。白噪声是指其频谱在整个频域里面分布均匀。它分布在整个频域,在每个频点的能量为常数,且基本恒定,其分布是随机,不具有相关性,所以不管用高通、低通还是带通滤波器均不能消除白噪声。而有色噪声则是在其频谱里面有相当宽的部分出现近似连续的分量,即能量分布不均匀。

如果根据与语音信号的关系,噪声可以分为加性噪声和非加性噪声。加性噪声就是独立于有用的信号,始终存在着。非加性噪声中绝大部分是乘性噪声,它是与有用信号共存在的,有用信号在,它就存在。而在语音信号处理里面主要讨论加性噪声,它们主要是周期性噪声,脉冲噪声,宽带噪声和同声道语音干扰噪声等。

以上介绍了语音处理里面常遇到的噪声,对其进行了分类,阐述了他们的特点,为下面的语音去噪提供了基础。

3.2 语音及人耳的听觉掩蔽特性

3.2.1 语音特性

语音信号是携带语音信息的语音声波。

语音信号是非平稳信号。人类发声系统在产生不同语音时生理结构的不同产生了非平稳的语音信号。但人的生理结构的局限性,其非平稳性的受到限制,在一段时间里面(10~30ms)可以认为人发声系统的特征基本不变,因此语音又可以认为有短时平稳性,语音增强正是利用的这一特性。

语音信号可以看作一个随机过程,可以利用统计论里面的方法来分析讨论。但由于语音信号的非平稳性、非遍历性,因此长时间的时域统计对语音增强没有什么意义。语音的短时谱幅度统计特征是时变的,只有当分析的一帧信号趋于无穷大时,才能近似的认为具有高斯分布,均值为0。在有限帧长时,高斯分布只是一种近似的描述。

语音信号基本上可以分为清音,浊音和爆破音,它们的产生机理和特征有很大的区别。当气流通过声门时,声带不振动,而在某处收缩,迫使气流以高速通过这一收缩部分产生了清音;如果声带的张力刚好使声带发生张弛振荡,产生一股准周期气流,这一气流激励声道就产生了浊音;如果使声道完全闭合,在闭合后建立起气压,然后突然释放就得到了爆破音。清音看上去类似于白噪声,没有明显的时域和频域特性,很难和白噪声区分开来。而浊音则有明显的周期性,其分布主要在低频段,而且在频谱上会出现共振峰。利用这一特性可以用梳状滤波器提取语音分量、消弱非语音信号达到语音增强的目的。

3.2.2 人耳的听觉掩蔽特性

人类的听觉系统是一个十分巧妙高端的语音信号处理系统^[32],它处理信号能力是现在任何一个语音处理系统所不能模拟的。

人耳主要分为外耳、中耳和内耳三个部分。外耳主要起到收集外来声音进入外耳道的作用;中耳是连接外耳和内耳的三个听小骨,其起到声音传输放大的作用;听力最重要的部分是内耳的耳蜗。耳蜗就是绕成两圈半的骨质细管,沿其长度以基底膜分为两半。按欧姆分析,后得生理学解剖证实,基底膜的作用如一频率分析器。从耳朵的蹬骨处高频,逐渐降低,到蜗孔处的低频,各点对不同频率共振。基底膜上与频率相应的部分发生共振,其上的毛细胞弯曲,发出神经脉冲,经耳蜗神经传入大脑,产生听觉感知。人耳产生听觉最简单,最基础的是响度级,听力觉察到声音的大小强弱与声音的强度有关,但不完全由强度决定。因为人耳的振动系统、传递系统、换能系统等都与频率有关,因而主观的强弱与客观的强度有关,但也受到频率的影响。

正常人的听觉系统非常的灵敏,人耳可以听到频率从 20Hz~16kHz 的声音,而年轻人的听力上限可以到达 20kHz,老年人则由于高频感知部分的细胞损伤听力上限下降到 10kHz。人耳听觉的主观感知主要是响度、音调和掩蔽效应。

测量声音强弱的物理量是声强,单位是 W/m^2 。当声音的强度小到人耳刚刚能够听到时,称为听阈。听阈是随着频率的变化而变化的,外界的声音变化也会影响到听阈的变化。如果加大声音的强度,使其大到人耳不能忍受的时,称为痛阈。听阈和痛阈之间就是人耳的听觉范围。从声强上来看,正常人的听觉范围可以达到 120dB。

音调是听觉分辨声音高低时,用于描述这种感受的一种特性。一般的频率低的声音音调低,频率高的音调高,但他们间是不成正比的,音调还与声音的波形和强度有关。

人耳除了上面介绍的有时频分析的能力,还有一个重要的特性就是掩蔽特性^[1]。当两个响度不同的声音作用于人耳时,则响度较高的频率成分的存在会影响到对响度较低的频率成分的感受,使其变得不易被人耳察觉,这种现象叫做掩蔽效应。一般来说,对于中等掩蔽强度,纯音最有效的掩蔽出现在它的频率附近,低频的纯音可以有效的掩蔽高频的纯音,而高频的纯音对低频的纯音的掩蔽作用较小,这是掩蔽效应的基本原理。

掩蔽效应分为同时掩蔽^{[1][2]}和异时掩蔽两种。同时掩蔽指掩蔽现象发生在掩蔽者和

被掩蔽者同时存在时,也称作频域掩蔽,声音能否听到取决于它的频率和强度。正常人听觉的频率范围为 20Hz~20kHz,强度范围为-5dB~130dB。在听觉域内,人耳对声音的响应随着频率而变化,最敏感的频率段时 2kHz~4kHz。听阈是人耳刚好能听到的最低声压级,他是声音频率的函数。图 3-1 中的虚线是人耳在安静时听阈曲线。

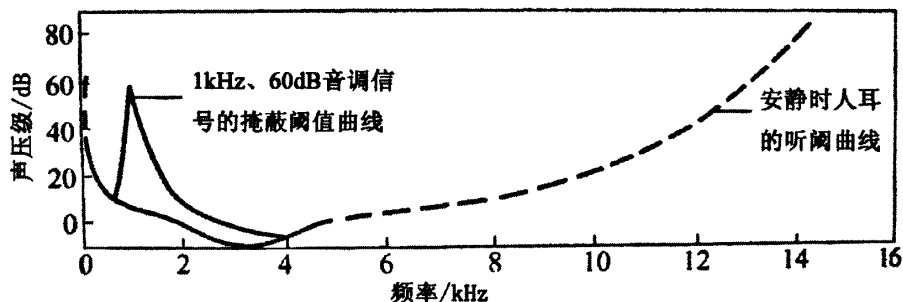


图 3-1 频率为 1kHz 且声压级为 60dB 的音调信号掩蔽阈值曲线

Fig.3-1 Signal of 1 kHz and 60 dB sound pressure level masking threshold curve

一个较强信号(掩蔽者)的存在,听力阈值不等于安静时的阈值。在掩蔽者频率的领域里,听力阈值被提高。如图 3-1 中 1kHz 所示的实线,60dB 的音调信号产生的掩蔽曲线。利用这一特性可以把被掩蔽掉的弱信号(如比较小的噪声信号)看作与人耳没有关系的信号,不对其进行处理;再对噪声的频谱进行适当的整形,使噪声低于掩蔽阈值,让噪声不被人耳所感知。

异时掩蔽发生在掩蔽者和被掩蔽者不同时存在时,分为前掩蔽和后掩蔽。若掩蔽效应发生在掩蔽者开始前的某段时间里称为前掩蔽;若发生在之后则为后掩蔽。图 3-2 给出的是同时掩蔽和异时掩蔽现象,从图中可以看出在掩蔽者存在时这种掩蔽现象一直有效,当掩蔽者消失后,掩蔽现象还持续了一段时间。这个特性主要用在计算各频段的掩蔽阈值之间的影响。

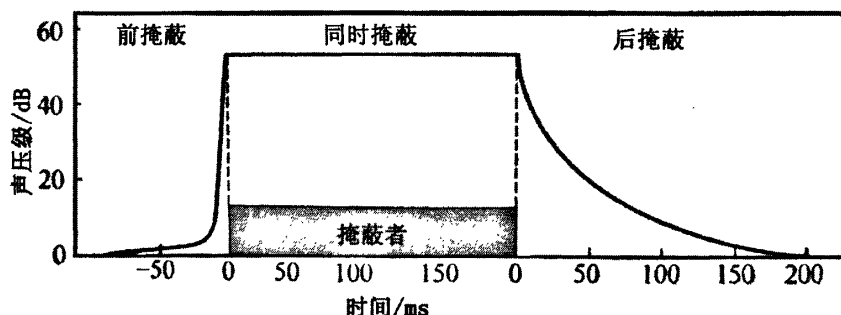


图 3-2 3 种掩蔽现象的强度以及持续时间

Fig 3-2 Three kinds of masking phenomenon and the continued strength of the time

人类听觉系统的掩蔽效应是用一个数学心理学模型来描述,依据该模型可以估算出掩蔽者的掩蔽阈值。掩蔽阈值取决于掩蔽者的音调性、频率、声压级和持续时间。图 3-3 描述的是一个掩蔽者产生的掩蔽阈值曲线。

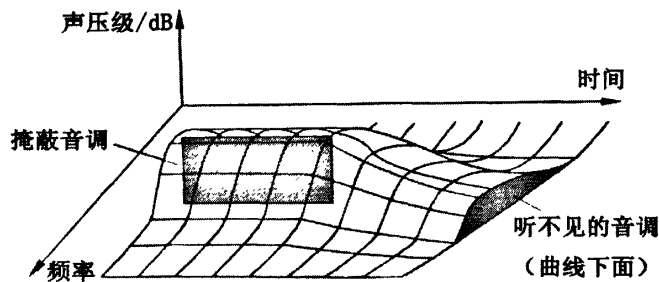


图 3-3 掩蔽阈值曲线
Fig.3-3 Masking threshold curve

从图中可以看出掩蔽阈值是时间、频率和声压级的函数，并且掩蔽阈值随着掩蔽音调的变化而变化。

声音的掩蔽者分为 3 种类型，纯音调、宽带噪声和窄带噪声。

(1) 纯音调信号间的掩蔽，这是指掩蔽者和被掩蔽者都市纯音调信号，这种掩蔽效应比较的简单，掩蔽阈值曲线的低频段陡峭，高频段比较的平坦。

(2) 宽带噪声对纯音的掩蔽，掩蔽者是宽带噪声，被掩蔽者是纯音调信号。虽然白噪声的功率谱是平坦的，但是它产生的掩蔽阈值却只在低频段保持着水平。在频率大约为 500Hz 以上时，掩蔽阈值随着频率的增大而提高，每 10 倍频程大约提高 10dB。在低频段，掩蔽阈值一般高于噪声功率谱密度 17dB。

(3) 前面两种掩蔽者对纯音的掩蔽效应比较简单，主要受声压级的影响；而窄带噪声对纯音的掩蔽是一种比较复杂的过程，掩蔽阈值随着声压级的不同而变化，并且随着窄带噪声的中心频率的变化而变化。中心频率不同的窄带噪声产生的掩蔽阈值曲线形状是不同的，在低频段曲线比较的窄，随着频率的增高，曲线逐渐变宽。

一个纯音可以被以它为中心频率且具有一定带宽的连续噪声所掩蔽，如果在这一频带内噪声功率等于该纯音功率，这时该纯音刚好处于能听到的临界状态，这一带宽为临界带宽。连续的临界频带序号为临界频带域，通常将 20Hz~16kHz 之间的频率分为 24 个频率群，表 3.1 所示的是临界带宽表。

表 3.1 临界带宽表
Tab.3.1 Critical bandwidth Table

频率群序号	中心频率/Hz	临界带宽/Hz	下限频率/Hz	上限频率/Hz
1	50	80	20	100
2	150	100	100	200
3	250	100	200	300
4	350	100	300	400
5	450	110	400	510
6	570	120	510	630
7	700	140	630	770
8	840	150	770	920

9	1000	160	920	1080
10	1170	190	1080	1270
11	1370	210	1270	1480
12	1600	240	1480	1720
13	1850	280	1720	2000
14	2150	320	2000	2320
15	2500	380	2320	2700
16	2900	450	2700	3150
17	3400	550	3150	3700
18	4000	700	3700	4400
19	4800	900	4400	5300
20	5800	1100	5300	6400
21	7000	1300	6400	7700
22	8500	1800	7700	9500
23	10500	2500	9500	12000
24	13500	3500	12000	15500

人耳的基底膜具有频谱分析的能力，频率群的划分相应地将基底膜划分为很多小的部分，每一个部分对应着一个频率群。掩蔽效应就是在这些频率群内发生，这是因为同一个频率群的声音作用于基底膜的相同部分，它们在大脑中似乎是叠加在一起来评价，如果这时发声，可以互相掩蔽。

本课题在谱减法的基础上结合听觉掩蔽效应对谱减系数进行修改，改进了语音质量。

3.3 基于听觉掩蔽效应和噪声实时估计的语音增强算法

在实际的环境中，外界的噪声总是对语音信号有干扰，环境的噪声会影响语音的质量，在高噪声的环境下，人耳对语音将无法分辨，就不能从环境中听取到想要的信息。所以语音增强技术应声而出，并且得到了很快的发展。

目前基于短时谱估计的方法有谱减法、维纳滤波和最小均方误差估计等方法。维纳滤波和谱减法出发点不同，但形式没有太大差别。谱减法可以去除相当一部份背景噪声，信噪比提高了，但是其还残留下来的音乐噪声较大；改进的谱减法通过增加两个参数 α 和 β 来调整噪声功率谱的大小，但由于两个参数是先验得到的，适应性较差，而且残留噪声大；P.lockwood&J.Boudy 提出根据信噪比来自适应调整谱减系数 α 和 β 的非线性谱减法，信噪比有很大的提高，但信噪比的提高不等于听觉质量的提高，语音最终的目的是要被人耳感知；最小均方误差(Minimum Mean Square Error ,MMSE)方法要预先估计语

音谱分布,因而计算量大且语音的频谱很难估计。针对上述算法的不足,本算法利用人耳的听觉掩蔽效应对谱减法的 α 和 β 参数进行修正。首先对噪声进行实时估计^[33],利用传统谱减法得到初步的增强信号,计算听觉掩蔽阈值^[34],根据听觉掩蔽阈值调整参数 α 和 β 。针对谱减法残留的音乐噪声,利用基于听觉掩蔽效应的感知滤波器进行滤波,进一步把音乐噪声滤除。这种方法可以很好的抑制残留的音乐噪声,同时也能提高语音的听觉质量。

3.3.1 参数可调的谱减法

含噪语音时域模型可以表示为

$$y(n) = s(n) + d(n) \quad (3.1)$$

式中 $y(n)$ 是含噪语音, $s(n)$ 是纯净语音信号, $d(n)$ 是加性噪声, $s(n)$ 和 $d(n)$ 统计不相关。一帧加窗后的语音信号可以表示为

$$y(m, n) = s(m, n) + d(m, n) \quad m = 1, 2, 3, \dots; n = 0, 1, \dots, N-1 \quad (3.2)$$

式中 m 是帧号, N 是帧长。进行FFT变换,在短时谱减法中,用含噪语音的短时谱幅度 $|Y(m, k)|$ 减去噪声谱的估计值 $|N(m, k)|$ 可以得到增强后的语音谱,如下所示:

$$|S(m, k)| = \left[|Y(m, k)|^2 - E(|N(m, k)|^2) \right]^{\frac{1}{2}} \quad (3.3)$$

其中 $|S(m, k)|$ 是增强后的语音功率谱。在得到增强后的语音功率谱后,进行反变换得到增强后的语音,这种谱减法简单易行但信噪比不高,残留的噪声大。在原有的谱减法基础上增加了两个谱减系数 α 和 β 来改善信噪比,其数学表达式如下

$$S(mk) = \begin{cases} \left(1 - \alpha \left[\frac{|N(mk)|}{|Y(mk)|} \right]^2 \right)^{\frac{1}{2}} \times Y(mk) & \left[\frac{|N(mk)|}{|Y(mk)|} \right]^2 < \frac{1}{\alpha + \beta} \\ \left(\beta \left[\frac{|N(mk)|}{|Y(mk)|} \right]^2 \right)^{\frac{1}{2}} \times Y(mk) & \text{其它} \end{cases} \quad (3.4)$$

其中 α 和 β 为谱减系数,可以通过适当的调整这两个系数值提高信噪比。但由于谱减系数是先验得到的,适应性较差,而且残留噪声大,影响听觉效果。

本文在调整 α 和 β 系数时,利用听觉掩蔽值来修正。当两个响度不同的声音作用于人耳时,则响度较高的频率成分的存在会影响到对响度较低的频率成分的感受,使其变得不易被人耳察觉,这种现象叫做掩蔽效应。利用听觉掩蔽阈值来调整谱减系数,可以

减少音乐噪声的影响，提高语音质量。听觉掩蔽阈值 ($T(m, k)$) 大说明在此频率段中人耳对其它相近的频率段的语音信号和噪声信号的抗干扰能力比较的强，此时采用较小的谱减系数，这样可以不减少影响语音质量的频率部分。如果掩蔽阈值比较小，则采用较大的谱减系数。则谱减系数的计算公式^[35]如下

$$\frac{T_{\max} - T(m, k)}{\alpha - \alpha_{\min}} = \frac{T(m, k) - T_{\min}}{\alpha_{\max} - \alpha} \quad (3.5)$$

$$\frac{T_{\max} - T(m, k)}{\beta - \beta_{\min}} = \frac{T(m, k) - T_{\min}}{\beta_{\max} - \beta} \quad (3.6)$$

实验表明在兼顾信噪比和听觉质量的前提下，选取 $\alpha_{\max} = 8$ 和 $\alpha_{\min} = 1$ ， $\beta_{\max} = 0.02$ 和 $\beta_{\min} = 0$ 。其中 T_{\max} 和 T_{\min} 为听觉掩蔽阈值 T 的最大值和最小值。

3.3.2 掩蔽阈值的计算和感知滤波器

(1) 噪声实时估计

在以往的谱减法中用的噪声估计都是采用的静音检测，算法对语音进行语音帧的语音/寂静段进行检测，用寂静段的噪声估计来代替语音段的噪声，这样就增加了估计误差。通过观测可以发现含噪的语音功率谱可以很快地衰减到噪声功率谱的水平，所以通过观察含噪语音的功率谱的最小值并对其进行修正来估计噪声功率谱。

先对含噪语音进行平滑处理，改善谱估计质量。

$$P(m, k) = \lambda P(m-1, k) + (1 - \lambda) |Y(m, k)|^2 \quad (3.7)$$

其中 $P(m, k)$ 为平滑后的语音谱， λ 为平滑系数。在一个长度为 τ 的时间窗内，找出平滑后的最小值 $P_{\min}(m, k)$ ，在每输入一帧语音信号更新其最小值。考虑到与最小值与噪声之间的偏差，采用偏差补偿算法，对最小值进行补偿，得到最后的噪声估计值 $N(m, k)$ 。

(2) 掩蔽阈值的计算

先将语音信号进行临界频带分析。在得到噪声估计值后，利用传统的谱减法得到近似的语音信号功率谱 $S(m, k)$ 进行频率域划分，通常划分为 24 个临界频带，语音信号在每个临界频带里的能量为

$$B_i = \sum_{k=bl_i}^{bh_i} S(m, k) \quad (3.8)$$

B_i 表示是各个临界频带的能量， bl_i 和 bh_i 分别表示临界频带的 i 的下限和上限， i 表示得是临界频带序号。考虑到各个频带间的相互掩蔽的关系，定义传递函数^[36]

$$SF_{ij} = 15.81 + 7.5(\Delta + 0.474) - 17.5\sqrt{1 + (\Delta + 0.474)^2} \quad (3.9)$$

其中 $\Delta = i - j$ 表示两个频带的频带号的差值. 考虑到相互的影响后给出了最后的能量公式

$$C_j = \sum_{i=1}^{i_{\max}} B_i \times SF_{ij} \quad j = 1, 2, \dots, i_{\max} \quad (3.10)$$

由于噪声和音调的掩蔽特性的不同, 先要判断各个频带里是噪声特性还是音调特性, 可以根据谱平坦度测度(Spectrum Flatness Measure, SFM)定义如下

$$SFM = \frac{G_i}{A_i} \quad (3.11)$$

G_i 为各个频带内功率谱的几何平均, A_i 为各个频带内功率谱的算术平均。 $SFM \in [0, 1]$, 为 0 表示频带的信号的具有纯音调特性, 纯音的掩蔽阈值偏移量为 $(14.5 + i)$ dB。为 1 具有白噪声特性, 偏移量为 5.5 dB。根据 SFM 定义如下的音调系数

$$\phi = \min\left(\frac{SFM}{-60}, 1\right) \quad (3.12)$$

相对的掩蔽阈值的偏移量为

$$O_i = \phi(14.5 + i) + 5.5(1 - \phi) \quad (3.13)$$

掩蔽阈值的计算式如下

$$T(m, i) = 10^{\log_{10}(C_i) - (O_i / 10)} \quad (3.14)$$

得到的掩蔽阈值再扩展到各个频谱上, 记为 $T'(m, k)$, 最终的掩蔽阈值公式如下

$$T(m, k) = \max(T'(m, k), T_a(m, k)) \quad (3.15)$$

其中, $T_a(m, k)$ 是绝对听觉阈值, 计算公式为

$$T_a(m, k) = 3.64f^{-0.8} - 6.5\exp(f - 3.3)^2 + 10^{-3}f^4 \quad (3.16)$$

其中 f 为信号频率

(3) 感知滤波器

在得到初步增强的语音信号后, 还有一些残余噪声, 而这些残余噪声可以利用人的听觉掩蔽特性掩蔽掉, 不需要全部消除, 如果完全消除会降低语音的可懂度, 增加语音的失真。在此基础上用基于听觉掩蔽效应的感知滤波器进行滤波。

在这里假设增强后的语音信号为 $\hat{S}(m, k) = G(m, k) \times S(m, k)$, $G(m, k)$ 为感知滤波器。

让残留的噪声功率谱低于掩蔽阈值，让噪声不被人耳所察觉。所以感知滤波器的函数要满足如下的条件

$$|G(m,k)|^2 |N(m,k)|^2 \leq T(m,k) \quad (3.17)$$

其中 $T(m,k)$ 为掩蔽阈值，在计算得到阈值后，可以设计出如下的感知滤波器。

模型如下

$$G(m,k) = \min \left[\theta \times \sqrt{\frac{T(m,k)}{|N(m,k)|^2}}, 1 \right] \quad (3.18)$$

θ 是一个小于 1 而大于 0 的系数，在实验的过程中发现加上了 θ 这个系数，可以很好的抑制音乐噪声，效果非常明显。在本实验中 θ 取值为 0.8。

3.3.3 算法步骤

通过上面的分析讨论下面给出基于听觉掩蔽效应和噪声实时估计的语音增强算法的实现过程：

- (1) 根据式 3.7 的语音平滑谱实时估计出噪声 $N(m,k)$ ；
 - (2) 利用式 3.3 计算出近似的纯净语音谱 $S(m,k)$ ，为下面听觉掩蔽阈值的计算打好基础；
 - (3) 利用步骤 2 中的近似纯净语音谱计算听觉掩蔽阈值 $T(m,k)$ ；
 - (4) 利用式 3.5 和式 3.6 计算谱减系数 α 和 β ；
 - (5) 把步骤 4 中计算得到的谱减系数利用式 3.4 进行语音增强，此时的谱减系数根据听觉掩蔽阈值自动调整，提高语音音质；
 - (6) 利用式 3.18 计算得到感知滤波器，进一步抑制音乐噪声，提高信噪比；
- 以上是基于听觉掩蔽效应和噪声实时估计语音增强算法的步骤分解。

3.3.4 实验仿真和结果分析

根据本文的分析，对所用的语音增强算法进行性能评测，并和传统的谱减法进行了比较。由于人耳对语音的相位感知不敏感，可以用含噪语音的相位来代替增强语音信号的相位，根据增强后的语音幅度谱，利用逆傅立叶变换就可以得到增强后的信号。

在本文中，纯净语音选取 863 中文语音识别语料库里面的一段语音，噪声选取的是高斯白噪声，语音数据的采样率为 8kHz，帧长为 256，窗口滑动为 128，数据窗采用的是汉明窗。本文将语音频率段分为 24 个临界频带来计算掩蔽阈值，语音谱的平滑系数 λ 取 0.96，经过多种信噪比的信号类型分析，时间窗长度 τ 取 0.4s 时，实验效果好。

采用的客观测试指标包括全局信噪比(Signal-to-Noise Ratio, SNR)和语音失真测度(Itakura-Saito, IS)，SNR 和 IS 的主要区别是 SNR 不能反映人们对语音信号的听觉质量而

IS 语音失真测度在一定程度上可以反应人们对语音信号的主观感受, IS 系数越小, 说明语音的品质越好。实验结果如表 3.2 所示。

表 3.2 算法的比较
Tab.3.2 Comparison algorithm

输入信噪比 (dB)		7.662	4.639	2.527
全局信噪比 (dB)	A	10.48	8.239	7.125
	B	10.23	8.363	7.117
	C	9.266	6.823	4.2
语音失真 测度 (IS)	A	1.126	1.222	1.297
	B	3.142	2.163	2.332
	C	3.216	2.781	2.438

A: 本文所用的算法; B: 改进的谱减法; C: 传统的谱减法

从表3.2可看出本算法在不同输入信噪比的含噪信号的全局信噪比和其他算法相差不多, 但在失真测度值上, 本算法的IS值要小于其他算法, 尤其是在输入信噪比较大的时候, 本算法的IS值要远远小于原有的算法。因为根据听觉掩蔽值来调整谱减系数, 在听觉掩蔽阈值较大时, 采用较小的谱减系数, 这样可以保证有用的语音能充分的保留下来, 保证语音的质量; 掩蔽阈值较小时, 采用较大的谱减系数。本文中尤其是采用了噪声的实时估计, 减少了噪声估计误差, 在最后的输出端加上了感知滤波器, 使得有用信号得到了最大的保留, 语音失真度达到了最低水平。

通过主观测听增强语音可发现, 本文的算法可以使音乐噪声达到很低的水平, 使增强后的语音质量有一定的提高, 且音乐噪声的残留量比较的小, 说明了IS失真测度与音质的主观评测有较大的相关性。

3.4 听觉掩蔽效应语音增强的改进算法

绝大部分语音增强系统在语音增强后都会残留下来音乐噪声, 这种音乐噪声会给人带来烦躁感。音乐噪声主要原因是噪声估计的误差, 在频谱上噪声的随机尖峰的很难完全消除, 这样在增强后的语音中就会夹杂着有节奏的音乐残留噪声。目前针对音乐噪声的处理方法有很多: 通过增加两个谱减系数来调整信噪比的改进谱减法, 但是谱减系数是先验得到的, 适应性较差; P.Lockwood&J.Boudy 提出了非线性谱减法, 根据语音的信噪比自适应调整语音的谱减系数, 增强了算法的鲁棒性, 但是增强的语音中还是残杂着音乐噪声; Nathalie Virag 将人耳的听觉掩蔽特性应用到非线性谱减法, Ching-Ta Lu 把听觉掩蔽特性应用到去除色噪声, 这些算法直接应用人耳的听觉特性, 从听觉感知上去除音乐噪声, 很好的解决了音乐噪声的问题, 但由于噪声估计的误差, 从而导致了听觉掩蔽阈值计算的误差, 增强语音中还是残留着音乐噪声。

本文实现的算法主要是去除音乐噪声的影响，消除人耳对增强语音的不适。本算法是以听觉掩蔽效应的谱减法为基础，在语音的预处理上，利用阈值判断，先去除无语音段的噪声影响，再利用听觉掩蔽阈值来调整谱减系数。由于掩蔽阈值的计算误差，本文提出了在计算谱减系数的时候，在计算公式中增加了修改参数 θ ，来抑制谱减系数的过分增大，提高了谱减系数适应性。在最后基于 MMSE 准则，在保证语音失真最小的前提下，计算出平滑系数来进一步消除音乐噪声。该算法使含噪语音的信噪比有了很大的提高，实时性好，鲁棒性强，而且音乐噪声几乎完全抑制了。

3.4.1 改进的听觉掩蔽效应算法

在实际的语音通信中，语音段中存在大量的寂声段，这些寂声段掺杂着噪声，在频域里面利用噪声的频谱特性判断出噪声，对其进行处理，抑制噪声。通过分析可知一帧噪声的短时能量要远小于语音的短时能量，这样可以设置一个阈值来作为判断噪声/语音的标准。如果一帧信号的能量小于阈值则判断为噪声，直接将其频域的幅度谱置为 0；如大于则进行下一步语音增强处理。在实验的过程中发现帧噪声能量和的最大值和最小值约为 2 倍的关系，设置阈值为最小值 2 倍时，噪声段判断准确。有语音段的语音增强，用含噪语音的短时谱幅度 $|Y(m,k)|$ 减去噪声谱的估计值 $|N(m,k)|$ 可以得到增强后的语音谱。

用式 3.4 来实现语音增强， $S(m,k)$ 是增强后的语音谱， m 为帧号， k 为离散频率。谱减系数 α 和 β 来改善信噪比，提高增强效果。其中谱减系数如果是先验得到，则算法的适应性较差，在本文中利用听觉掩蔽阈值来自适应调整谱减系数 α 和 β ，提高了算法的鲁棒性。

这里把增强后的语音 $S(m,k)$ 与纯净语音 $X(m,k)$ 的误差定义为语音失真 $E_s(m,k)$ 和残留噪声 $E_r(m,k)$ 的和，表达式如下

$$E(m,k) = S(m,k) - X(m,k) = E_s(m,k) + E_r(m,k) \quad (3.19)$$

其中 $E_s(m,k)$ 和 $E_r(m,k)$ 可用如下公式^[37]代替

$$\begin{aligned} E_s(m,k) &= [g(m,k) - 1]X(m,k) \\ E_r(m,k) &= g(m,k)D(m,k) \end{aligned} \quad (3.20)$$

其中 $D(m,k)$ 为噪声谱，在这里用 $N(m,k)$ 来代替， $g(m,k) = (1 - \alpha \frac{N^2(m,k)}{Y^2(m,k)})$

利用人耳的生理特性，人耳对声音的是否感知，听觉阈值在其中起着决定性的作用。当声音高于听觉阈值时人耳才对其有感知，而小于时是不能感知。利用听觉掩蔽阈值来调

整谱减系数,可以减少音乐噪声的影响,提高语音质量。听觉掩蔽阈值($T(m,k)$)大说明在此频率段中人耳对其它相近的频率段的语音信号和噪声信号的抗干扰能力比较的强,此时采用较小的谱减系数,这样可以不减少影响语音质量的频率部分。如果掩蔽阈值比较小,则采用较大的谱减系数。则传统的谱减系数的计算式 3.5 和 3.6 如下

$$\frac{T_{\max} - T(m,k)}{\alpha - \alpha_{\min}} = \frac{T(m,k) - T_{\min}}{\alpha_{\max} - \alpha} \quad \frac{T_{\max} - T(m,k)}{\beta - \beta_{\min}} = \frac{T(m,k) - T_{\min}}{\beta_{\max} - \beta}$$

选取 $\alpha_{\max} = 8$ 和 $\alpha_{\min} = 1$, $\beta_{\max} = 0.02$ 和 $\beta_{\min} = 0$ 。

但在计算掩蔽阈值时首先要得到纯净的语音谱,而通常的是用近似的语音谱来代替纯净语音谱。要从含噪语音谱中得到纯净谱的估计,首先要估计出噪声谱,利用语音谱和噪声谱的不相关性,从含噪谱中减去噪声谱得到纯净语音谱的估计。

在得到近似的纯净语音谱后,进一步估计听觉掩蔽阈值。噪声估计的误差会引起听觉阈值与真实值之间的误差。图 3-4 给出了一帧信号听觉掩蔽阈值估计值与真实值的比较。

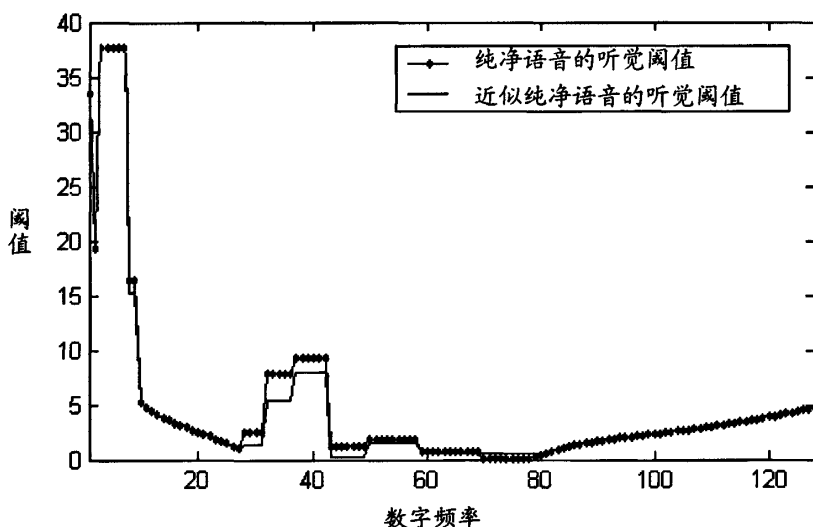


图 3-4 一帧语音信号听觉掩蔽阈值的近似值与准确值比较

Fig.3-4 The compare of the reality and the estimate about noise masking

由图可以很明显的看出听觉阈值的误差,听觉阈值的误差从而导致谱减系数的误差。本文针对谱减系数的计算公式进行了改进。将式 3.6 变换成如下形式

$$\alpha = -T \frac{\alpha_{\max} - \alpha_{\min}}{T_{\max} - T_{\min}} + \frac{T_{\max} * \alpha_{\max} - T_{\min} * \alpha_{\min}}{T_{\max} - T_{\min}} \quad (3.21)$$

实验表明,有些频率区域的听觉阈值要比语音的谱值小很多,同时在这些频率段内计算出来的掩蔽阈值要比真实值小,产生这一现象的原因是在语音谱丰富的频域里面语音变化快,噪声的实时估计误差偏大,从而导致为计算掩蔽阈值而初步得到的纯净语音谱在这些频率段里语音谱偏小,最终导致掩蔽阈值偏小。如果直接用原来的公式去求解谱减

系数, 那么会导致谱减系数的误差增大。为了减少计算 α 的误差, 本文在式 3.21 中加入了修改系数 θ , 得到公式如下

$$\left\{ \begin{array}{l} \alpha = (-T \frac{\alpha_{\max} - \alpha_{\min}}{T_{\max} - T_{\min}} + \frac{T_{\max} * \alpha_{\max} - T_{\min} * \alpha_{\min}}{T_{\max} - T_{\min}}) / \theta, \quad S(m, k) - T > \xi \\ \alpha = -T \frac{\alpha_{\max} - \alpha_{\min}}{T_{\max} - T_{\min}} + \frac{T_{\max} * \alpha_{\max} - T_{\min} * \alpha_{\min}}{T_{\max} - T_{\min}}, \quad \text{其它} \end{array} \right. \quad (3.22)$$

其中修改参数 θ 可按定义 $\theta = \frac{S(m, k) - T}{|S - T|_{\max}} + 1$ 来计算。本文根据听觉阈值小于真实值时听

觉阈值要比语音谱值小很多, 定义修改参数 θ , 使 θ 随着听觉阈值与语音谱值差值的增大而增大。在本实验中取 $\xi = 10$, 如果 $S(m, k) - T > \xi$, 即语音谱值与掩蔽阈值相差很大, 说明掩蔽阈值计算值偏小, 则导致式 3.4 中 α 的过大, 这样在谱相减的时候就过分衰减信号谱了。由于式 3.22 中 θ 和 α 成反比例, 可通过式 3.22 的修改系数 θ , 来抑制了 α 的增大。 $S(m, k) - T$ 越大, θ 越大, 而 θ 和 α 成反比关系, 这样就阻止了 α 的增大。通过 θ 可以抑制谱减的时候由于阈值曲线的误差导致信号的过分衰减, 降低了语音的失真。

3.4.2 语音谱平滑

然而增强的语音中还残留着一些音乐噪声, 语音还需要进一步进行处理。由于音乐噪声是频谱上残留的随机尖峰引起的, 音乐噪声可以用语音谱平滑的来抑制。平滑后可发现音乐噪声大幅度衰减了, 而在基于 MMSE 准则平滑后的语音没有大的变化, 几乎和原始语音相同。平滑语音谱公式如下

$$|\hat{S}(m, k)|^2 = \lambda(m, k) |\hat{S}(m-1, k)|^2 + [1 - \lambda(m, k)] |S(m, k)|^2 \quad (3.23)$$

$\lambda(m, k)$ 是平滑系数。增强的语音谱可以表示成纯净语音谱和误差谱 $R(m, k)$ 的和

$$|S(m, k)|^2 = |X(m, k)|^2 + R(m, k) \quad (3.24)$$

如果误差谱 $R(m, k)$ 小于 0, 则说明在 $S(m, k)$ 在式 3.4 中衰减的太大了, 如果大于 0, 则反之。式 3.23 的平滑可以使平滑语音谱 $\hat{S}(m, k)$ 接近纯净语音谱 $X(m, k)$, 而且消除了音乐噪声的影响。平滑系数 $\lambda(m, k)$ 根据 MMSE 准则来估计, 这样可以定义一个函数 $J_\alpha(m, k)$

$$J_\alpha(m, k) = E \left\{ \left[|\hat{S}(m, k)|^2 - |X(m, k)|^2 \right]^2 \left| \hat{S}(m-1, k) \right|^2 \right\} \quad (3.25)$$

把式 3.23 代入式 3.25 可以得到

$$J_{\alpha}(m, k) = E \left\{ \left[\lambda(m, k) \left| \hat{S}(m-1, k) \right|^2 + [1 - \lambda(m, k)] \left| S(m, k) \right|^2 - \left| X(m, k) \right|^2 \right]^2 \left| \hat{S}(m-1, k) \right|^2 \right\} \quad (3.26)$$

假设 $E[\left| X(m, k) \right|^2 R(m, k)] = 0$ ，用近似语音谱的数学期望代替纯净语音谱的数学期望

$$E[\left| S(m, k) \right|^2] = E[\left| X(m, k) \right|^2]。由于 E[\left| X(m, k) \right|^2] = \sigma_X^2(m, k)，E[\left| X(m, k) \right|^4] = 2\sigma_X^4(m, k)$$

把式 3.24 代入式 3.26，化简得到

$$J_{\alpha}(m, k) = \lambda^2(m, k) \left\{ \left[\left| \hat{S}(m-1, k) \right|^2 - \sigma_X^2(m, k) \right]^2 \right\} + \sigma_X^4(m, k) + \sigma_R^2(m, k) - 2\lambda(m, k) \sigma_R^2(m, k) + \sigma_R^2(m, k) \quad (3.27)$$

这里 $\sigma_R^2(m, k) = E[\left| R(m, k) \right|^2] = E_R(m, k)$ ，要求 $J_{\alpha}(m, k)$ 的最小值，对式 3.27 右边进行

$\lambda(m, k)$ 一阶求导，并令求导后的式等于零，得到 $\lambda(m, k)$ 的估计值

$$\lambda^*(m, k) = \frac{\sigma_R^2(m, k)}{\left[\left| \hat{S}(m-1, k) \right|^2 - \sigma_X^2(m, k) \right]^2 + \sigma_X^4 + \sigma_R^2(m, k)} \quad (3.28)$$

对分子分母同时除以 $\sigma_R^2(m, k)$ 得

$$\lambda^*(m, k) = \frac{1}{[\hat{\gamma}(m-1, k) - \hat{\gamma}(m, k)]^2 + \gamma^2(m, k) + 1} \quad (3.29)$$

这里 $\hat{\gamma}(m-1, k) = \frac{\left| \hat{S}(m-1, k) \right|^2}{\sigma_R(m, k)}$ ， $\gamma(m, k) = \frac{\sigma_X^2(m, k)}{\sigma_R(m, k)}$ 。 $\gamma(m, k)$ 为信号和谱失真比 (Signal to Spectral Deviation Ratio, SSDR)

从式 3.29 中可以看出， $\gamma(m, k)$ 很高时，则平滑系数

$\lambda(m, k)$ 趋于零，说明在此帧信号里失真小，音乐噪声很小，不需要进行语音平滑；如果

SSDR 低时，则 $\lambda(m, k)$ 将取较大值，这样残留的音乐噪声才能被平滑掉。根据最小均方误差准则估计的平滑系数使语音的残留音乐噪声和谱失真降低到最小值。

3.4.3 算法步骤

上面讨论了听觉掩蔽阈值计算误差产生原因，对这种方法进行了改进。通过上面的分析，下面是听觉掩蔽效应的语音增强改进算法的实现过程：

(1) 利用估计出的噪声来计算出近似的纯净语音谱 $S(m, k)$ ；

(2) 计算听觉掩蔽阈值 $T(m,k)$;

(3) 对原有的谱减系数 α 和 β 的计算公式进行改进, 利用改进的谱减系数计算公式

(3.22) 计算得到更为精确系数;

(4) 利用式 3.4 去除噪声, 实现语音增强, 得到增强后的语音谱;

(5) 用式 3.23 对增强后语音谱来实现基于 MMSE 准则下的语音谱平滑, 得到语音谱 $\hat{S}(m,k)$, 使音乐噪声完全抑制, 提高语音的听觉质量。

上述的是听觉掩蔽效应语音增强的改进算法的步骤分解, 通过上述的步骤实现了语音增强算法。

3.4.4 实验仿真和结果分析

根据本文的分析, 对所用的语音增强算法进行性能评测, 并和其他改进的谱减法进行了比较。由于人耳对语音的相位感知不敏感, 可以用含噪语音的相位来代替增强语音信号的相位, 根据增强后的语音幅度谱, 利用逆傅立叶变换就可以得到增强后的信号。在本文中, 纯净语音选取 863 中文语音识别语料库里面的一段语音, 噪声选取的是高斯白噪声, 语音数据的采样率为 8kHz, 帧长为 256, 窗口滑动为 128, 数据窗采用的是汉明窗。本文将语音频率段分为 22 个临界频带来计算掩蔽阈值。

首先对本文所用的方法计算出来的谱减系数和原来的方法进行比较。下图 3-5 给出了, 一帧信号谱减系数原始计算值和改进后的计算值比较图。

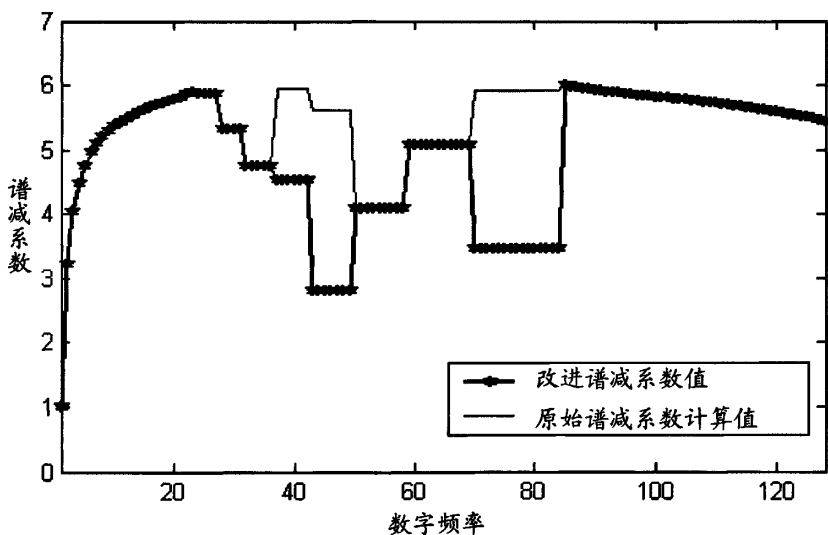


图 3-5 一帧信号谱减系数原始计算值和改进后的计算值比较图

Fig.3-5 The compare of the new and the old arithmetic

从图中可以看出在高频部分的谱减系数和原来没有变化, 变化的主要是中低频部分。因为在中低频部分的语音比较丰富, 变化较快, 这样就增大了噪声估计的误差, 引起谱减系数计算的误差。从图中可以看出本文的计算方法比原始的计算方法适应性提高

了。

再对原始听觉掩蔽效应语音增强算法和本文改进算法的语音谱进行比较，如图 3-6 所示。下图所示的是本文所用语音的第 210 帧，数字频谱 k 从 30 到 60 的纯净语音频谱，原始听觉掩蔽语音增强后的语音频谱和改进算法的语音频谱的局部放大比较图。

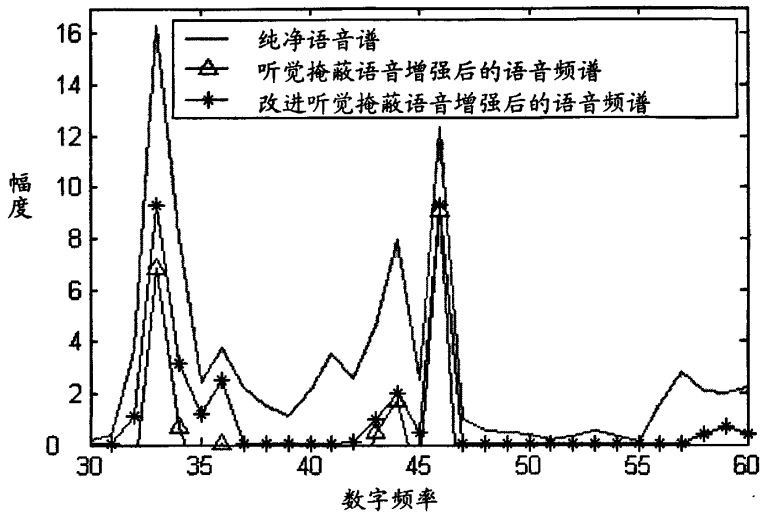


图 3-6 语音谱比较图
Fig.3-6 The compare of spectrum

从图中可以看出，本文所用算法的语音谱要更加接近于纯净语音谱，说明在谱减系数计算公式中加入了修改参数 θ ，对降低语音失真有明显的效果。

从客观测试指标来分析，采用了全局信噪比(SNR)和语音失真测度(IS)，SNR 和 IS 的主要区别是 SNR 不能反映人们对语音信号的听觉质量而 IS 语音失真测度在一定程度上可以反应人们对语音信号的主观感受，IS 系数越小，说明语音的品质越好。

文中对他们的信噪比和语音失真侧度值分别进行了比较，同时还对增强语音进行了非正式听力测试。 比较实验结果如表 3.3 所示。

表 3.3 算法的比较
Tab.3.3 The compare of the arithmetic

输入信噪比 (dB)		7.662	4.639	2.527
全局信噪比 (dB)	A	15.9	12.84	9.376
	B	12.95	9.327	7.439
	C	9.373	6.943	4.5
语音失真 测度 IS	A	0.823	1.762	2.359
	B	0.921	1.857	2.582
	C	3.135	2.892	2.593

A: 本文所用的算法; B: 听觉掩蔽效应算法; C: 传统的谱减法

从表 3.3 中可以看出，本文所用算法的信噪比有了很大的提高，主要原因是本算法

比其他算法的谱减系数适应性提高了很多,能更有效地去除噪声。从 IS 失真测度值上看,本算法的 IS 值要远远小于传统的谱减法,因为本算法利用听觉掩蔽阈值来调整谱减系数,根据听觉掩蔽原理,在听觉掩蔽阈值较大时,采用较小的谱减系数,这样可以保证有用的语音能充分的保留下来,保证语音的质量;掩蔽阈值较小时,采用较大的谱减系数,保证信噪比的提高。本算法和原始的听觉掩蔽法比较可以看出,本算法的 IS 值要比原始的听觉掩蔽法要小,这说明本算法的语音失真小了,这是本文对谱减系数的计算进行了改进,减少了语音失真。

最后对语音进行了主观测听,结果表明本文加入了语音平滑后,音乐噪声有了很大抑制,尤其在高信噪比的时候,音乐噪声几乎被完全消除了,而且听觉质量上和听觉掩蔽法没有什么改变。在低信噪比的情况下,其音乐噪声也得到了很好的抑制。本文针对以往增强算法中谱减系数计算局限性,提出了一种改进的听觉掩蔽效应语音增强算法,减少了语音失真,同时对语音谱进行了平滑处理,消除了语音谱上的残留尖峰,抑制了音乐噪声。客观性能测试和主观性能测试都表明该算法具有很强的优越性。

3.5 本章小结

本章主要分析了听觉掩蔽效应语音增强的改进算法。本章首先分析了噪声和语音的特性,重点分析了人耳的听觉掩蔽特性,对其产生的原因进行了分析。提出了两种改进的听觉掩蔽效应语音增强算法,一种是对噪声进行实时估计,提高噪声估计的准确性;一种是对在计算掩蔽阈值时产生的误差进行修正。针对这种误差,对原有的算法计算公式进行了改进,提出了一种改进算法。实验仿真结果表明语音的去噪效果优良,听觉感知舒适,算法性能指标表明了本章所用语音增强算法的优越性。

第四章 基于共振峰估计的宽动态压缩算法

宽动态压缩算法是数字助听器核心算法之一,其通过对语音的不同频域的频谱进行补偿进而达到听力补偿的目的,提高耳病患者对语音的感知。传统的宽动态压缩算法在进行频谱增益时忽略了语音共振峰的存在,引起了语音共振峰的非线性失真,降低了语音听觉舒适度。针对此不足点,本章提出了基于共振峰提取的宽动态压缩算法,在对语音进行宽动态压缩的时候,本算法充分考虑了语音共振峰对人耳感知的重要性,在同一个共振峰内使用相同的频谱增益值,使共振峰尽量保持不失真变形,降低了语音的失真度,提高了语音的辨别率和听觉舒适度。

4.1 共振峰提取

4.1.1 共振峰

在语音信号分析技术中,表征语音信号主要特征的两个最基本的语音参数是共振峰和基频。共振峰是反映声道谐振特性的主要特征,它代表了发音信息的最直接来源,而且人在语音感知中也利用了共振峰信息。

人类的声道和鼻道都可以看作是非均匀截面的声管,再发音时起共鸣器的作用,当准周期脉冲激励进入到声道时会引起共振特性,产生一组共振频率,称为共振峰频率或简称共振峰^[2]。共振峰参数包括共振峰频率和频带带宽,它们是区别不同韵母的重要参数。共振峰与发声器官的确切位置有很大的关系,即共振峰频率与声道的形状和大小有关,每种形状都有一套共振峰频率作为其特性。改变声道的形状就产生了不同的声音,因此,当声道形状改变时,语音的频谱特性也将作改变。共振峰频率由低到高排列依次为第一共振峰,第二共振峰,第三共振峰,……,相应的频率为 $F1$, $F2$, $F3$, ……表示。图 4-1 显示的是一段语音的频谱图,从频谱图中可以清楚的看到 4 个共振峰。

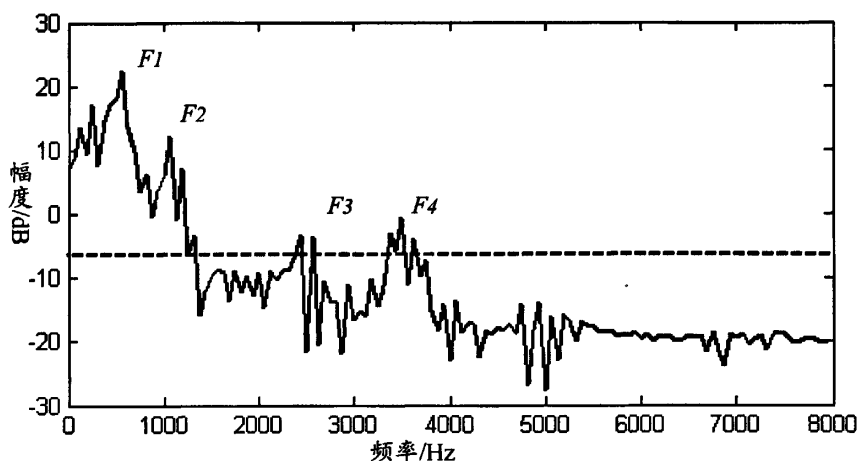


图 4-1 信号频谱图

Fig. 4-1 Signal frequency spectrum

语音合成研究表明一般的浊音中可以辨别的共振峰有 5 个，表示浊音信号最主要是前面三个共振峰。所以前面对 3 个区别不同语音及其重要，对语音信号进行处理的时候要充分保留这三个共振峰的特性，这样经过处理后的语音才能充分保留原语音的基本信息。本算法就是充分考虑到了共振峰对对耳感知的重要性，减少了频谱增益对原始信号共振峰的影响。

4.1.2 共振峰的提取

共振峰是一个最基本的表征语音信号特征的参数。在语音信号分析里面，分析知道共振峰信息包含在语音频谱包络中，因此共振峰参数提取的关键是估计自然语音的频谱包络，并认为谱包络中的最大值就是共振峰。目前针对共振峰的提取主要有带通滤波器法、基于功率谱的峰值提取法、倒谱法^[39]和线性预测法^[40] (Linear Prediction Code, LPC)。本算法针对共振峰的提取选用了 LPC 法。

从线性预测导出的声道滤波器是频谱包络估计器的最新形式，其提供了一个优良的声道模型。LPC 法对频谱包络低频部分能精确的提取共振峰，而本算法主要就是考虑语音信号低频部分的三个共振峰，所以此方法是本算法中共振峰提取最廉价最有效的一种方法。

线性预测分析的基本思想是用过去的 p 个样点来预测现在和未来的样值点

$$\hat{s}(n) = \sum_{i=1}^p a_i \cdot s(n-i) \quad (4.1)$$

预测误差为 $\varepsilon(n)$ 为

$$\varepsilon(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^p a_i \cdot s(n-i) \quad (4.2)$$

$\hat{s}(n)$ 为通过过去的 p 点预测的一个样点，通过某种准则下使预测误差 $\varepsilon(n)$ 达到最小值的方法来决定唯一的一组线性预测系数 $a_i (i=1,2,\dots,p)$ 。

求出一组预测系数后，就可以得到语音产生的全极点模型频率响应，形式如下

$$H(e^{j\omega}) = \frac{G}{1 - \sum_{i=1}^p a_i e^{j\omega i}} = \frac{G}{\sum_{i=0}^p a_i e^{j\omega i}} = \frac{G}{A(e^{j\omega})} \quad (4.3)$$

G 是增益因子，在实际的语音信号处理中最常用就是上述的全极点模型。这主要有两个方面的原因：(1) 如果不考虑鼻音和摩擦音，那么语音的声道的传递函数就是一个全极点模型；而对于鼻音和摩擦音，声学理论表明其声道的传递函数既有极点又有零点，但如果 p 阶数比较高的时可以用极点来代替零点。(2) 可以用线性预测分析的方法估计全极点模型参数。在共振峰频率上其频率响应特性会出现峰值，其频率响应 $H(e^{j\omega})$ 即

称为 LPC 谱。也是序列 $(1, a_1, a_2, \dots, a_p)$ 傅里叶变换的倒数。

LPC 谱的特点是在信号能量较大的区域内(与信号的低频部分对应)接近信号谱的峰值处, LPC 谱和信号谱接近;而在能量较低的区域即接近谱的谷底处,相差较大。而本算法主要考虑是低频部分的共振峰,用 LPC 法完全满足本算法共振峰提取的需要。 p 值选取的越大,其对语音估计就越精确,计算出来的信号包络就越接近原来信号的频谱包络,共振峰提取更加准确。但是这样会增加计算量和存储量,而且当 p 增加到一定的程度后,其性能的改善就不那么明显了,所以 p 的取值一般在 8~14 间,本算法选用的 p 值为 12。图 4-2 所示的是一帧语音的频谱图和 LPC 谱图

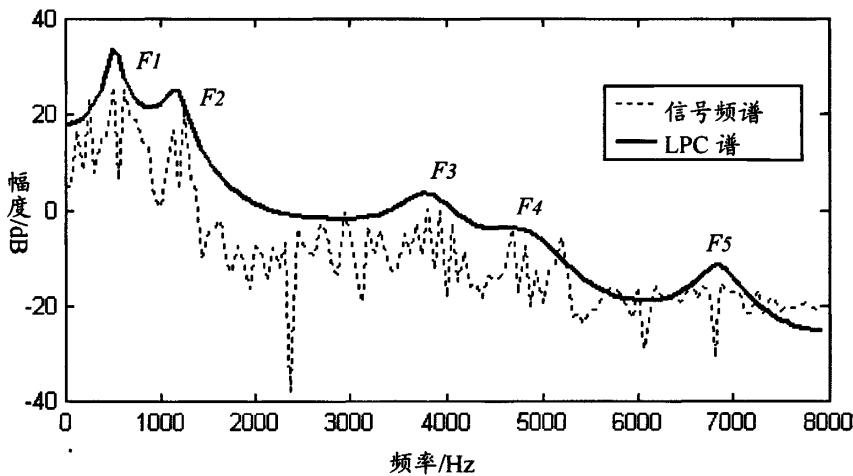


图 4-2 LPC 谱

Fig.4-2 LPC spectrum

图中的 LPC 谱可以看出其谱的走向和原信号谱几乎吻合,从 LPC 谱中可以很明显的看出共有 5 个峰值点,而这 5 个峰值点正好对应着信号频谱的 5 个共振峰点。原信号的频谱图中有很多的尖峰,不容易判断共振峰的频率点,而 LPC 谱图则去除了这些尖峰,只保留了共振峰点,可以很准确的找到共振峰。

在找共振峰点的时候,运用了一个比较简单的峰值检测算法。本算法在峰值检测时候,是利用 LPC 谱上的一频率点幅度值与前后频率点幅度值的差值正负符号的变化来判断是否是极值点,如果正负符号有变化,则为极值点,即所求的共振峰点。下图 4-3 是 LPC 共振峰提取算法的流程图

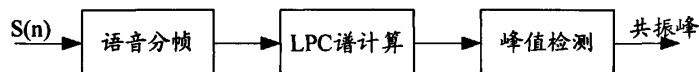


图 4-3 LPC 共振峰提取算法的流程图

Fig. 4-3 LPC resonance peak flow chart of the Algorithm

通过峰值检测算法提取到了一帧语音信号的峰值点,这些峰值点为此帧语音信号的共振峰。提取好的共振峰是为了对原始的宽动态压缩算法的增益值进行改进,降低语音的失真。

4.2 改进的宽动态压缩算法

4.2.1 宽动态压缩算法

在听力损失的情况下，人耳的听觉普遍下降，最终造成了听觉动态范围的减小，而且动态范围减小的程度和频率有关，一般耳病患者的高频听力损失要高于低频听力损失，这种不同频域间的听力损失影响了人耳的听力感知。宽动态压缩^[42-44]的目的就是将正常听力动态范围内的语音压缩到患者的听力动态范围里面。

宽动态压缩算法是数字助听器的核心算法之一^[45]，算法的最终目标是保证提高听觉的舒适度和语言的辨识率，所以必须不能破坏语言结构，不降低辨识率为前提。本算法在目前数字助听器通用算法的基础上，针对在进行频谱增益时引起的共振峰非线性失真提出了改进，把共振峰的存在考虑到算法中去，降低了语音的失真，提高了语音感知。

宽动态压缩算法是将整个语言动态范围按一定的比例均匀的压缩到患者的残余听力中。传统的多通道压缩算法在第二章中已有详细的介绍，多通道压缩放大算法就是利用滤波器组将信号分频段后再合并，声音信号可以被分为数个独立的频率区域，在各个不同的频域区域内，根据患者听力损失的情况，对不同频段加以不同的放大处理，对不同的频率成分运用不同的增益比，最终将合成的声音将会吻合患者的听力范围。但此类算法的不足点是在频段交叠区处理带来了新的问题，如当语音的共振峰正好处在频段的交叠处时，会被劈成两个部分，而他们将在不同的频段内进行放大，这会造成共振峰的移动、变形，甚至会消失^[48]。由上节的讨论可知共振峰的变动会影响人耳对语音的感知。针对这点不足，又有学者提出了宽动态压缩算法，把此类算法带回到频域里面进行处理，在频域对语音进行处理，减少了对语音共振峰的影响。

人的耳朵是一个生物滤波器，其作用可以当成一个线性时不变系统。设听力健康者的耳朵响应函数为 $C(j\omega)$ ，输入声音信号为 $S(j\omega)$ ，通过人耳到大脑后输出的信号为 $O(j\omega)$ ，耳病患者的响应函数为 $\tilde{C}(j\omega)$ ，助听器的频响函数为 $H(j\omega)$ 。

正常人的耳朵对声音的响应如下

$$O(j\omega) = S(j\omega) \cdot C(j\omega) \quad (4.4)$$

此时人的大脑对输入的声音响应 $O(j\omega)$ 能做出正确的判断，对语音有正确的理解感知。如果是耳病患者，其耳朵的频响函数是 $\tilde{C}(j\omega)$ ，则对输入的语音 $S(j\omega)$ 响应后其输出信号不等于 $O(j\omega)$ ，那么人的大脑对此信号则不能做出正确的理解，影响了人的语音感知。解决此问题就是在耳朵的响应函数前面级联一个助听器响应函数，使他们级联后对语音的响应能等于 $O(j\omega)$ ，响应公式如下

$$O(j\omega) = S(j\omega) \cdot H(j\omega) \cdot \tilde{C}(j\omega) \quad (4.5)$$

正常人耳的听力响应模型和耳病患者的听力响应模型如下图 4-4 所示

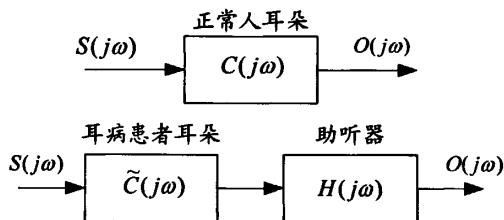


图 4-4 助听器和人耳的响应函数模型

Fig. 4-4 Ear hearing aids and the response function model

要使语音能很好的被耳病患者所理解，关键是要正确的计算出助听器的频响应函数 $H(j\omega)$ ，由式 4.4 和式 4.5 可知

$$H(j\omega) = C(j\omega) / \tilde{C}(j\omega) \quad (4.6)$$

求出正常人耳和耳病患者耳朵的频响应函数，就可以得到数字助听器的频响应函数 $H(j\omega)$ 。通过纯音测试，就可以得到正常人和患者的听力曲线图，从而得到正常人和患者的听力动态范围，进而得到正常人和患者的响应函数，用式 4.6 就可以计算出助听器频响应函数。

4.2.2 频域增益

频响应函数 $H(j\omega)$ 计算得到后实际上就是不同频点的增益补偿。根据患者的听力曲线可以计算得到整个频域的增益值，而整个频域的增益又按照输入信号的声压级实时的变化，即每一频点上面的频域增益，要根据输入信号的声压级进行调整。对于宽动态压缩算法，最终的目标是将声音均匀的压缩到患者的听力范围内，要保证输入声音的最适合声压级，提高听觉舒适度。图 4-5 是一个频点的输入输出曲线图。

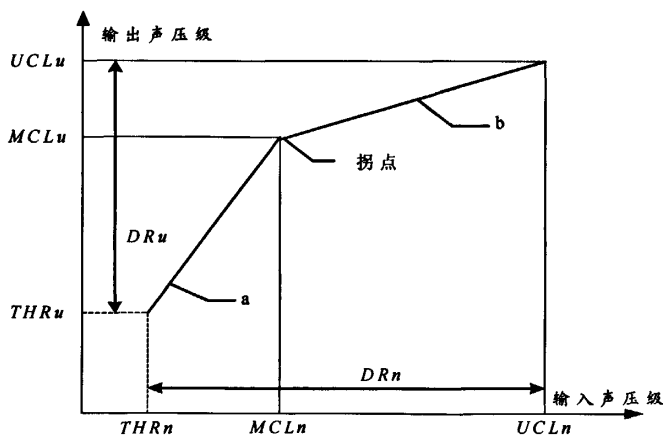


图 4-5 宽动态压缩输入输出曲线

Fig.4-5 I/O curve of WDRC

图中的 THR_n 和 THR_u 分别为正常人和耳病患者的听阈, MCL_n 和 MCL_u 分别是正常人和耳病患者的听觉最适级, UCL_n 和 UCL_u 正常人和耳病患者的痛阈。通过测得的正常人和耳病患者的听力动态范围, 计算出频域增益 $CR(f)$ 。图中增加的一个拐点的目的是使一个频点的不同声压级的语音能更好的吻合患者的听力范围, 声压级大的增益较小一点, 上半增益为 CR_b ; 而声压级小的则增益大点, 下半增益为 CR_a , 使语音能变成最佳的声压级输出。

增益计算公式如下

$$CR_a = \frac{MCL_n - THR_n}{MCL_u - THR_u} \quad (4.7)$$

$$CR_b = \frac{UCL_n - MCL_n}{UCL_u - MCL_u} \quad (4.8)$$

对于听阈的测听, 不可能对频域上的每个频点进行测听, 只是对特征频点进行测听。测定的特征频点一般为 250Hz、500Hz、1000Hz、2000Hz、4000Hz 和 8000Hz。通过测得的特征频域点的听阈值、最适阈值和痛阈值, 就可以求特征点的频域增益, 对于频率点以外的频点对其进行插值, 进而得到每点的增益。下图 4-6 是根据一个中耳炎患者的听力曲线值计算得到了每一点频率所对应的增益 $CR(f)$ 。

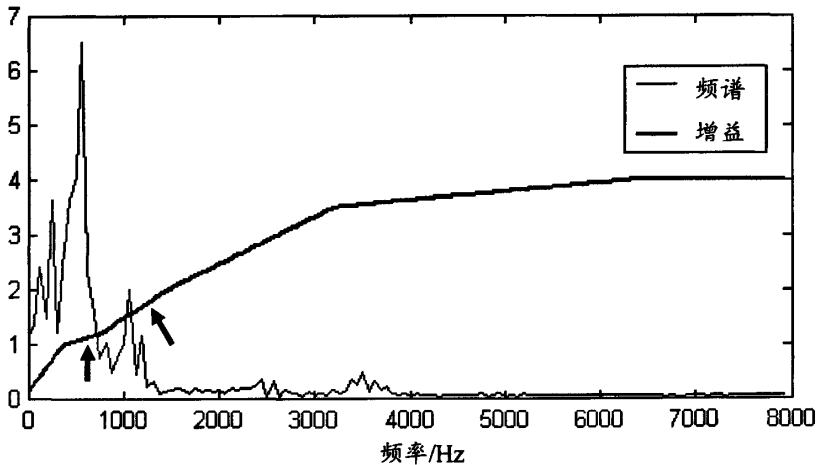


图 4-6 频谱增益图

Fig. 4-6 Enrichment plans of spectrum

从图中可以看出, 高频部分的增益要大于低频部分, 而且增益不是一直不变的. 由于增益是变化的, 那么在同一个共振峰中, 共振峰的增益也是非线性变化的, 就引起了共振峰的变形。可以从图中黑色箭头指示的增益部分可以看到, 在这些增益频域里面, 有语音的共振峰出现了, 而在语音共振峰出现的频段里面, 增益是变化的。在同一个共振峰频段间, 增益的变化, 即直接导致在此频段里面进行增益后, 共振峰变形非线性变形了。由前面的讨论知道, 共振峰是语音感知的重要参数。共振峰非线性变形了, 人耳对语音的感知同样也下降了。

本算法主要针对共振峰非线性变形提出改进, 降低了共振峰的非线性变形。主要思

想是在同一个共振峰频段里面，使用同一个频谱增益，使共振峰能线性变化，减少语音的失真。具体的算法是如下：

首先计算得到整个频谱的原始增益值 $CR(f)$ ；在整个频段里面寻找共振峰，如果有共振峰，则求出此共振峰频段增益的最小值 $CR(f_1)$ ；把此共振峰频段的增益全部用代替 $CR(f_1)$ ，这样可以保证在同一个共振峰中增益是相同的，消除了共振峰的非线性变化，最终得到了新的频谱增益值 $\overline{CR}(f)$ 。

由语音信号处理理论和人耳感知原理可知，对语音的辨识，起主要作用的是前面三个共振峰，所以在对增益进行改进的时候主要的是考虑前面的三个共振峰。通过上面一节运用 LPC 法提取的共振峰参数，选取前面三个共振峰 $F1$ 、 $F2$ 和 $F3$ 。对增益改变就此三个共振峰的频段里面进行调整。

4.3 算法的步骤

通过上面的分析，对宽动态压缩算法进行了改进，充分考虑了语音中的共振峰信息。下面是本章所提出的基于共振峰估计的宽动态压缩算法的实现过程：

- (1) 根据听力曲线，运用式 4.7 和式 4.8 计算得到原始的频谱增益 $CR(f)$ ；
- (2) 运用 LPC 法计算每帧语音的频谱包络；
- (3) 运用峰值检测算法，找到语音的前三个共振峰；
- (4) 在同一个共振峰频段内，把原来的频谱增益全部用此频段内的增益最小值代替，得到新的增益值 $\overline{CR}(f)$ ；

- (5) 对输入语音在频域里面进行频谱增益，得到了宽动态压缩后的语音信号。

上述的是本章基于共振峰估计的宽动态压缩算法的步骤分解。通过对原始频谱增益的改进得到了新的频谱，改进后的增益充分考虑了语音共振峰的信息，最大限度的限制了语音的失真。

4.4 实验仿真和结果分析

根据分析，对本章的基于共振峰估计的宽动态压缩算法进行性能评测，并和原始的宽动态压缩算法进行了比较。在本文中，语音选取 863 中文语音识别语料库里面的语音，语音数据的采样率为 16kHz，帧长为 256，窗口滑动为 128，数据窗采用的是汉明窗。下面所用的听力曲线实验数据是利用一个典型的中耳炎患者的听力曲线值，根据听力曲线值计算得到了每一点频率所对应的增益。

首先对本算法所计算出来的增益值和原始宽动态压缩计算出来的增益值进行比较。下图 4-7 是在一帧语音信号中，用本算法和原始算法计算出来的增益值比较。

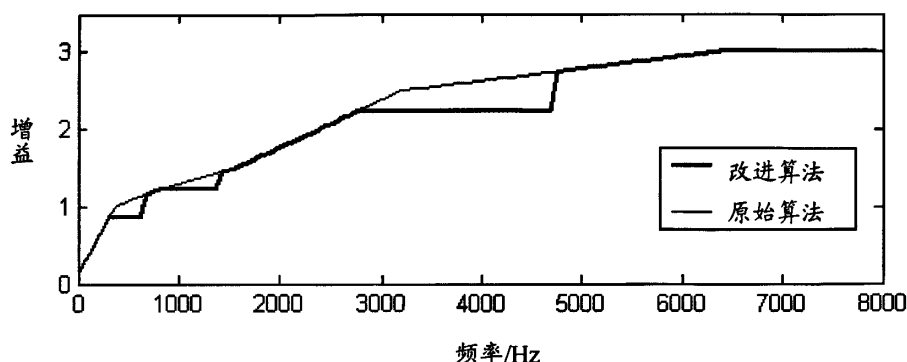


图 4-7 改进的和原始的宽动态压缩算法增益值比较图

Fig. 4-7 Wide dynamic compression algorithm gain value comparison chart

从图 4-7 中可以看出在一帧语音信号中, 改进后的宽动态压缩算法计算出来的增益值相对于原始算法计算出来的增益值有三处明显的直线段。这主要是本算法考虑到了共振峰的问题, 在同一个共振峰内使用同一个频谱增益值, 而且本算法重点考虑了对语音理解有重要影响的第一第二第三共振峰, 所以有三个直线段。用统一的增益值代替原来的变化的增益值, 可以减少语音共振峰的非线性失真。增益在高频部分的值要大于低频部分, 这主要是神经型听力损失患者的高频听力损失要高于低频, 在进行增益补偿的时候, 高频部分的增益要比低频部分大, 这样听者才能更好的感知高频部分的语音, 提高语音的感知舒适度和理解率。下图是 4-8 是一帧语音输入输出的频谱比较。

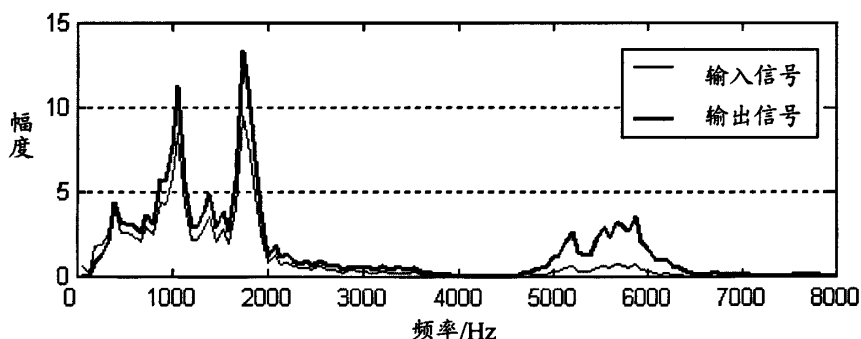


图 4-8 输入输出信号频谱比较

Fig. 4-8 Input and output signals spectrum comparison

图 4-8 中可以很明显的看到高频部分的增益要大于低频部分, 通过人耳测听, 输出语音变的响亮了。这主要是高频部分的语音得到了比较大的增益, 人耳才感知到声音的变的响亮。

下面通过客观测试指标 Itakura-Saito 语音失真测度(IS)来比较改进后的宽动态压缩算法和原始算法的性能。IS 语音失真测度在一定程度上可以反应人们对语音信号的主观感受, IS 系数越小, 说明语音的品质越好。实验时用了中文语音库中的三段语音, 分别用两种不同的算法进行实验比较, 实验结果如表 4.1 所示

表 4.1 算法比较

Tab.4.1 Comparison algorithm

输入语音		语音 1	语音 2	语音 3
语音失真测度	A	4.25	5.64	3.24
(IS)	B	3.94	4.57	2.73

A: 原始的宽动态压缩算法 B: 改进的算法

从表 4.1 中可以看出改进的算法对不同的输入语音处理后, 其语音失真测度值要比原来的宽动态压缩算法的语音失真测度值要小, 这表明本算法对语音的失真有了限制。这主要是本文在对语音进行频谱增益时, 充分考虑的到了语音共振峰的存在, 在同一个共振峰内使用统一的增益值, 降低了共振峰的非线性失真, 提高了语音的听觉舒适度。

4.5 本章小结

本章主要研究了基于共振峰估计的宽动态压缩算法。针对原始的宽动态压缩算法在语音进行频谱增益的时候, 忽略共振峰存在的问题, 在进行增益时, 引起了共振峰的非线性失真, 降低了语音的感知舒适度。本算法针对此问题对原算法进行了改进, 首先用 LPC 法估计语音的共振峰, 在同一个共振峰内选用同一个增益值, 减少了语音共振峰的非线性失真。实验表明, 此算法语音失真测度值要小于传统算法, 降低了语音失真。在实验中唯一不足的就是由于条件的限制, 没有让耳病患者真正试听一下利用本算法处理后的语音, 检验本算法在实践中的应用效果, 但是就从客观性能分析, 本算法要比原算法优越。

第五章 数字助听器算法的验证

数字助听器的核心算法主要是语音增强、宽动态压缩和移频压缩三种语音处理算法。前面三章是对三种算法进行理论上的研究,本章主要是对这三种算法进行实际验证。在算法的验证过程中,选用了合众达的 SEED-DEC5416 开发板,在 CCS 软件开发平台上编写三个语音处理程序,最终把编译好的程序代码下载到开发平台上运行。通过 PC 机上的语音输出接口输出标准库中的语音信号到开发板上的语音输入端口,语音处理后通过开发板上的语音输出口把处理好的语音输出,并通过扬声器播发出来。实践证明本章编写的算法能够实现数字助听器语音处理预期的效果。

5.1 硬件平台 SEED—DEC5416

合众达的 SEED-DEC5416 硬件开发平台主要集成了 DSP、SRAM、FLASH、CODEC、UART、USB 等外设以及开放给用户的扩展总线。这样的硬件设计可以使它应用在语音处理及其相关的领域里面。主要硬件组成如下:采用 TMS320VC5416@160MHz^[51]能实现高速运算;外扩 SRAM,最大容量为 512k×16 位,基本配置为 64k×16 位;外扩 Flash,最大容量为 1M×16 位,基本配置为 256×16 位;AC97 标准的 Audio 音频接口,双声道、立体声输入/输出,最高采样频率 96kHz;2 路 UART 接口;标准化扩展总线。

54 系列 DSP 具有很高的编程灵活性和运行速度。它具有先进的哈佛结构、专门硬件逻辑的 CPU、片内存储器、片内外设和专用的指令集、将 C54DSP 的 CPU 和片内存储器与外设配置组合在一起的螺旋结构,使得它可以满足电子众多领域的应用要求。其增强型的哈佛结构具有一条程序总线、三条数据总线和四条地址总线,这样的设计可以将程序和数据存储在不同的存储空间中,即程序存储器和数据存储器是两个独立的存储器,每个存储器可以独立的编程、独立的访问。与两个存储器相对应的是系统中设置了程序总线和数据总线,从而使数据的吞吐率提高了一倍,加上其的流水线操作大大提高了数据处理的能力。

音频的采集选用了低功率 Σ - Δ 型的 16 位 A/D、D/A 音频接口芯片 TLV320AIC23B^[52]。TLV320AIC23 是 TI 公司推出的一款高性能的立体声 Codec 芯片,内置耳机输出放大器,输入输出的增益可以通过内部寄存器编程来调节,内部高度集成了 ADC 和 DAC,采样频率可以从 8K 到 96K 编程来实现,其 ADC 和 DAC 的信噪比可以分别达到 90dB 和 100dB,同时还具有低功耗的优点。它和 TMS320C54xDSP 的 I/O 电压兼容,可以实现串口的无缝连接。

C5416 通过 McBSP 来设置 AIC23 的内部寄存器时,每个字的前 7 位为寄存器的地址,后 9 位为寄存器的内容。

AIC23 的控制寄存器如下表 5.1 所示:

表 5.1 AIC23 的控制寄存器

Tab. 5.1 The control register of AIC23

地址	需要初始化的寄存器
0000000	左声道的音量控制寄存器
0000001	右声道的音量控制寄存器
0000010	左声道耳机的音量控制寄存器
0000011	右声道耳机的音量控制寄存器
0000100	模拟音频的路径控制寄存器
0000101	数字音频的路径控制寄存器
0000110	省电方式控制寄存器
0000111	数字音频的接口格式寄存器
0001000	采样率的设置寄存器
0001001	数字接口设置寄存器
0001111	复位寄存器

当 BFSX2 发出使能信号控制 AIC23, 通过访问上面的各个寄存器来设置 AIC23 的工作状态。SDIN 接收一个 16 位的字, 前 7 位为所要访问的寄存器地址, 后 9 位为要写入各个寄存器的数值。

5.2 软件整体设计

本程序主要分三部分: 开发板初始化, 语音输入输出和语音处理算法。

软件设计和编程主要是使用 TI 公司推出的开集成发环境 (Code Composer Studio, CCS)。TI 公司为 DSP 开发者提供了一套 DSP 集成开发环境。CCS^[53]提供了软件开发、程序调试和系统仿真环境, 可以帮助用户通过 CCS 平台对硬件进行调试、程序仿真和开发, 可以大大提高编程的效率。CCS 不但能支持汇编软件的开发, 而且还支持 C/C++ 语言进行软件开发。CCS 支持 C 语言和汇编混合编程, 采用这种方法, 既可缩短软件开发的周期, 提高程序的可读性和可移植性, 又能满足系统实时运算的要求。

DSP 软件开发流程主要分为以下几个步骤:

(1) 首先编写汇编源程序, 按照 DSP 规定的汇编语言指令来编写源代码, 也可用 C/C++ 来编写源程序, CCS 支持汇编语言和 C/C++ 语言混编。

(2) 写好的 C/C++ 源程序通过 CCS 的 C/C++ 编译器编译成汇编源程序, 汇编的效率可以通过优化功能来提高。

(3) 生成的汇编源程序通过调用编译器生成一个共用目标文件格式 COFF(Common Object File Format, COFF)的目标文件。

(4) 通过调用连接器把生成的全部目标文件集合在一起, 生成可执行的目标代码。

(5) 生成的目标代码通过下载器或者十六进制转换程序, 把代码装载入 DSP 系统的存储器中去, 执行编写的程序。

软件整体结构根据其功能可以分为语音算法^{[54][55]}和数据采集^[56]两个大部分, 图 5-1 所示的是程序设计流程图。本文实现的是数字助听器的语音处理算法, 程序主要分为如

下几个部分：(1) DSP 和 AIC23 的初始化，设置 DSP 的工作时钟，初始化 McBSP0 和 McBSP2 的工作状态，打开 AIC23 的数据接口，通过 McBSP2 来设置 AIC23 的各个工作参数。(2) 初始化好系统后，系统进入等待状态，判断数据是否有输入，如有数据输入，则把数据进入到信号处理程序。(3) 把收到的数据先通过 FFT 变换，变换到频域里面进行语音算法处理。(4) 语音数据输出。

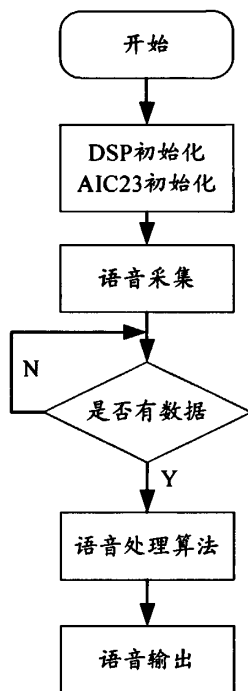


图 5-1 程序设计流程图

Fig.5-1 Flowchart design procedures

5.3 信号输入输出设计

在设计信号输入输出时主要考虑了 C5416DSP 的工作频率，McBSP 的工作模式和 AIC23 的寄存器设置。C5416DSP 工作频率设置为 160MHz。McBSP0 用于与 AIC23 的数据接口，用来实现 A/D 数据的输入 D/A 数据的输出。这里 McBSP0 应该设置为从模式，用来与 AIC23 进行通讯。同样对 McBSP2 进行初始化工作状态，McBSP2 是用来配置 AIC23 的寄存器，控制 AIC23 的工作状态。DSP 通过 McBSP2 把要设置的值写入到 AIC23 的寄存器中，所以 McBSP2 应设置为主模式，来控制 AIC23。

AIC23 可以通过表 5.1 所列的 11 个控制寄存器来设置其工作状态。每个控制字为 16 位，前 7 位寄存器的地址，后 9 位写入寄存器的内容。通过设置寄存器的值可以控制左右声道的输出输入增益，改变 A/D 和 D/A 的采样率。在本文中设置采样率为 8kHz，符合 CCITT(国际电报电话咨询委员会)提出的标准建议采样频率 8kHz。AIC23 的采样频率控制寄存器地址 0001000B，表 5.2 所示的是采样频率控制寄存器的 9 位控制字。

表 5.2 采样频率控制寄存器

Tab.5.2 Sample rate control registers

位	D8	D7	D6	D5	D4	D3	D2	D1	D0
功能	×	CLKOUT	CLKIN	SR3	SR2	SR1	SR0	BOSR	USB/Normal
默认	0	0	0	1	0	0	0	0	0

CLKIN 时钟输入分频 0=MCLK 1=MCLK/2

CLKOUT 时钟输出分频 0=MCLK 1=MCLK/2

SR[3:0] 采样频率控制

BOSR 基过采样频率

USB 模式 0=250fs 1=272fs

标准模式 0=256fs 1=384fs

本文选择 AIC23 的工作模式为 USB 工作模式，查 AIC23 的工作手册可知，A/D 和 D/A 工作频率都设为 8kHz 时，9 位控制字为 000001101B，写入采样频率控制寄存器的值为 0xd。AIC23 有很强的编程可控制能力，从表 5.1 中可以看到其左右声道音量和输入输出增益都可通过编程来改变，如需具体的设置，可以查询其工作手册，本文不在做详细的介绍。关于 DSP 其他部分的初始化都采用了默认的设置值。

5.4 算法的实现

5.4.1 语音分帧处理

通过上面的各个部分的初始化，系统已经进入到就绪状态，等待语音信号的输入。由于使用的 McBSP0 作为数据通讯口，语音数据进入到 DSP 中是以串行的形式进入。McBSP0 每次只读一个 16 位的语音数据和语音处理算法建立在语音的短时平稳性上，要对读进的语音数据进行算法处理，需要对语音进行分帧处理。这里把连续的 256 个语音数据作为一帧信号，然后对一帧信号进行处理。

本文在程序设计时，开辟了缓冲区域 buffer 和 input，这里 buffer 共分为两个部分第一缓冲区 buffer1 和第二缓冲区 buffer2，input 和 buffer 大小设置为 512。图 5-3 是其缓冲区工作的原理图。

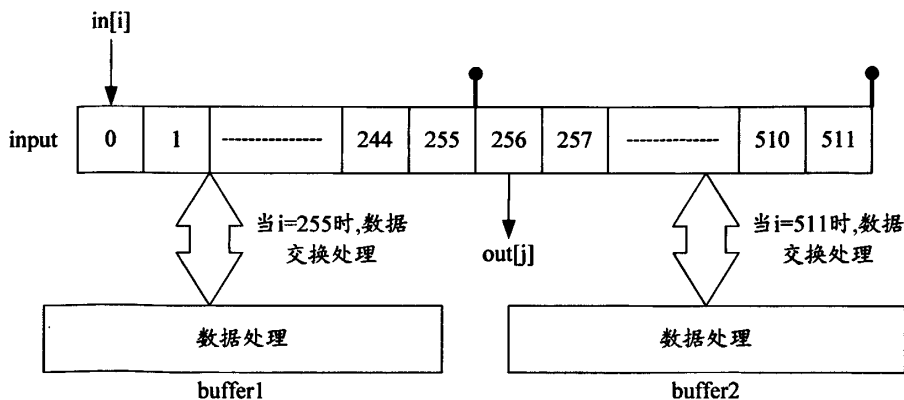


图 5-3 缓冲区工作原理图

Fig.5-3 Buffer schematic work

图中 input 是一个大小为 512 的环形缓冲区，当输入数据 in 输入数据到 input 的最后一位后，就重新回到 input 的起始位置，输出数据 out 延迟输入数据 256 个，它和 in 一样也是一个循环操作。

当 in 的数据写到 input 的 255 位时，就进入到处理程序中来。首先把 input 中的前 256 个数据搬移到 buffer1 中，对 buffer1 中的 256 个数据进行数据处理，处理好的 256 个数据再放回到 input 的前 256 位中，等待 out 的读出。

当 in 的数据在写 input 的后 256 位时，out 同时进行着把刚刚处理好的前 256 个数据依次读出。当 in 的数据写到 input 的 511 位时，把 input 的后 256 个数据搬移到 buffer2 中进行数据处理，处理好的 256 个数据再放回到 input 的后 256 位中，等待 out 的读出。in 写到 input 的 511 位时，out 已把前 256 个数据读出了，准备读下面的数据。

最后 in 再由 input 的 511 位回到 0 位继续向 input 中写数据，out 开始读下面 256 个后处理好的数据，这样不断的循环下去，这就是缓冲区的工作原理。

为了避免分帧带来的截断效应，需要对语音进行加窗处理，应用在谱减法中的各种各样的窗函数有着不同的性能特性，综合考虑选用 Hamming 窗对语音进行加窗处理。

5.4.2 算法验证

本章在算法验证的过程中使用了 CCS 强大的图形显示功能，对输入输出数据进行了分析比较。本文在进行数据处理时，要把时域的离散信号变换到频域里面进行处理。对数据进行离散傅立叶变换，本文 FFT 运算的最小阶数选取和语音帧的长度相同。FFT 和 IFFT 的运算量占整个语音运算量中的 70%~80%，因此 FFT 和 IFFT 算法的性能直接决定整个语音算法能否实时处理。FFT 的运算周期随着其阶数的增大而迅速变长，运算占用的系统空间迅速增加，综合考虑保证系统实时性本文采用了 256 点的 FFT。

在进行 FFT 和 IFFT 运算的时候选择了 DSP 库函数中的 RFFT 和 RIFFT 实傅里叶变换函数，DSP 库函数采用的是汇编语言来编写，每层计算的时候使用了正弦表来查询，虽然空间有所增加，但是正弦值用查表比用公式计算要快的多，这里用空间来换取时间是值得的。DSP 库函数都进行了精简优化，使代码最精简，运算速度力求最快。

本文先采用了对正弦信号进行 RFFT 变换为频域信号，然后再通过 RIFFT 变换到时域，下图 5-4 是输入输出信号的时频域图。

从图 5-4A 和 5-4C 可以看出输入信号经过 RFFT 和 RIFFT 变换后其波形几乎相同，只是在幅度上有了变化，这主要是因为 RFFT 和 RIFFT 在变换的过程中对输入输出数据进行了防溢出处理，如果想把输出波形和输入完全一样，可以在输出波形上乘以一个适当的增益，就可以完整的恢复原始信号。

图 5-4B 是输入信号的频域图，在低频附近有一个明显的峰值，这和经过 RFFT 变换后的输入信号频谱图 5-4D 几乎完全一样。这说明在 DSP 上实现的 RFFT 和 RIFFT 变换是完全有效的。

在频域里面对信号进行处理，这时语音算法用 C 语言来编写，提高算法的可读性。

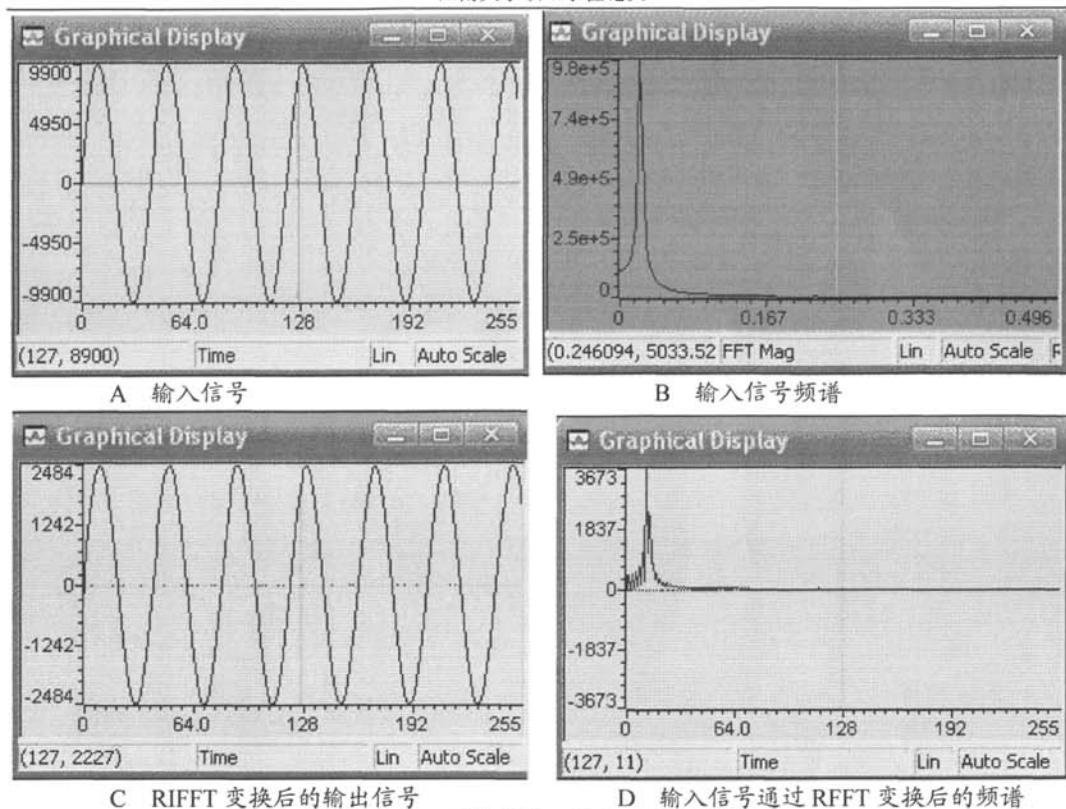


图 5-4 正弦信号输入输出的时频域图

Fig.5-4 The input and output sina signals in the frequency domain plans

用 C 语言编写的算法有语音增强、宽动态压缩和移频压缩等, 这些算法首先用 MATLAB 语言编写, 在 PC 机上进行仿真, 验证算法的可行性。这些算法在 PC 机上仿真通过后, 在 CCS 开发平台上用 C 语言编写后加入到主函数中去。

在语音进行算法处理时涉及到一个实时的问题: 就是一帧语音处理算法运行的时间是否小于 A/D 的转换时间, 如果大于 A/D 转换的时间那么就会出现一帧 256 个数据不能全部读出, 出现时延现象。下面来讨论一下 DSP 运行一帧语音所需的时间。

TMS320VC5416DSP 的运行速度达到 160MHz, 本文算法主要分为噪声估计, 掩蔽阈值计算, 谱减系数计算和 FFT 运算等几个部分, 由于 DSP 其特殊的结构和专用的指令, 能够高速的实现 FFT 和 IFFT, 每帧数据处理运行的时间大约为:

$$T_{dsp} \approx \frac{20 \times 256}{1.6 \times 10^8} = 0.032ms$$

AIC23 采样频率为 8kHz, 完成一次采样所需要的时间为:

$$T_{A/D} = \frac{1}{8000} = 0.125ms$$

从上面计算的 T_{dsp} 和 $T_{A/D}$ 可以看出, DSP 每处理一帧的数据所需的时间要远远小于 A/D 采样一次的时间, 即在 A/D 采集一次数据时 DSP 就能够把一帧的语音数据处理完毕, 这样能够保证系统的实时性。

(1) 语音增强算法的验证

下面把听觉掩蔽效应语音增强的改进算法在 DSP 平台上实现。此算法先是在 MATLAB 上仿真通过, 验证了此算法的可行性, 再用 C 语言对 MATLAB 源程序进行改写, 把算法源程序改写为 CCS 开发平台上能运行的 C 源程序。

在 DSP 平台上实现语音增强算法, 采样率为 8kHz, 语音信号每帧为 256 点。语音的输入选用了 863 中文语音识别语料库里面的一段标准语音。语音通过 PC 机上的播放软件播放, 然后声音通过 PC 上耳机输出接口输出到开发板上的麦克风输入接口实现语音标准输入。把输入的一段语音选取其中两帧进行分析, 先把原始标准语音在主程序中加入噪声后存储起来变成含噪语音, 然后把处理后的去噪语音同样保存用以观察处理后的效果, 下面是语音的分析。

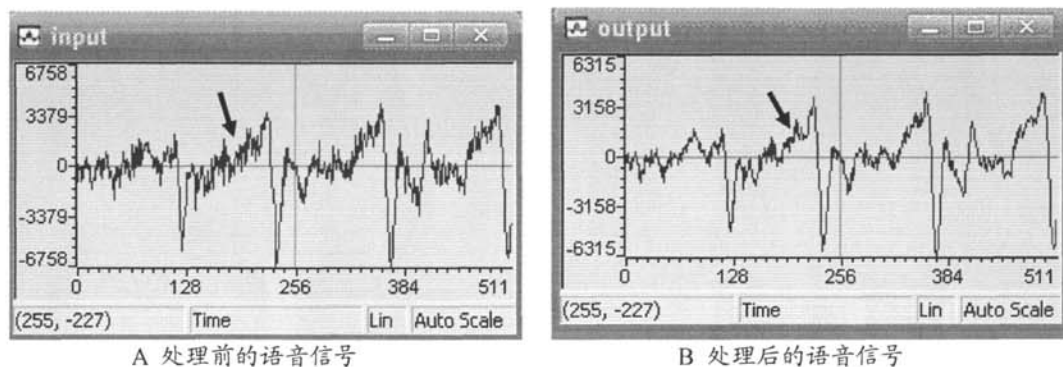


图 5-5 语音增强前后的语音比较图

Fig.5-5 Comparison chart in time domain of speech enhancement

图 5-5 是一段语音经过语音增强算法前后的时域比较图。图中只比较了两帧语音信号的时域波形, 从图 5-5A 中 (黑色箭头指示) 可以看出处理前时域的语音信号中含有很多的噪声毛刺, 这些噪声毛刺的存在影响了人的听觉感知。图 5-5B 是处理过后的语音信号, 图中 (黑色箭头指示) 可以很明显的看出, 噪声毛刺少了很多, 语音平滑了, 噪声得到了很好的抑制。把处理前和处理后的这段语音数据通过 CCS 上的 data/save 数据保存功能保存为数据文件, 再把这段数据调用到 MATLAB 里面, 利用 MATLAB 强大的数据计算功能计算输入输出信噪比和语音失真测度值。图 5-6 是本文算法的性能比较图, 分别比较不同输入信噪比的情况下输出信噪比和语音失真测度值。

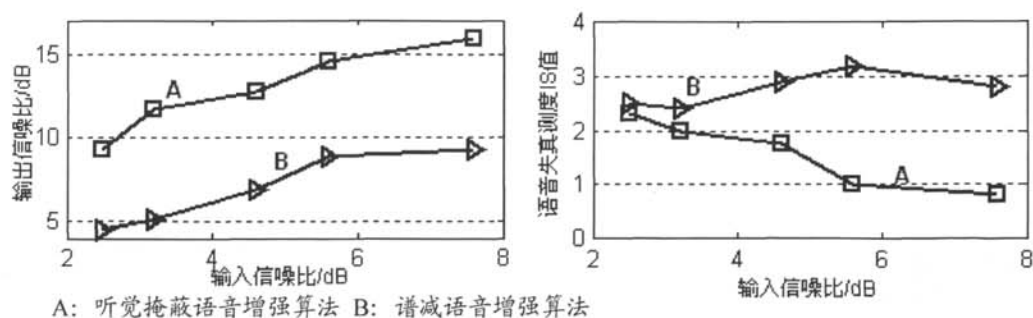


图 5-6 算法性能比较图

Fig. 5-6 Algorithm performance comparison chart

图 5-6 中可以看出听觉掩蔽语音增强算法对语音进行处理后, 信噪比有了很大的提高, 且语音失真比谱减法要小的多。通过耳机对输出语音进行测听, 语音的噪声明显的降低, 语音清晰度提高, 由于算法中利用了人耳的听觉掩蔽效应, 去噪后的语音中残留音乐噪声几乎消失了, 听觉质量有了明显的提高。

(2) 宽动态压缩算法的验证

把在 MATLAB 上仿真通过的宽动态压缩算法源程序加载到 DSP 开发板上, 对算法进行验证。利用 PC 机上的标准语音输出到开发板上的语音输入端口。利用第二章中提到的患者听力曲线来确定每点的频域增益。把纯净的语音和处理过后的语音分别存储起来, 通过 CCS 的图形显示功能把输入输出语音进行显示。采集的一帧语音数据经过 DSP 进行信号处理, 处理前后的结果比较如下图 5-7 所示。

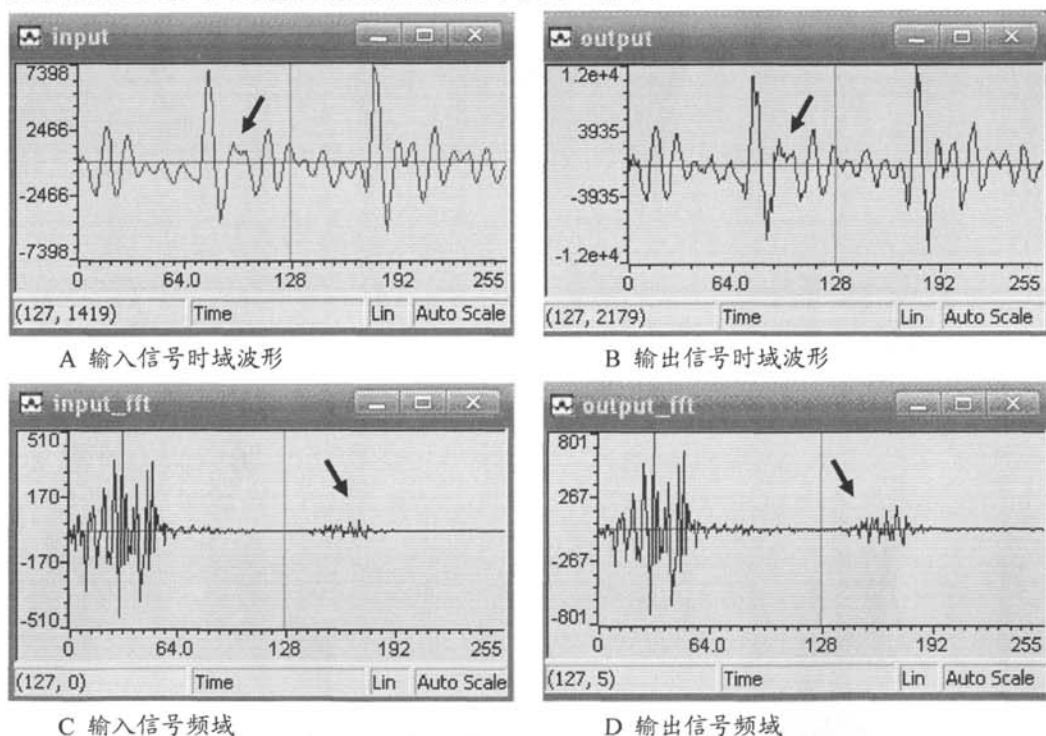


图 5-7 宽动态压缩在 DSP 中实现的结果比较

Fig. 5-7 Wide dynamic compression in DSP achieved results

从图 5-7A 和图 5-7B 一帧信号的时域图中可以看出信号在各点的能量都有所加强, 信号的细节变化 (黑色箭头指示) 比原信号要多了, 这表明信号的高频部分能量增强了。通过耳机的输出对输出信号通过测听, 声音比原来要响亮, 音调提高了。图 5-7C 和 5-7D 是信号的频域图, 从图 5-7D 中可以很明显的看出信号的高低频部分幅度都有所加强, 但是在高频部分的增益要比低频部分要大的多 (黑色箭头指示), 这也证明了先前说高频部分能量增强, 音调变高。在本实验验证的过程中, 利用 PC 机播放不同的音频, 通过测听音频的音调都有的提高, 从而证实了本算法的通用性。

(3) 移频压缩算法的验证

本实验同样是利用标准语音库里面的语音作为标准语音输入。在 DSP 平台上对语音进行压缩比率为 2 的移频压缩, 将输入的语音信号由原来的 0~8kHz 频域段按比率压缩到 0~4kHz 的频域段。同样对 256 点的一帧语音信号的输入输出进行分析, 保存输入语音和处理后的语音利用 CCS 的图形显示功能把保存的语音显示出来, 结果比较如下图 5-8 所示。

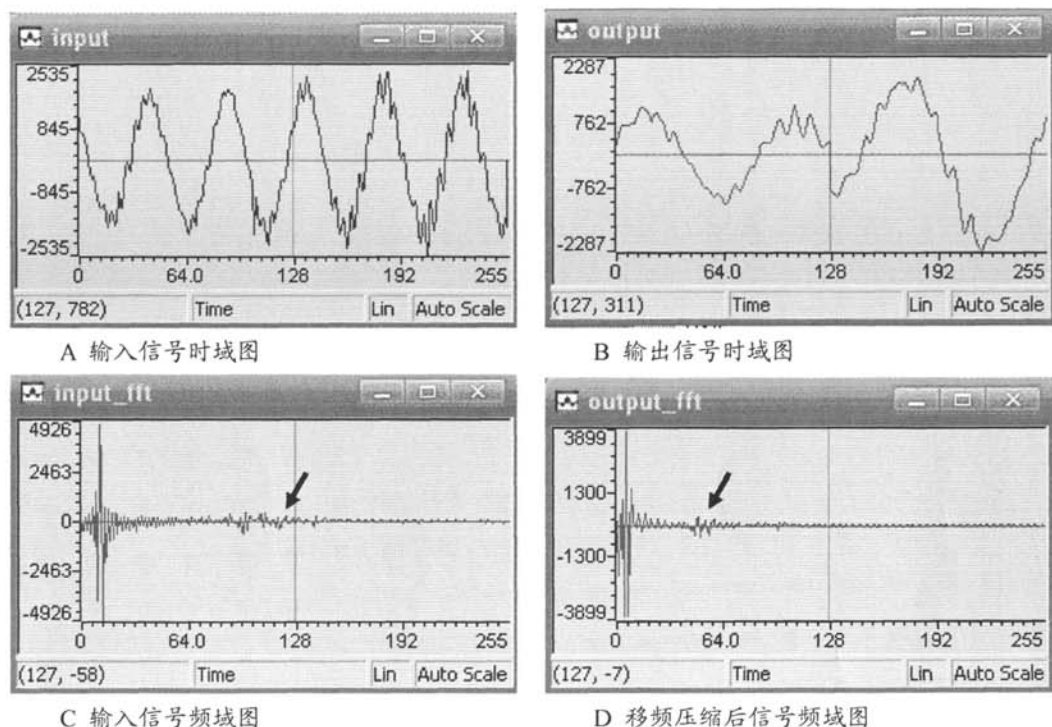


图 5-8 移频压缩在 DSP 中的实现的结果比较

Fig. 5-8 Shift frequency compression on DSP in the realization of the results

图 5-8A 和 5-8B 是移频前后的一帧信号时域比较, 从图中可以看出信号的变化减慢了, 信号被拉伸了, 语音频率降低; 图 5-8C 和 5-8D 是它们的频域图, 从频域图中可以很明显的看出频域被压缩了, 原来在频点 3kHz 附近的峰值 (黑色箭头指示) 频移到 1.5kHz 频点上。整个输出信号的最高频率由原来的 8kHz 变换为现在的 4kHz, 通过耳机测听, 声音的音调降低了很多, 声音低沉, 几乎不能辨认原来的声音音质, 但是由于是在频域按比率压缩信号的, 信号的能量峰关系保持不变, 语音的语意仍能辨认, 这会给重度耳聋患者听力水平带来提高。

针对本算法的编程, 把 8kHz 的语音数据压缩到 4kHz, 把语音变换到频域里面可以实现等比例压缩, 但在时域里面也可以实现, 这样可以省掉了 FFT 的正反变换, 节省了系统的运行时间和资源。由数字信号理论可知, 频域压缩后相对应的时域信号被拉伸了, 通过计算分析得出频域压缩前后相对应的时域信号变换规律, 利用此规律直接在时域里面面对语音数据进行处理。实验表明在时域里面处理的效果和在频域里面处理的效果一

致，但程序的运行效率提高了。

上面通过把数字助听器的三个语音算法应用到 DSP 开发板上，实现了语音处理系统。此语音系统是实现数字助听器的基础，把数字助听器的所有算法应到的到此系统上，就做成了数字助听器的雏形。所编的算法加到主程序中运行后发现几乎没有延时，算法的可靠性得到了验证。

5.4.3 算法实现过程中遇到的问题

下面简述了一下算法在 DSP 上实现的过程中遇到的问题。把遇到了的问题和解决的方法做一下简单的讨论。

(1) 系统的实时性问题

在编写程序时首先要注意的是输入输出信号实时性的问题。一帧语音信号输入到 DSP 中进行语音算法处理的时间是否小于一帧信号采集的时间，直接影响着输入的语音信号能否完全处理完并且即时输出。如果不能满足上面的要求则语音处理不完整和语音输出时序发生混乱，将导致输出语音无法辨认。此问题在上节中作了详细的讨论，由于本系统的核心处理器使用的是主频达到 160MHz 的 C5416DSP，一帧语音信号处理的时间远远小于 AD 采样一次的时间，所以系统的实时得到了保障。

(2) RFFT 和 RIFFT 变换的问题

在把数据进行 FFT 变换的时候，选用了 DSP 库函数中的 RFFT 和 RIFFT 函数。此函数采用的是汇编语言编写，运行效率很高。但在进行正反变换的时候可以发现，变换后的时域信号幅度比原信号要小，这主要因为在进行变换的时候函数进行了防溢出处理，数值上变小了，但是波形和原来保持了一致。要想恢复到原信号只要在信号上加上一个适当的增益即可。本文用是实数傅里叶变换，所以要求输入的数为实整数。

(3) 程序加载的问题

当把编写好多的程序再次加载到硬件开发板上时，有时会发现程序加载不进去，加载发生中止现象。这主要是由先前加载到存储器中的数据和现在加载的数据发生了冲突，引起了数据加载中断。解决这一问题的最简单手段就是在加载好一次程序运行后，如果想把重先修改的程序加载到板子上运行并确保数据加载运行的可靠性，需要对开发板进行复位清零，让开发板恢复到初始状态。如果在程序加载的过程中出现中止的现象，最先想到的解决方法是对开发板进行复位，如果问题还没有解决那就需要检查其写的 CMD 文件是否和开发板上的存储空间有冲突或者其他程序编写错误。

(4) 程序运行后,起初的杂音问题

本文所编写的程序在开发板上正常运行后，如果此时停止运行后再运行程序会发现从耳机里面会传来一段不规则的杂音，然后才会听到正常的语音。通过分析查看程序停止运行后缓冲区中的数据，发现程序停止运行后本文所设置的缓冲区中的半个缓冲区中的数据不正常，另外半个是正常的。这可能是程序停止的时候，数据正往其中的一个缓冲区写数据，此时中断了数据的传输，缓冲区中的数据发生了不预料的现象，引起了数据的不正常，再次运行程序时会把这段不正常数据输出到耳机中，产生了不规则杂音。

5.5 本章小结

本章主要研究了数字助听器算法的实现。先介绍了算法实现所用的硬件平台 SEED-DEC5416 的基本配置，重点介绍了 AD 芯片 AIC23。然后设计了系统软件的整体流程，McBSP 和 AIC23 的初始化和语音算法的编程。在算法的实现部分分别把数字助听器的三个核心算法改写为 C 语言，通过 CCS 强大的编译功能把程序加载到开发板上运行。同时通过 CCS 上的图形显示功能显示输入输出数据和存储数据功能把处理后的数据加入到 MATLAB 上进行性能分析。在本章的最后分析了算法的实时性问题和实践中遇到问题的体会，通过分析表明本算法能够实现数字助听器的预期处理效果。

第六章 总结与展望

6.1 全文内容总结

本文主要研究了数字助听器的核心算法及算法在 DSP 上的实现。将人耳的听觉掩蔽效应应用到数字助听器的核心算法中,直接从人耳的听觉生理上出发,改善人耳对声音的舒适度,提高了人耳对语音的听觉感知。在算法的实现上,利用了合众达的 SEED-DEC5416 硬件平台,编写数字助听器语音算法程序,并进行了实验验证。

本文首先从数字助听器的常用三个算法入手:语音增强算法、宽动态压缩算法和移频压缩算法。从他们的原理深入分析和阐述,对原有算法处理方案的优点和不足点作出了分析,尤其是语音增强算法,传统用的是自适应滤波增强,而本文从人耳的生理特性出发,用听觉掩蔽效应的语音增强算法来代替自适应算法。

其次,对人耳的生理特性—听觉掩蔽效应和听觉机理进行了详细的分析,对语音增强中所遇到的噪声和语音特性进行了介绍,利用听觉掩蔽效应和噪声实时估计实现了语音增强算法。

再次,针对传统听觉掩蔽效应语音增强中听觉掩蔽阈值计算误差导致语音增强算法性能降低的现象,深入分析导致误差的原因,对原有的算法公式进行了改进,提出了一种听觉掩蔽效应语音增强的改进算法。实验表明该算法的信噪比和听觉舒适度得到了改善。

然后,针对传统的宽动态压缩算法,忽略了语音共振峰的存在,在实现频谱增益时引起了共振峰的非线性失真,提出了一种基于共振峰估计的宽动态压缩算法。在同一个共振峰频段内使用相同的增益,减小了语音共振峰的非线性失真。实验表明该算法降低了语音的失真度,改善了听觉舒适度。

最后,对数字助听器的三个核心算法进行了验证。利用合众达的 SEED-DEC5416 开发板,把数字助听器的三个语音处理算法移植到 DSP 中,并对其中出现的问题进行了分析。在算法的验证过程中利用了 PC 机上的标准语音通过音频输出端口输入到开发板上,通过耳机对输出语音进行测听,测听结果表明在此开发板上实现了数字助听器的三种语音处理功能,达到了数字助听器雏形的要求。

6.2 后续研究方向

数字助听器的优越性就在于其能够加入不同的语音处理算法,其有很大的市场前景。而语音处理算法可以不断的进行优化改进,提高算法的性能。本文由于研究时间比较的短,针对数字助听器的两个核心算法进行了改进,并且只是在 DEC5416 开发平台上验证了语音处理算法,所以本文的工作还有待于进一步研究改善。其主要体现在下面几个方面:

- (1) 进一步从人耳的生理特性出发,把听觉掩蔽效应应用到数字助听器的其他几个

算法中去，充分利用人耳的听觉感知特性，使算法处理后的语音能够更好被耳病患者所接受。

(2) 宽动态压缩算法中，共振峰估计选用了 LPC 法，此算法能很好的提取共振峰但算法计算的耗时比较大，有必要研究共振峰提取的快速算法。

(3) 本文只是在开发板上验证了数字助听器的三个算法，由于条件的限制，没有真正的让患者来体验。下一步如有条件可以让患者来体验一下，从实际应用中检验算法的可靠性和使用性。

致谢

论文完成之际，衷心感谢对我的课题及论文进行指导和帮助的所有老师和同学！

本文的全部研究成果，都是在导师于凤芹教授的精心指导下完成的。于老师不仅耐心地引导我进入了一个对我来说陌生而又全新的研究领域，而且为我创造了优越的科研与学习环境，使我的科研能力有了显著提高。于老师深厚的专业功底、严谨求实的治学精神、勤奋认真的工作作风、严以律己宽以待人的处事态度都在潜移默化地影响着我、鞭策着我，使我不仅在学业和科研方面深受教益，而且在为人处事及个人修养方面都受益匪浅。

两年来，江南大学和通信与控制工程学院所提供的各种学习资源和科研设施，是我顺利完成学业和科研工作的重要平台。通过聆听各位任课老师的课程，为我的科研能力的进步积累了扎实的专业基础知识。

感谢同实验室的师兄张正凯、李玉、肖志、李江、王淦平、范树凯对我学习和生活上的帮助。

感谢同实验室的师兄弟黄叶权、丁祝高、袁佳能、郭昕、孔磊、田源和宋倩倩在我们共同学习探讨中给了我不断努力进取的动力，尤其感谢黄叶权，他在我编写 MATLAB 程序有困难时给了我莫大的帮助。

感谢上海交通大学的严勇同学，他给我提供了很多宝贵的学习资料。

感谢含辛茹苦的父母对我的养育之恩，感谢我女朋友谢小倩及其家人的理解，他们对我的学业一贯支持和鼓励，给了我战胜困难的信心和勇气。他们永远是我强大的精神支柱和坚强的后盾。

本文所引用的全部参考文献，都是我学习和研究得以提升的重要源泉。他们的理论、思想与方法为我的学位论文提供了有益的思路和启迪。

常怀一颗感恩的心，真诚地感谢所有帮助过、关注过我的人！

最后，向百忙之中审阅论文的各位专家、教授致以崇高的敬意和深深的感谢！

参考文献

1. 韩纪庆,张磊,郑铁然.语音信号处理[M].北京:清华大学出版社,2004.25-32
2. 张雄伟,陈亮,杨吉斌.现代语音处理技术及应用[M].北京:机械工业出版社,2003.9-11,250-252
3. Berouti M,Schwartz R,Makhoul J.Enhancement of speech corrupted by acoustic noise[C]. IEEE International Conference on Acoustics,Speech,and Signal Processing,1979,(4): 208-211
4. Lockwood P,Boudy J.Experiments with a nonlinear spectral subtractor (NSS), hidden markov models and projection, for robust speech recognition in cars [J]. Speech Communication,1992,11(2):215-228
5. Nathalie V.Signal channel speech enhancement based on masking properties of human auditory system[J].IEEE Transactions on Speech and Audio Processing,1999,7(2):126-137
6. Lu C.Reduction of musical residual noise for speech enhancement using masking properties and optimal smoothing[J].Pattern Recognition Letter,2007,28(11):1300-1306
7. Zhong L,Rafik.G, Richard M.Noise estimation using speech/non-speech frame decision and subbed spectral tracking[J], Speech Communication,2008, 49(7): 542-557
8. Yoshifumi C,Kotaro M,Katsumori H,Hidetoshi N,etc.Real-time processing using the frequency domain binaural model[J].Applied Acoustics,2007, 68(8):923-938
9. Kotta M, Preeti R.Speech enhancement in nonstationary noise environments using noise properties [J]. Speech Communication,2006, 48(1):96-109
10. 陶智,赵鹤鸣,龚呈卉.基于听觉掩蔽效应和 Bark 子波变换的语音增强[J].声学学报, 2005,30(4):367-372
11. 陶智,赵鹤鸣,吴迪.基于临界频带子波变换的语音增强[J].微电子学与计算,2004, 21(12):50-52
12. 徐望,王炳锡,丁琦.一种基于信号子空间和听觉掩蔽效应的语音增强方法[J].信号处理,2005,20(2):112-116
13. 王莉,胡剑波,徐盛.基于听觉掩蔽效应得语音增强算法的研究[J].电声技术,2006, 4(7):39-43
14. 刘海斌,吴镇杨,赵力,等.非平稳环境下基于人耳听觉掩蔽特性的语音增强[J].信号处

理, 2003,19(4):303-307

15. 谢志文,尹俊勋.同时掩蔽效应的实验研究[J].声学技术,2006,25(5):446-450
16. 同鸣,郝重阳,刘晓军.基于听觉掩蔽特性的扩频信息隐藏方法[J].计算机工程,2006,32(10):137-139
17. 清源科技.TMS320C54x DSP 硬件开发教程[M].北京:机械工业出版社, 2003.195-202
18. 清源科技.TMS320C54x DSP 应用程序设计教程[M].北京:机械工业出版社,2003.2-5
19. Texas Instruments Incorporated 著梁晓雯等编译.TMS320VC54 系列 DSP 的 CPU 与外设[M].北京:清华大学出版社,2006.5-10
20. Sen M.Kuo BobH.Lee 著.卢伯英译.实时数字信号处理[M].北京:中国铁道出版社,2004.340-343
21. 赵明忠,顾斌,王钧铭,等.DSP 应用技术[M].西安:西安电子科技大学出版社,2004.198-200
22. 汪安民,程昱.DSP 应用开发实用子程序[M].北京:人民邮电出版社,2005.252-285
23. 贾彬彬.数字助听器产品的市场现状和技术展望[EB/OL].http://www.mednet.cn/News/NewsText/News_557.htm,2005-5-9
24. 吴星华.基于 DSP 的数字助听器开拓[J/OL].<http://www.icembed.com/info-15355.htm>, 2007-2-10
25. 库奇.了解数字助听器[EB/OL].<http://www.kuqiqi.com/jtjy/tsjy/tzjy/211933553.html>, 2007-10-16
26. 肖宪波,王宁宇,胡广书.数字助听器中若干主要算法的发展和现状[J].生物医学工程学,2004,21(4):694-698
27. 何振亚.自适应信号处理[M].北京:科学出版社,2002.239-241
28. 陈广飞,应俊.数字助听器宽动态压缩算法研究[J].北京生物医学工程,2006,25(5):454-456
29. 应俊.基于 DSP 数字助听器关键技术的研究[D]:[硕士学位论文].北京:中国人民解放军军军医进修学院,2006
30. 梁勇.移频助听技术与移频助听器[J].中国听力语言康复科学杂志,2005,3(2):25-27
31. 赵力.语音信号处理[M].北京:机械工业出版社,2003.77-80,272-273
32. 马大猷.现代声学理论基础[M].北京:科学出版社,2004.50-51
33. Rainer M.Noise power spectral density estimation based on optimal smoothing and minimum statistics[J].IEEE Transactions on Speech and Audio Processing,2001,9(5):

504-512

34. Kathryn H, Arehart J, Stephen G, etc. Evaluation of an auditory masking threshold noise suppression algorithm in normal-hearing and hearing-impaired listeners[J]. *Speech Communication*, 2003, 40(4): 575-592
35. 蔡汉添, 袁波涛. 一种基于听觉掩蔽模型的语音增强算法[J]. *通信学报*, 2002, 23(8): 93-98
36. Ted P, Andreas S. Perceptual coding of digital audio[J]. *Proceedings of the IEEE*, 2000, 88(4): 451-512
37. Lu C, Wang H. Speech enhancement using perceptually constrained gain factors in critical-band wave-packet transform[J]. *Electronics Letters*, 2004, 40(6): 394-396
38. Marka B, Elim S. Auditory masking of anuran advertisement calls by road traffic noise[J]. *Animal Behaviour*, 2007, 74(6): 1765-1776
39. 王晓亚. 倒谱在语音基音和共振峰提取中的应用[J]. *无线电工程*, 2004, 34(1): 57-61
40. 郁伯康, 郁梅. LPC 方法提取语音信号共振峰的分析[J]. *电声技术*, 2000, 3(1): 3-8
41. 何峰, 陈晓清, 李国锁等. 一种新的语音信号共振峰提取的算法[J]. *信号处理*, 2007, 23(4): 618-621
42. Audiologic I, Boulder C. The continuous frequency dynamic range compressor[C]. *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, 2003, (2): 19-22
43. Teiero-Caladol J, Rutledge J, Nelson P. Preserving spectral contrast in amplitude compression for hearing aids[C]. *Engineering in Medicine and Biology Society in Proceedings of the 23th Annual International Conference of the IEEE*, 2001, (2): 1453-1456
44. Teiero J, Bernal S, Hidalgo J. A digital hearing aid that compensates loudness for sensorineural hearing impairments[C]. *International Conference on Acoustics, Speech, and Signal Processing*, 1995, (5): 2991-2994
45. Shinichi S, Katsuhiko G, Makoto T. Frequency compression hearing aid for severe-to-profound hearing impairments[J]. *Auris Nasus Larynx*, 2006, 27(4): 327-334
46. Brian C. Speech processing for the hearing-impaired: successes, failures, and implications for speech mechanisms[J]. *Speech Communication*, 2003, 41(2): 81-91
47. Jean-Baptiste M, Liesbeth R, Jan W. Comparison of adaptive noise reduction algorithms in dual microphone hearing aids[J]. *Speech Communication*, 2007, 48(8): 957-970

48. Harold K,David B.Hearing aids:A review of what's new[J].Otolaryngology-Head and Neck Surgery,2006,134(6):1043-1050
49. Jong W,Steven M,Ward R,ect.Improving performance in noise for hearing aids and cochlear implants using coherent modulation filtering[J].Hearing Research ,2008,239(2):1-11
50. 应俊,陈广飞,柴学宁.基于 DSP 的全数字助听器设计和实现[J].医疗卫生装备, 2006,27(7):8-9
51. Texas Instruments Incorporated.TMS320VC5416 Fixed-point Digital Signal Processor [DB/OL]. <http://focus.ti.com/lit/an/spra095d/spra095d.pdf>,2004
52. Texas Instruments Incorporated. TLV320AIC23B Stereo Audio Codec [DB/OL]. <http://focus.ti.com/lit/ds/sgls240a/sgls240a.pdf>, 2004
53. Texas Instruments Incorporated. TMS320C54X Code Composer Studio Tutorial [DB/OL]. <http://focus.ti.com/lit/an/spra327d/spra327d.pdf>,2004
54. 王剑峰,马艳萍,刘云.McBSP 在语音信号处理中的应用与实现[J].微计算机应用, 2006,27(3):349- 351
55. 蔡斌,郭英,李虹伟,等.应用 TMS320C54X 系列 DSP 实现语音增强算法[J].仪器仪表学报, 2003,24(4):440-442
56. 杨银堂,马峰,刘毅.DSP 应用开发从实践到提高[M].北京:中国电力出版社,2007.50-56

附录：作者在攻读硕士学位期间发表的论文

已发表或录用的论文：

1. 阚仁根,于凤芹. 基于听觉掩蔽效应和噪声估计的语音增强算法[J]. 电声技术, 2008,32(2):55-57
2. 于凤芹, 阚仁根. 听觉掩蔽效应语音增强的改进算法[J]. 声学技术, (已录用)
- 3 于凤芹, 阚仁根. 基于共振峰估计的宽动态压缩算法[J]. 生物医学工程杂志(已投稿)

附录：部分源程序

注：下面的 matlab 源程序，需要下载相对应的应用工具箱，否则下面程序无法运行。

```
% speech enhancement by noise masking
% 增强语音平滑
% speech threshold
% 2007 年 10 月 26 日
clear;
close all
clc;
var1=0.01;
lamta=0.98;
[signalAll,fs,bit]=wavread('D:\matlab\work\voice
library\Male.wav');
signal=signalAll(:,1);
noise=var1*randn(size(signal));
%%%%%%%%%%%%使用噪声库
%length_noise=length(signal);
%[noise_all,fs_noise,bit_noise]=wavread('D:\matlab\work\voice library\white_2s_16khz.wav');
%noise=noise_all(1:length_noise);
%%%%%%%%%%%%
y=signal+noise;
length_y=length(y);
length_over=mod(length_y,256);
length_Y=length_y-length_over;
% Take the FFT of overlapping frames
wind=hanning(256);
i=1;
for t=1:128:(length_Y-256);
    sp(:,i)=fft(wind.*y(t:(t+255)));
    sp_signal(:,i)=fft(wind.*signal(t:(t+255)));
    noise_sp(:,i)=fft(wind.*noise(t:(t+255)));
    abs_sp_signal(:,i)=abs(sp_signal(:,i)).^2;
    abs_noise_sp(:,i)=(abs(noise_sp(:,i))).^2;
    abs_sp(:,i)=(abs(sp(:,i))).^2;%%%
    angle_sp(:,i)=angle(sp(:,i));
    i=i+1;
end;
%%%%%%%%%%%% noise speech smooth
num_frame=length_Y/128-2;
NoiseSpeechSp=abs_sp;
NoiseSpeechSmooth=zeros(256,num_frame);
```

```
for i=2:num_frame;
NoiseSpeechSmooth(:,i)=lamta*NoiseSpeechSmooth(:,i-1)+(1-lamta)*NoiseSpeechSp(:,i);
end;
NoiseSpeechSmooth(:,1)=NoiseSpeechSp(:,1);
%%%%%%%%%%%% noise estimation
NoiseSp=zeros(256,num_frame);
for i=1:num_frame

NoiseSp(:,i)=min(NoiseSpeechSmooth(:,i))*ones(256,1);
end
%%%%%%%%%%%% estimate the speech threshold
SumSp=sum(abs_sp);
SumSpmin=min(SumSp);
SpeechThreshold=(SumSpmin*2.1)*ones(1,num_frame);
%plot(1:num_frame,SumSp,'b-*,1:num_frame,SpeechThreshold,'r-');
%axis([0 num_frame 0 10]);
%%%%%%%%%%%% reduce the noise by the speech threshold
for i=1:num_frame;
    if sum(abs_sp(:,i))<SpeechThreshold;
        SpeechEnhanceFirst(:,i)=zeros(256,1);
    else
SpeechEnhanceFirst(:,i)=abs(abs_sp(:,i)-1*NoiseSpeech(:,i));
    end
end
%%%%%%%%%%%% 纯净语音的掩蔽阈值
signal_masking(:,i)=abs_sp_signal;
%%%%%%%%%%%%噪声阈值计算 分为 24 个临界频带
varr1=[1 2 3 5 6 8 10 12 15 17 20 24 28 32 37 43 50 59 70 85 102 123];
varr2=length(varr1);
BB_compare=zeros(2*(varr2-1),num_frame);
meanArith=zeros(2*(varr2-1),num_frame);
meanGe=zeros(2*(varr2-1),num_frame);
for i=1:varr2-1
BB_compare(i,:)=sum(signal_masking(varr1(i):var
```

```

r1(i+1,:));
meanArith(i,:)=mean(signal_masking(varr1(i):varr1(i+1),:));
meanGe(i,:)=(prod(signal_masking(varr1(i):varr1(i+1),:)).^(1/(varr1(i+1)-varr1(i)+1)));

BB_compare(2*(varr2-1)+1-i,:)=BB_compare(i,:);
    meanArith(2*(varr2-1)+1-i,:)=meanArith(i,:);
    meanGe(2*(varr2-1)+1-i,:)=meanGe(i,:);
end;

for i=1:num_frame;
    if sum(meanArith(:,i))>0;
        SFM(:,i)=meanGe(:,i)/meanArith(:,i);
        SFMdb(:,i)=10*log10(SFM(:,i));
    else SFMdb(:,i)=zeros(2*(varr2-1),1);
    end;
end;

df=SFMdb./(-60);
dff(find(df>1))=1;
Oi=zeros(2*(varr2-1),num_frame);

for i=1:2*(varr2-1);
    Oi(i,:)=dff(i,:)*(14.5+i)+(ones(1,num_frame)-dff(i,:))*5.5;
end;

for i=1:num_frame;
    if BB_compare(:,i)>0;

        dh(:,i)=log10(BB_compare(:,i))-(Oi(:,i)/10);
        Ti(:,i)=abs(10.^dh(:,i));
        else Ti(:,i)=zeros(2*(varr2-1),1);
        end;
end;

TI=zeros(256,num_frame);
for i=1:varr2-1;

    TI(varr1(i):varr1(i+1),:)=ones(varr1(i+1)-varr1(i)+1,1)*Ti(i,:);
end;

for i=1:128;
    TI(257-i,:)=TI(i,:);

```

```

end
Tlclearspeech=TI;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%绝对掩蔽阈值
Tla=zeros(256,num_frame);
for k=1:128
    fsK=62.5*k/1000;

    Tla(k,1)=3.64*(fsK)^(-0.8)-6.5*exp(-0.6*(fsK-3.3)^2)+0.001*fsK^4;
    Tla(257-k,1)=Tla(k,1);
end
Tla=Tla(:,1)*ones(1,num_frame);
Tlclearspeechfinal=max(Tlclearspeech,Tla);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%noise masking
signal_masking(:,:)=SpeechEnhanceFirst;
%plot(signal_masking(:,180))
%axis([120 135 0 0.001])

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%噪声阈值计算 分为 24 个临界频带
varr1=[1 2 3 5 6 8 10 12 15 17 20 24 28 32 37 43 50 59 70 85 102 123];
varr2=length(varr1);
BB=zeros(2*(varr2-1),num_frame);
meanArith=zeros(2*(varr2-1),num_frame);
meanGe=zeros(2*(varr2-1),num_frame);
for i=1:varr2-1

    BB(i,:)=sum(signal_masking(varr1(i):varr1(i+1),:));

    meanArith(i,:)=mean(signal_masking(varr1(i):varr1(i+1),:));

    meanGe(i,:)=(prod(signal_masking(varr1(i):varr1(i+1),:)).^(1/(varr1(i+1)-varr1(i)+1)));
    BB(2*(varr2-1)+1-i,:)=BB(i,:);
    meanArith(2*(varr2-1)+1-i,:)=meanArith(i,:);
    meanGe(2*(varr2-1)+1-i,:)=meanGe(i,:);
end;

for i=1:num_frame;
    if sum(meanArith(:,i))>0;

```

```

SFM(:,i)=meanGe(:,i)./meanArith(:,i);
SFMdb(:,i)=10*log10(SFM(:,i));
else SFMdb(:,i)=zeros(2*(varr2-1),1);
end;
end;

df=SFMdb./(-60);
df(find(df>1))=1;
Oi=zeros(2*(varr2-1),num_frame);

for i=1:2*(varr2-1);
Oi(i,:)=df(i,:)*(14.5+i)+(ones(1,num_frame)-df(i,:))
)*5.5;
end;

for i=1:num_frame;
if BB(:,i)>0;
dh(:,i)=log10(BB(:,i))-(Oi(:,i)/10);
Ti(:,i)=abs(10.^dh(:,i));
else Ti(:,i)=zeros(2*(varr2-1),1);
end;
end;

end;

TI=zeros(256,num_frame);
for i=1:varr2-1;

TI(varr1(i):varr1(i+1),:)=ones(varr1(i+1)-varr1(i)+
1,1)*Ti(i,:);
end;

for i=1:128;
TI(257-i,:)=TI(i,:);
end
TI=max(TI,TIa);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% 计算谱减系数
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
修正系数
CC=ones(256,num_frame);
space=TIclearspeechfinal-TI;
spaceMax=max(space);
for i=1:num_frame;
for j=1:256
if space(j,i)>0;
CC(j,i)=space(j,i)/spaceMax(i)+1;
end

```

```

end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
AAcompare=zeros(256,num_frame);
AA=zeros(256,num_frame);
for i=1:num_frame;
if sum(TI(:,i))>0;
TImax=max(TI(:,i))*ones(256,1);
TImin=min(TI(:,i))*ones(256,1);

AAcompare(:,i)=(6*TImax-TImin-5*TI(:,i))./(TIm
ax-TImin);
%%
AA(:,i)=(6*TImax-TImin-5*TI(:,i).*CC(:,i))./(TIm
ax-TImin);

AA(:,i)=(6*TImax-TImin-5*TI(:,i))./((TImax-TIm
in).*CC(:,i));
end;
end;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%谱减系数比较图
%AACompare_frame=130;
%%subplot(211)
%plot(1:256,AA(:,AACompare_frame),'b-','1:256,
AAcompare(:,AACompare_frame),'r-');
%axis([1 128 0 7]);
%legend('改进谱减系数值','原始谱减系数计算
值')

%%subplot(212)
%plot(1:256,TI(:,AACompare_frame),'r',1:256,TIc
learspeechfinal(:,AACompare_frame),'b');
%axis([1 128 0 10])
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for i=1:num_frame;
if sum(abs_sp(:,i))<SpeechThreshold;
SpeechEnhance(:,i)=zeros(256,1);

SpeechEnhanceCompare(:,i)=zeros(256,1);
%else
SpeechEnhance(:,i)=abs(abs_sp(:,i))-AA(:,i).*Noise
Sp(:,i));
else
SpeechEnhance(:,i)=abs_sp(:,i)-AA(:,i).*NoiseSp(

```

```

,i);

SpeechEnhanceCompare(:,i)=abs_sp(:,i)-AAcompare(:,i).*NoiseSp(:,i);
    end;
end;
SpeechEnhance(find(SpeechEnhance<0))=0.0001;
SpeechEnhanceCompare(find(SpeechEnhance<0))=0.0001;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%计算平滑系数
for i=1:num_frame;
    varNoise(:,i)=sum(NoiseSp(:,i));
    varSpeech(:,i)=sum(SpeechEnhance(:,i));
end
VARnoise=ones(256,1)*varNoise;
VARspeech=ones(256,1)*varSpeech;
Ymw1=VARspeech./VARnoise;
for i=2:num_frame;

Ymw2(:,i)=SpeechEnhance(:,i-1)./VARnoise(:,i);
end
Amw=ones(256,num_frame)./((Ymw2-Ymw1).^2+Ymw1.^2+ones(256,num_frame));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%平滑语音
la=Amw;
SpeechEnhancefinal=zeros(256,num_frame);
SpeechEnhancefinalCompare=zeros(256,num_frame);
for i=2:num_frame;

SpeechEnhancefinal(:,i)=la(:,i).*SpeechEnhancefinal(:,i-1)+(ones(256,1)-la(:,i)).*SpeechEnhance(:,i));

SpeechEnhancefinalCompare(:,i)=la(:,i).*SpeechEnhancefinalCompare(:,i-1)+(ones(256,1)-la(:,i)).*SpeechEnhanceCompare(:,i);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%纯净语音,原始听觉掩蔽效应算法的语音,改进后语音谱的频谱比较值
frameNumber=210;
plot(1:256,abs_sp_signal(:,frameNumber),'r-',1:256,SpeechEnhancefinalCompare(:,frameNumber),'b
-^',1:256,SpeechEnhancefinal(:,frameNumber),'k-*');
axis([30 60 0 0.17]);
legend('纯净语音谱','听觉掩蔽语音增强后的语音频谱','改进听觉掩蔽语音增强后的语音频谱');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%ifft
X_hat_FB=sqrt(SpeechEnhancefinal);
X_hat=zeros(1,length_Y);
j=sqrt(-1);
i=1;
for t=1:128:(length_Y-256);

X_hat(:,(t:t+255))=X_hat(:,(t:t+255))+real(ifft(X_hat_FB(:,i).*exp(j*angle_sp(:,i))));
    i=i+1;
end;

xx=1:length(signal);
xx=xx/fs;
subplot(411)
plot(xx,signal);title('纯净信号')
axis([0 xx(end)-1 1]);
subplot(412)
plot(xx,y);title('含噪信号')
axis([0 xx(end)-1 1]);
out=X_hat;
subplot(413)
plot(xx,out);title('输出信号')
axis([0 xx(end)-1 1]);
subplot(414)
plot(xx,noise);title('噪声信号')
axis([0 xx(end)-1 1]);

wavwrite(y,fs,'a');
wavwrite(out,fs,'speechout');

```

注：下面是在 SEED-DEC5416 开发板上运行的主程序，其相对应的调用库函数没有列出。

.CMD 文件

```
* ===== dec5416.cmd =====

MEMORY
{
    PAGE 0: /*软件仿真*/
    /*VECT: origin = 0x1000, len = 0x0200
    IPROG2: origin = 0x1200, len = 0x1000*/
    /*EMU 仿真*/
    VECT: origin = 0x28100, len = 0x100
    IPROG1: origin = 0x200, len = 0x7c00
    IPROG2: origin = 0x28200, len = 0x7c00
    IPROG3: origin = 0x38000, len = 0x8000
    /*当 FLASH 处于低端的程序空间时*/
    FLASHL: origin = 0x40000, len = 0x40000
    /*当 FLASH 处于高端的程序空间时*/
    FLASHH: origin = 0x4000000, len = 0x80000

    PAGE 1: /*软件仿真*/
    /*    SPRAM: origin = 0x60, len = 0x20
    DARAM: origin = 0x2000, len = 0x1000*/
    /*EMU 仿真*/
    SPRAM: origin = 0x60, len = 0x20
    DARAM: origin = 0x80, len = 0x3f80
    /*做相应设置后也可以是 FLASH 空间*/
    SARAM: origin = 0x8000, len = 8000
}

SECTIONS
{
    .vector: {} > VECT PAGE 0
    .text: {} > IPROG2 PAGE 0
    .cinit: {} > IPROG2 PAGE 0
    .switch: {} > IPROG2 PAGE 0
    .const: {} > IPROG2 PAGE 0

    .sys_buffer: {} > SPRAM PAGE 1
    .stack: {} > DARAM PAGE 1
    .motor: {} > DARAM PAGE 1
    .data: {} > DARAM PAGE 1
    .uartb_rsect: {} > DARAM PAGE 1
```

```
.uartb_rsect: {} > DARAM PAGE 1
.bss: {} > DARAM PAGE 1
.test: {} > DARAM PAGE 1
}
```

Code.c 文件

```
/******
*****
*语音
*****
*****

#include "type.h"
#include "codec.h"
#include "sysreg.h"
#include "mcbbsp54.h"
#include "memory.h"
#include "tms320.h"
#define FFT_SIZE 128
#define NUM_buffer FFT_SIZE/2
#define NUM_throld1 NUM_buffer-1
#define NUM_throld2 FFT_SIZE-1

DATA input[FFT_SIZE]={0};
#pragma DATA_SECTION (input,".test")

HANDLE codec_command=0;
HANDLE codec_data=0;

unsigned int i= 0;
unsigned int k= 0;

main()
{
    unsigned int i= 0;
    unsigned int j= 0;
    DATA input[FFT_SIZE]={0};
    DATA buffer1[NUM_buffer]={0};
    DATA buffer2[NUM_buffer]={0};

    /*设置系统时钟*/
    sys_clk(CLK160);
    /*初始化 DEC5416 板卡*/
    dec5416_init();
    /*打开 codec 数据接口*/
```


作者: [阚仁根](#)
学位授予单位: [江南大学](#)
被引用次数: 1次

本文读者也读过(10条)

1. [崔海徽](#) 全数字助听器的数字信号处理技术研究[学位论文]2004
2. [应俊](#) 基于DSP数字助听器关键技术的研究[学位论文]2006
3. [尹向雷](#) 数字助听器中语音增强算法研究及DSP实现[学位论文]2010
4. [王青云](#), [赵力](#), [乔杰](#), [邹采荣](#), [WANG Qing-yun](#), [ZHAO Li](#), [QIAO Jie](#), [ZOU Cai-rong](#) 符合人耳听觉特征的数字助听器子带响度补偿[期刊论文]-应用科学学报2008, 26 (6)
5. [王青云](#), [赵力](#) 一种新的数字助听器自适应回波抵消方法[会议论文]-2008
6. [马晓虹](#), [Ma Xiaohong](#) 基于DSP的助听器语音处理系统设计[期刊论文]-科技广场2009 (7)
7. [应俊](#), [陈广飞](#), [柴学宁](#), [郝瑞庭](#), [YING Jun](#), [CHEN Guang-fei](#), [CHAI Xue-ning](#), [HAO Rui-ting](#) 基于DSP的全数字助听器设计和实现[期刊论文]-医疗卫生装备2006, 27 (7)
8. [马强](#), [吴荣彬](#), [曲天书](#), [吴玺宏](#), [迟惠生](#) 一种数字助听器系统研究开发平台的实现[会议论文]-
9. [王青云](#), [赵力](#), [赵立业](#), [邹采荣](#), [Wang Qing-yun](#), [Zhao Li](#), [Zhao Li-ye](#), [Zou Cai-rong](#) 一种数字助听器多通道响度补偿方法[期刊论文]-电子与信息学报2009, 31 (4)
10. [赵立业](#), [ZHAO Li-ye](#) 数字助听器系统中的回声消除方法[期刊论文]-传感器与微系统2009, 28 (11)

引证文献(1条)

1. [尹向雷](#), [郑恩让](#), [马令坤](#), [马晓虹](#) 基于掩蔽效应的维纳滤波器语音增强及DSP实现[期刊论文]-电子技术应用2010 (4)

本文链接: http://d.wanfangdata.com.cn/Thesis_Y1397321.aspx