分类号　　密　级

太原理工大学

硕士学位论文

题　　目　改进相位谱信息及相位重构的语音增强算法研究

英文并列题目　Speech Enhancement Research Based on Improved Phase Spectrum Information and Phase Reconstruction

研究生姓名：吉慧芳

学　　号：2016510297

专　　业：信息与通信工程

研究方向：语音信号处理

导师姓名：贾海蓉

职　　称：副教授

论文提交日期　2019/05

学位授予单位：太原理工大学

地　　址：山西·太原

太　原　理　工　大　学

声　　明

本人郑重声明：所呈交的学位论文，是本人在指导教师的指导下，独立进行研究所取得的成果。除文中已经注明引用的内容外，本论文不包含其他个人或集体已经发表或撰写过的科研成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律责任由本人承担。

论文作者签名：吉慧芳　　日期：2019.5.29

关于学位论文使用权的说明

本人完全了解太原理工大学有关保管、使用学位论文的规定，其中包括：①学校有权保管、并向有关部门送交学位论文的原件与复印件；②学校可以采用影印、缩印或其它子复制手段复制并保存学位论文；③学校可允许学位论文被查阅或借阅；④学校可以学术交流为目的，复制赠送和交换学位论文；⑤学校可以公布学位论文的全部或部分内容（保密学位论文在解密后遵守此规定）。

签　　名：吉慧芳　　日期：2019.5.29

导师签名：[illegible]　　日期：2019.5.29

改进相位谱信息及相位重构的语音增强算法研究

摘 要

在过去的三十年里，语音增强一直是一个活跃的研究领域，它能将被各种噪声干扰下的带噪语音中的“纯净语音”提取出来，达到提高语音感知质量和可懂度的目的。语音增强可以应用在语音通信、人机交互和人工智能等领域。

目前，研究者们已经提出了各种基于谱幅值信息的语音增强算法，然而这些算法在抑制背景噪声的同时引入了一些语音失真，从而降低了语音可懂度。而近年来的研究表明相位信息在提升语音信号的可懂度方面有显著影响。对此，本文在具体介绍相位谱补偿及相位重构的语音增强算法的基础上，提出了改进相位谱信息及相位重构的语音增强算法。本文的主要工作如下：

1.简要介绍了语音增强的目的、意义及国内外发展状况，阐述了传统的语音增强算法原理，分析了噪声估计方法、基频估计并进行了相关实验仿真。

2.针对传统的相位谱补偿算法中补偿因子固定，无法对含噪语音相位谱进行灵活补偿的问题，提出了一种改进相位谱信息的语音增强方法。首先，该算法提出了基于每帧语音信号输入信噪比的*Sigmoid*型相位谱补偿函数，能够根据噪声的变化来灵活地对带噪语音的相位谱进行补偿；然后，结合改进DD的先验信噪比估计与语音存在概率算法(SPP)来估计噪声功率谱；最后，在维纳滤波中结合新的语音存在概率噪声功率谱估计与相位谱补偿来提高语音的增强效果。相比传统相位谱补偿(PSC)算法而言，改进算法可以有效抑制音频信号中的各类噪声，同时增强语音信号感知质量，提升语音的可懂度。

3.针对传统相位重构算法对语音信号清浊音段使用相同的方法重构相位，使得相位估计效果不明显的问题，提出了用信噪比信息与时频特征改进相位重构的新方法。首先，引入与相位失真有关的时频特征并计算决策阈值；接着，利用信噪比信息计算带噪语音与纯净语音的相位偏差，两项比较进一步估计清音段与浊音段的语音相位，能有效提高语音可懂度；最后，将重构的相位与改进二元假设模型的幅值估计结合并进行语音增强。经在不同噪声情况下、对不同带噪语音信号进行实验表明，新算法信噪比明显提高的同时、语音感知评价指标也优于传统算法，表明降低了语音失真、提高了语音可懂度。

关键词：语音增强，相位谱补偿，维纳滤波，相位重构，MMSE-LSA

SPEECH ENHANCEMENT RESEARCH BASED ON IMPROVED PHASE SPECTRUM INFORMATION AND PHASE RECONSTRUCTION

ABSTRACT

In the past three decades, speech enhancement has been an active research field, which can extracts the “pure speech” from the noisy speech affected by various kinds of noise to improve the perceived quality and intelligibility of speech. Speech enhancement can be applied in areas such as voice communication, human-computer interaction, and artificial intelligence.

At present, researchers have proposed various speech enhancement algorithms based on spectral amplitude information. However, these algorithms introduce some speech distortion while suppressing background noise, thus reducing the speech intelligibility. Recent studies have shown that phase information has a significant impact on improving the intelligibility of speech signals. Based on the introduction of the speech enhancement algorithms with phase spectrum compensation and phase reconstruction, a speech enhancement algorithm with improved phase spectrum information and phase reconstruction is proposed. The main work of this paper is as follows:

1. Briefly introduce the purpose, significance and development of speech

enhancement, expounds the principle of traditional speech enhancement algorithm, analysis the noise estimation method, fundamental frequency estimation and related experimental simulation.

2. Aiming to the problem that the compensation factor in the traditional phase spectrum compensation algorithm is fixed and the phase spectrum of the noisy speech can not be flexibly compensated, a speech enhancement algorithm with improved phase spectrum information is proposed. Firstly, the Sigmoid phase-spectral compensation function which is based on the signal to noise ratio of each frame input speech is presented in this paper, it can flexibly compensate the phase spectrum of the noisy according to the change of noise; next, Estimate noise power spectrum through combine a priori SNR estimation of the improved DD and speech presence probability algorithm (SPP) . Finally, Wiener filtering is applied to improve speech enhancement effect by combining the new speech presence probability noise power spectrum estimation and phase spectrum compensation. Compared to the traditional phase spectrum compensation (PSC) algorithm, the improved algorithm can effectively suppress the noise in the audio signal, and then enhance the perceived quality of the speech and improve speech intelligibility. Firstly, the Sigmoid phase-spectral compensation function which is based on the signal to noise ratio of each frame input speech is presented in this paper, it can flexibly compensate the phase spectrum of the noisy according to the change of noise; next, Estimate noise power spectrum through combine a priori SNR estimation of the improved DD and speech

presence probability algorithm (SPP) . Finally, Wiener filtering is applied to improve speech enhancement effect by combining the new speech presence probability noise power spectrum estimation and phase spectrum compensation. Compared to the traditional phase spectrum compensation (PSC) algorithm, the improved algorithm can effectively suppress the noise in the audio signal, and then enhance the perceived quality of the speech and improve speech intelligibility.

3. Aiming to the problem that the traditional phase reconstruction algorithm uses the same method to reconstruct the phase of the unvoiced and voiced segments of the speech signal, which makes the effect of phase estimation not obvious, a new method for improving the phase reconstruction by using the signal-to-noise ratio information and the time-frequency feature is proposed. Firstly, the time-frequency characteristics related to phase distortion are introduced and the decision threshold is calculated. Then, the signal-to-noise ratio information is used to calculate the phase deviation between the noisy speech and the pure speech. The two comparisons further estimate the speech phase of the unvoiced segment and the voiced segment, which can effectively improve the speech intelligibility. Finally, the reconstructed phase is combined with the amplitude estimation of the improved binary hypothesis model and speech enhancement is carried out. Experiments on different noisy speech signals under different noise conditions show that the signal-to-noise ratio of the new algorithm is significantly improved, and the speech perception evaluation

index is better than the traditional algorithm, which shows that the speech distortion is reduced and the speech intelligibility is improved.

KEY WORDS: speech enhancement, phase spectrum compensation, wiener filtering, phase reconstruction, MMSE-LSA

目录

第一章 绪论

1.1 研究背景及意义

无论身处何处，我们都必须通过语音通信来进行信息交流。然而在现实生活中，语音通信往往受到各式各样的噪声影响，导致语音感知质量与可懂度的降低。语音增强能够有效解决现实生活中的噪声干扰问题，其目的是将有用语音信号从各种背景噪声中提取出来，在改善语音信号质量的同时提高其可懂度，既降低了听觉疲劳，又提高了谈话时语音识别的准确性。

在许多领域我们都需要对语音进行增强，包括语音通话、场景录音、军事窃听、助听器、耳蜗植入物和说话人识别等[1]。在蜂窝电话标准中，使用声码器来对含噪语音进行压缩编码，语音增强系统能够用作其前端的预处理器来提高接收端的语音质量；在电话会议系统中，某一个终端的回响可能会被传送到其他所有接收端，语音增强系统能够在语音广播到其他接收端之前对其进行增强，提高系统的性能；在军事通讯系统中，周围空间的噪声会对任务的准确接收产生严重干扰，语音增强系统能够抑制背景噪声，改进语音质量及可懂度。语音增强技术不仅可以有效抑制音频信号中的噪声，增强语音信号感知质量，还会使语音通信设备变得更加便携和人性化。因此随着科学技术的发展，语音增强技术在现实生活中的意义愈加重要。

1.2 语音增强研究的历史和现状

自20世纪60年代开始，随着科学技术与信息通信技术的不断发展，推动了人们对语音增强技术的研究，使其成为一个活跃的研究领域。

谱减法[2]是早期Boll提出的一种变换域语音增强方法。在该方法中，输入语音信号被分解为短时段，在语音不存在时，利用带噪信号的频谱减去估计的噪声谱进而得到纯净语音谱。虽然这种方法简单有效，但缺点是残留噪声大，易引起音调失真，即所谓的音乐噪声。针对这一问题，许多学者对传统谱减法进行了改进。Berouti等人提出了一种降低音乐噪声的方法[3]，在谱减过程中引入过减和谱平滑因子，通过从带噪语音频谱中减去对噪声谱的过估计，同时避免频谱分量低于预先定义的最小值。频率自适应谱减法[4,5]是在对噪声谱过估计的基础上提出的，在一般情况下，噪声不影响均匀地分布在整个

谱域上的语音信号的频谱。然而，根据噪声的频谱特性，某些特定的语音频率范围会受到噪声的严重影响。Lockwood和Boudy针对这一问题提出了非线性谱减法(Nonlinear spectral subtraction, NSS)[4]，其中过减因子取决于每帧语音的频率。在信噪比较低的频率下减去较大的估计噪声值，在信噪比较高的频率下减去较小的估计噪声值。在非线性谱减基础之上，Jennifer C Saldanha等人提出了采用多波段谱减法(Multi-band Spectral Subtraction, MBSS)[5]对带噪语音信号进行增强，将语音频谱分成N个不重叠的频带，其中每个波段的过减因子是独立计算的。文献[6]提出了一种方法，通过将谱减语音在清音和浊音段分为多个子帧，并将每个子帧的相位随机化到一个均匀的子帧上，来分别降低清音区和浊音区的音乐噪声。

维纳滤波算法是 Hansen 和 Jensen 提出的参数估计语音增强方法[7]。Doclo 和 Moonen 进一步扩展了 Wiener 方法在多通道情况下的应用[8]。维纳滤波方法推导纯净信号在均方意义上的最优复离散傅里叶变换(Discrete Fourier Transform, DFT)系数，利用语音信号的复谱生成一个线性估值器，在噪声信号与语音信号的 DFT 系数均是独立高斯随机变量的情况下，该估计值从最小均方误差意义上视为最优。当信噪比很大时维纳滤波器不会产生语音失真，而信噪比很小时维纳滤波器的输出将会被严重衰减，使语音信号产生失真。为此学者们提出了许多改进的维纳滤波算法。学者 Nayan Modhave 等人提出了一种用于助听器语音处理的矩阵维纳滤波器算法[9]，通过增加辅助通道的数量可以提高设备的性能。其中矩阵维纳滤波器是一种考虑语音和噪声相关性的矩阵组合滤波器，其性能优于传统的多通道维纳滤波器。学者 Aishwarya Yelwande 等人提出了基于自适应维纳滤波的语音增强算法[10]，采用自适应维纳滤波器对加性噪声干扰下的语音信号进行增强，其中在时间域设计自适应维纳滤波器采用 NLMS(Normalized Least Mean Square)算法抑制语音信号中的噪声。考虑到 DCT(Discrete Cosine Transform)域噪声信号的两种状态：破坏性和构造性干扰 ，Jing Wei 等人提出了一种基于纯净语音和噪声系数的拉普拉斯-高斯混合分布模型的双增益维纳滤波器[11]用于语音增强。

基于最小均方误差 MMSE(Minimum Mean-Square Error)是 Ephraim 与 Malah 等学者于 1985 年提出的基于统计模型的语音增强方法[12]。该方法假设纯净语音信号 DFT 系数的实部与虚部都服从高斯分布。该假设对噪声信号是成立的，但对语音信号而言并不成立。因此，通常将语音信号的 DFT 系数构建为非高斯模型。Borgstrom,Bengt J 等人采用伽马分布对 DFT 系数的实部和虚部建模[13]; Ephraim 和 Y Cohen,I 假设 DFT 系数为对

方差的条件高斯分布，同时假设方差为指数分布，最后得到 DFT 系数的概率分布为拉普拉斯分布[14]；文献[15]中推导了一个噪声 DFT 系数基于拉普拉斯分布，而语音 DFT 系数基于伽马分布的估计器。相对于基于高斯分布的功率谱估计器，基于非高斯分布的功率谱估计器的处理结果分段信噪比要稍微高一点。

自适应小波变换是Yao和Zhang提出的语音增强方法[16]，专门设计用来模拟人类听觉系统。小波分解既有感知尺度特性又有适应性[17]。小波变换通过选择合适的小波分解尺度，把信号按照不同的频段进行分解，再用恰当的阈值来处理小波系数，以实现去噪功能。但是因为噪声存在时变特性，所以只是通过固定阈值来实现去噪效果明显不可靠，许多学者对这个问题进行深入研究，例如Jia H R等提出了一种基于改进小波阈值的语音增强方法[18]，相比于现有的小波阈值语音增强算法而言，能够有效提高增强语音的分段信噪比和语音质量感知评价。Talbi Mourad提出了一种融合平稳仿生小波变换(Smooth Bionic Wavelet Transform, SBWT)和最大幅度平方谱后验估计(Maximum a Posterior Estimator of Magnitude-Squared Spectrum, MSS-MAP)的语音增强算法[19]，在降噪的同时不会产生明显的音乐噪声。

子空间算法[20]是Ephraim提出的语音增强算法，基于的原理是纯净语音信号能够被看作带噪语音信号Euclidean空间中的一个子空间。与谱减法不同，子空间方法使用KLT变换(Karhunen－Loeve Transform)而不是DFT变换来分解带噪信号向量空间。KLT变换是信号相关变换，即特征向量(基)需要根据每一帧语音重新计算，因此计算效率低于DFT变换。然而，当协方差矩阵为循环矩阵时，DFT与KLT变换等价。因此，在渐进意义上，功率谱减法具有最优均方误差，且两种方法(谱减和子空间)等价。基于感知的子空间语音增强方法的主要思想是以某一种方式来对残留噪声的频谱进行整形，使其低于掩蔽阈值，从而使得残留噪声不可被人耳听见。这就需要通过一种方式将傅里叶变换域与(KLT变换)特征域联系起来。文献[21,22]都是将听觉模型集成到信号子空间的语音增强算法，基于由特征域-频域变换导出的频谱来估计掩蔽阈值，再通过频域-特征域变换而将掩蔽阈值变换回特征值域，最后将其集成到信号子空间方法中。Surendran Sudeep等人提出了一种基于听觉掩蔽特性的感知子空间语音增强方法[23]，在确定该算法的增益参数时，采用了人耳听觉系统的掩蔽特性。与现有的一些语音增强方法相比，该方法在可懂度方面具有更好的性能。

以上算法都仅仅是对语音信号的幅度谱进行估计，而令其相位谱保持不变，这是由于1982年，Wang 和Lim得出结论：相位信息在语音增强中是可以被忽略的[24]。Vary

推导出相位失真阈值和局部信噪比之间为6分贝的关系[25]，在这里，带噪语音相位可以被认为是纯净相位谱的一个不错的估计。Ephraim和Malah表明，假设相位服从均匀先验分布，那么带噪语音频谱相位是对纯净相位的最小均方误差(MMSE)估计[12]。

在同样的假设下，Lotter和Vary 表明带噪语音相位是纯净相位的最大后验(Maximum A Posteriori, MAP)估计[26]。最近的研究[27]表明，语音与所选的幅度分布无关，语音相位的MMSE估计由噪声相位谱给出。这一发现是根据Ephraim和Malah对高斯语音和噪声模型的早期观察得出的[12]。之后的一些研究陆续证明相位信息有助于提升语音信号的可懂度，推翻了先前的观点。Alsteris 和Paliwal分析了短时相位谱在语音信号处理中的应用[28]，证明了相位谱在自动语音识别、人类听力和语音可懂度方面的有用性[29]。Mowlaee 和 Kulmer对语音增强中的相位研究进行了对比估计，论述了相位估计方法的潜力和局限性[30,31]。他们的主要贡献在于从带噪语音中估计出相位，并将其用于重构语音信号。

2012年，在一次电子电气工程师会议中，Martin Krawczyk等学者根据相位估计对于改进语音信号性能的有用性而提出相位重构理论[32]，阐述了近十几年来在单通道语音增强算法中很少重视相位信息的原因，并提出通过构建语音信号浊音段的谐波模型，将时域和频域相结合从而得到重构后的相位信息。2014年，Timo Gerkmann等学者在语音信号浊音段谐波模型的基础上做出了改进，提出通过对带噪语音信号进行清浊音段的划分并在浊音段利用谐波模型重构纯净语音相位，其中阐述了使用时域和频域方法重构相位信息的条件[33]。Dang X等通过高分辨率的STFT(Short Time Fourier Transform)估计出相位谱的近似值，提出了一种语音增强的相位谱估计方法[34]，实验结果表明将相位谱补偿与传统语音增强方法相结合，能够有效且显著地提升语音质量和清晰度；2018年，Barysenka S Y等学者基于谐波信号间相位不变性、相位准不变性和双相位约束条件，提出了新的相位估计方法[35]；贾海蓉等学者则在谐波模型的基础之上，采用带噪语音与纯净语音在清音段的几何关系式来估计清音段的语音相位[36]，并结合基于二元假设模型的幅值估计来得到最终的重构语音。

近年来，使用相位信息重构语音信号的方法不断被提出，经典方法包括基于谐波模型的方法[33]，仅从基频和噪声观测中重建语音频谱相位，在一定信噪比范围内提高了语音质量，但同时引入了语音失真；未封装相位的时间平滑[37](Temporal Smoothing of Unwrapped Phase, TSUP)，通过对带噪语音的瞬时相位谱进行相位分解，然后进行时间平滑，以减小噪声相位，从而重构出增强的瞬时相位谱用于信号重构，提高了语音质量

和可懂度；具有相位约束的几何方法[38]，利用基频和相位畸变特征进行谐波增强的相位重构方法[39]，通过考虑谐波相位谱的关系来估计语音相位谱，提高了各种噪声条件下的语音质量；闭环迭代相感知的语音增强方法[40]，利用几何约束[41]和相感知幅度估计器[42]来估计相位，在降噪的同时提高了语音信号质量；Md Tauhidul Islam等应用信噪比相关方法确定相位谱上的补偿量[43]，修正带噪语音的相位谱[44]，与谱减法结合后在不同信噪比下的性能优于传统语音增强方法。

本文对传统相位谱补偿算法及相位信息重构算法进行深入研究分析，在原有理论的基础上进行新的理论创新，并使用SNR与PESQ作为客观标准进行评价，结合时域波形图与语谱图等形式以呈现最终的实验效果。

1.3 语音信号的特性和噪声特性

1.3.1 语音特性

1. 语音信号的频谱特性

语音信号具有不平稳的特性，其二阶统计量也就是功率谱是随着时间变化而不断发生变化的。然而，研究发现，语音信号具备短时平稳特性，一般认为在 10~30ms 的时间段内语音的短时谱是相对稳定的[45]，所以能够截取一小部分语音信号进行频谱分析。所谓频谱图，就是以语音信号的频率值为横轴，幅度值为纵轴，把通过短时傅里叶变换得到的信号的幅值画在其所对应的频率上。频谱图主要研究的是短时谱，因此只能反映语音信号的静态频谱特性，故而还需要研究语谱图。语谱图不仅可以清晰地反映出信号的频谱特性和时间之间的关系，而且可以通过不同的灰度条纹以及深浅不同的颜色刻画出频率在固定时间间隔内的能量强弱，颜色越深说明该点的信号能量越强。图 1-1 为选自 863 语音库中的 SP15 语音的语谱图：

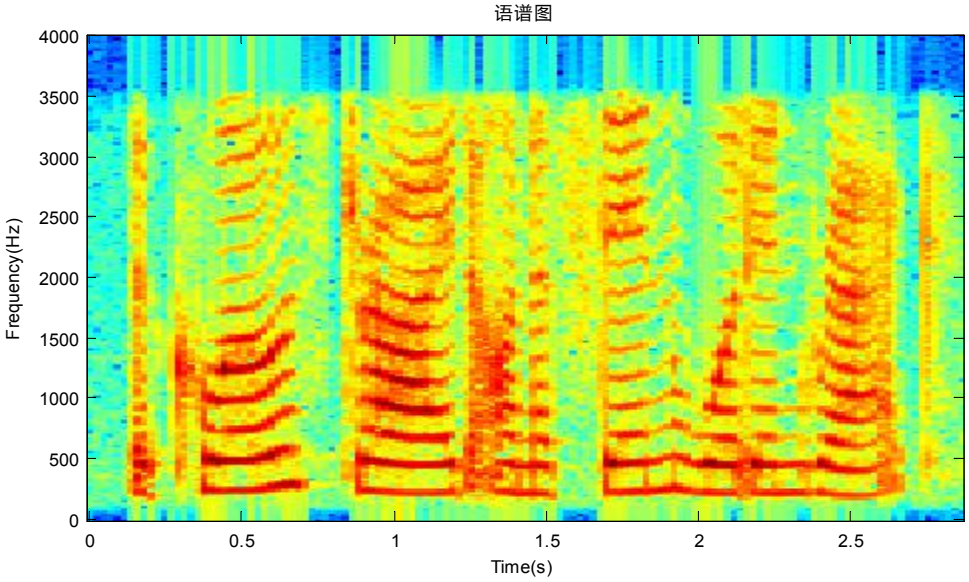


图 1-1 SP15 语音的语谱图

Fig.1-1 The speech spectrogram of SP15

2. 语音有清音和浊音之分

与滤波器的作用相似，声道通过调整声门气流的频谱从而得到不同的音色。声道的激励源是由声带所提供的，激励的周期与否由声带的状态决定。浊音(例如元音)是由声带的振动产生的，其输入为周期的声门气流序列，在频域上存在明显的共振峰结构，因此大部分能量主要集中在相对较低的频段。清音(例如辅音)是在声带不产生振动的情况下形成的，它的激励源与噪声高度相似，因此增加了区分清音和噪声的难度。

1.3.2 噪声特性

在实际生活中常会出现语音与其他声源混合的情况，把混合声音中人们所不需要的声音称为噪声。在生活中能够将噪声分成加性噪声与乘性噪声两种。加性噪声是指噪声与语音不但在时域上的干扰为相加关系，而且在频域上也为相加关系。我们生活环境中的背景噪声一般能够视作加性噪声。因为背景噪声在现实生活中普遍存在，所以对加性噪声的研究在语音增强领域中至关重要。乘性噪声是指噪声与语音在时域上是卷积关系的同时在频域上是相乘关系。麦克风传输中的电话信道和无线信道上的噪声属于此类噪声。此类噪声能够通过某种变换(同态滤波)转换成加性噪声，故能够用加性噪声的方法来处理。本文中的噪声源选自 NOIZEUS 语音库。

1.4 语音增强算法性能评估

如图 1-2 描述了语音增强系统的原理：

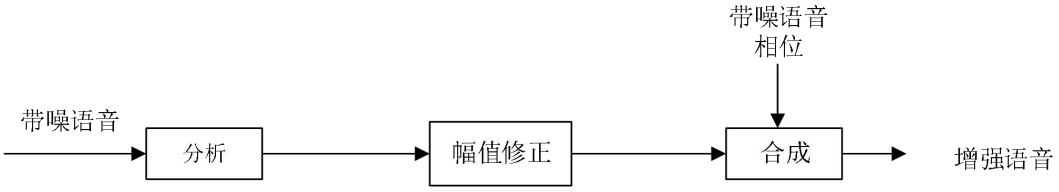


图 1-2 语音增强原理图

Fig.1-2 The principle diagram of the speech enhancement

由图 1-2 可知，在调制阶段，带噪语音谱幅值被修正，同时直接使用带噪语音信号的相位重建信号，但是考虑到带噪语音和纯净语音相位之间存在差异，故而对纯净语音信号的相位进行准确估计非常重要。如图 1-3 为相位估计算法的原理图。

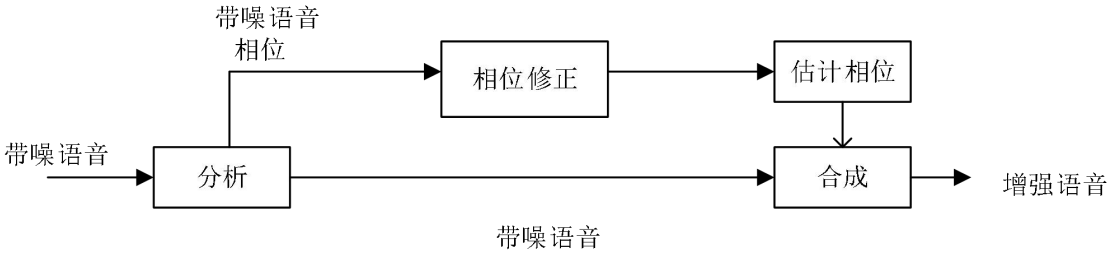


图 1-3 相位估计算法原理图

Fig.1-3 The principle diagram of the phase estimation algorithms

语音增强算法的性能评估主要考虑主观和客观两个方面。其中主观评测就是参考预先设置好的等级标准，让试听者凭借自己的主观感受及舒适度对测试的语音质量进行评价。而客观评测就是以一些评测指标为标准，量化增强语音与待测语音之间的数值“距离”，从而得出客观结论。以下将介绍一些常用的主观和客观评测方法。

1.4.1 主观评价

主观评测方法是以人的主观印象来对增强前后的语音质量进行评价，其优势是与人们听话时对语音质量的感觉相符，劣势在于进行主观评价的耗费巨大且历经时间较长，因此评判的误差较大。在主观听音测试中，试听者按照 5 分制离散量表(Mean Opinion Score, MOS 测试)或 0~100 连续量表(Diagnostic Acceptability Measure, DAM 测试)对增强语音质量进行评分[46]。

1. 5 分制离散量表(MOS)

最广泛使用的主观评价标准是由 CCITT 提出的 MOS 得分方法，MOS 使用五级评分制，又称 5 分制离散量表。其评分标准大致分为极差、差、一般、良与优五个等级。

MOS 测试包括训练阶段和评估阶段。训练阶段中，被试人员通过参考一些代表高、中等和低语音质量等级的参考信号，来均衡他们对语音质量评级的主观范围。这也就意味着在训练过程中，应该平衡全部试听人员对语音良好度的评判标准。评估阶段中，被试人员通过聆听语音信号样本且依照评分标准中的5个等级来对语音信号的质量进行评分。由于人类对声音的主观听感存在差异，因此应该对所有被试人员对语音良好度的评价结果取均值。

在数字语音通信中，MOS 测试的语音质量评测使用离散的五个等级来量化，所以试听人员一定要将对语音信号质量的主观印象用这五个等级来精确表示。

2. 0~100 连续量表(DAM)

由于 MOS 测试是一维的评价标准，单独使用 MOS评分不能准确判断语音信号对评分造成的影响和最终的打分结果。基于此，Voiers 提出了 DAM 多维音质评估方法[46]，它是在多种条件下对语音质量的接受程度的一种综合评测。

DAM 测试利用下面三个不同的尺度对语音信号质量的等级做出评测：可懂度，悦耳度与接受度。相比于 MOS 测试，DAM 测试首先要对试听者进行严格的训练，其次每次试听一定要让试听者充分聆听语音样本，也就是要给试听者提供一个参照系。它的缺点是非常浪费时间，因此多数语音增强研究并不采用主观的质量测度方法。

1.4.2 客观评价

理想的客观评测方法能够高度准确地预测通过具有正常听觉的试听者所得到的主观听音测试的结果。语音质量客观测度的计算，首先要将语音信号分割成 10~30ms 的帧，接着计算原始语音信号与经过增强后的语音信号之间的失真测度。并通过每个语音帧的失真测度的平均值来得到一个总的语音失真测度。失真测度的计算在时域上(如信噪比测度)和频域上(例如基于 LPC 谱距离测量)都能够进行。普通的客观评测方法包括：信噪比，分段信噪比，感知语音质量评估(PESQ)，坂仓距离等。

(1)信噪比(Signal-to-Noise Ratio, SNR)

信噪比[47]是衡量宽带噪声失真的语音增强的普遍方法，通常用于纯净语音和噪声都

已知的算法仿真中，其定义式为：

$$SNR = 10\log_{10}\frac{\sum_{n=0}^{M} s^2(n)}{\sum_{n=0}^{M}\left[s(n)-\hat{s}(n)\right]^2} \tag{1-1}$$

其中，$s(n)$ 代表纯净语音，$\hat{s}(n)$ 代表增强后的语音信号，M 代表帧数。

(2)因为语音信号一般是短时平稳的，所以在不同时间段内的信噪比是不同的。使用分段信噪比(SNRseg)[48]对其进行改善，具体表达式为：

$$SNR_{seg} = \frac{10\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{n=Nm}^{Nm+N-1} s^2(n)}{\sum_{n=Nm}^{Nm+N-1}\left[s(n)-\hat{s}(n)\right]^2}}{M} \tag{1-2}$$

其中，N 代表帧长。据上式可知分段信噪比是在信噪比计算的基础上对所有的语音信号帧求平均值。分段信噪比的计算存在一个缺陷，也就是在语音的静音间隙空间内信号的能量将会特别小，所以导致结果为负值，造成对分段信噪比的整体测量结果存在偏差。针对这个问题，采用限制 SNRseg 的值在[-10dB，35dB]的范围内从而避免静音帧对计算结果造成不良影响。

(3)PESQ 评价

感知语音质量评估(Perceptual Evaluation of Speech Quality, PESQ)[49]是一种针对输入-输出方式采用的典型算法，效果很好。该算法使用一个带噪语音信号与一个参考语音信号，首先把这两个需要比较的语音信号进行电平调整、输入滤波器滤波、时间对准与补偿、听觉变换，其次依次提取两路语音信号的参数，再结合其时频特征计算出 PESQ 分数，最后把该分数映射为主观平均意见分(MOS)。PESQ 使用线性评分标准，得分在-0.5~4.5 范围之内，得分越高对应其语音质量越高。其具体框图如图 1-4 所示：

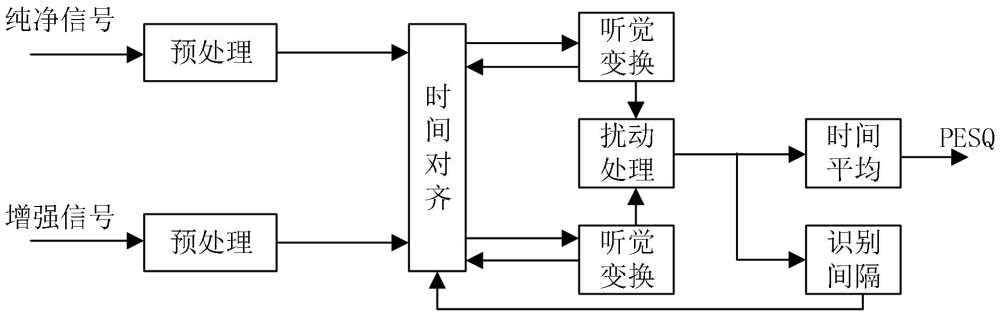


图 1-4 PESQ 流程图

Fig.1-4 Flow chart of the PESQ

(4)坂仓距离(Itakura-Saito Distance, ISD)

坂仓距离[50]测度使用的是线性预测方法，分别从纯净语音信号和增强语音信号的同步帧中得到质量评价，IS 距离需要考虑模型增益，其取值范围限制在[0-100]之间，计算公式为：

$$ISD=\sum_{n=1}^{M}1/\left\{2\left[\log_{10}\left(\frac{N}{2}\sum_{k=1}^{N/2}\frac{X(n,k)}{\hat{X}(n,k)}\times\frac{N}{2}\sum_{k=1}^{N/2}\frac{\hat{X}(n,k)}{X(n,k)}\right)\right]\right\} \tag{1-3}$$

其中，$X_{(n,k)}$和$\hat{X}_{(n,k)}$分别是对纯净语音与增强语音做出STFT 变换后所获得的第n帧第k个频率分量。

1.5 语音增强中的窗函数

离散序列的傅里叶变换的定义基于以下假设：对所有的n，离散时间序列$x(n)$均为已知。但实际上我们观察到的$x(n)$在时域上总是有限的，这相当于用一个有限时宽的函数对$x(n)$进行截短，我们把这个函数称为窗。基于语音信号的短时平稳性，我们可以认为帧长在 10~30ms 内语音信号的特性相对不变，在此基础上能够对分帧后的语音信号做相应的短时傅里叶变换。语音增强中使用的窗函数通常具有低通的性质，选择时需要考虑分辨率和频谱泄露之间的折中。在语音增强算法中较常使用的窗有矩形窗与汉明窗两种，窗函数的定义如下：

矩形窗

$$w(n)=\begin{cases}1 & 0\le n\le L-1\\ 0 & else\end{cases} \tag{1-4}$$

汉明窗

$$w(n)=\begin{cases}0.54-0.46\cos(2\pi n/(L-1)) & 0\le n\le L-1\\ 0 & else\end{cases} \tag{1-5}$$

其中L为窗长，在语音信号处理中，可以依据不同情况使用不同的窗函数。

下图 1-5 与 1-6 分别对应的是矩形窗与汉明窗的时域波形和幅频特性图。

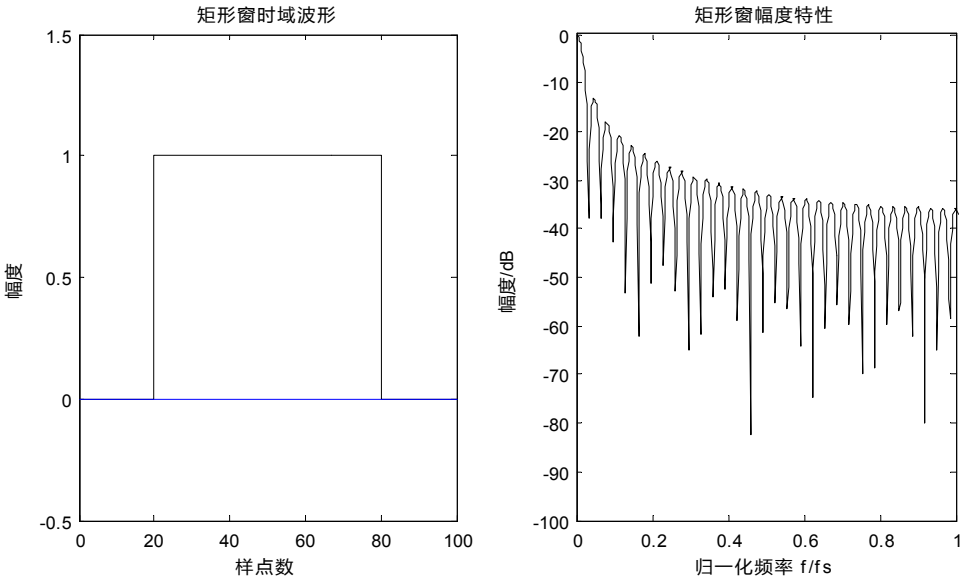


图 1-5 矩形窗及其频谱图

Fig.1-5 Rectangular window and its spectrum

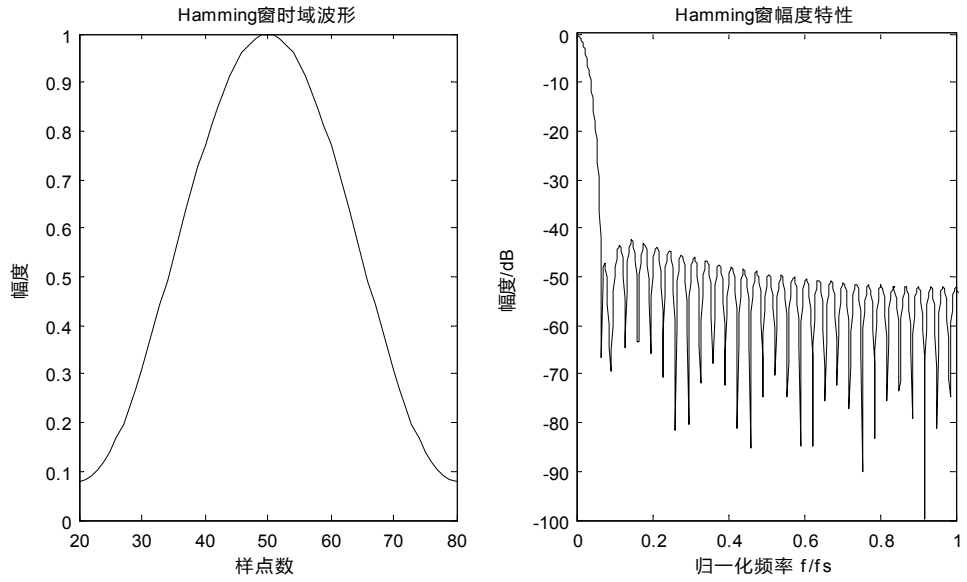


图 1-6 汉明窗及其频谱图

Fig.1-6 Hamming window and its spectrum

与矩形窗相比而言，汉明窗的主瓣宽度相对较宽，频谱分辨率相对较低，其旁瓣衰减较大，低通特性更为平滑，因此可以较好地体现短时语音信号的频率特性。

1.6 本文的主要工作

本文在传统单通道语音增强算法的研究忽略相位信息造成语音失真和可懂度低的情况下，分析了传统语音增强算法的原理，在此基础上提出了改进相位谱信息及相位重构的语音增强方法。本文的具体内容介绍如下：

第一章阐述了传统语音增强算法的研究背景与意义，并介绍了其发展现状，在此基础上对语音质量的性能评估和语音信号处理中常用的窗函数作了简单介绍。

第二章重点阐述了语音增强方法的基本理论知识，重点介绍了噪声估计、传统幅值估计及基频估计的基本原理，并对噪声估计算法中的最小值跟踪法和非平稳噪声估计法及相位谱语音增强算法中所用到的基频估计进行了相关的实验仿真。

第三章提出了一种用 *Sigmoid* 型相位谱补偿函数改进相位谱信息的语音增强方法，引入基于每帧语音信号输入信噪比的 *Sigmoid* 型相位谱补偿函数有效解决了传统相位谱补偿方法中补偿因子固定，无法对含噪语音相位谱进行灵活补偿的问题。相比传统相位谱补偿(Phase Spectrum Compensation, PSC)算法而言，改进算法可以有效抑制音频信号中的各类噪声，同时增强语音信号感知质量，提升语音的可懂度。

第四章提出了一种用信噪比信息和时频特征改进相位重构的语音增强算法，使用信噪比信息和时频特征重构谐波相位模型，解决了传统基于谐波模型的相位重构算法在提高语音质量的同时引入语音失真、导致语音可懂度低的问题。实验表明，该算法在降低语音失真的同时提高了语音的可懂度。

第五章进行总结和展望，总结了本文的研究内容及成果，并指明下一步的研究工作。

第二章 语音增强方法基本理论

2.1 噪声估计算法

在绝大多数的基于幅值信息的语音增强算法中，通常都会预先估计出带噪语音信号中的噪声谱。通过估计出的噪声谱可以计算出 MMSE 算法中语音信号的先验信噪比，也可以计算出子空间算法中噪声信号的协方差矩阵。对语音信号噪声谱的估计过小的话，会产生残留噪声；反之过高的估计语音信号的噪声谱会引起语音失真，进而导致语音信号的可懂度降低。所以，在语音增强系统中往往要对噪声进行准确估计。估计噪声最常用的方法就是话音活动检测算法(Voice Activity Detection，VAD)[51]，也就是在语音信号的清音段估计并更新出噪声功率谱，使用似然检验准则检测当前帧中语音是否存在。该方法在平稳噪声环境中效果良好，但是受到实际生活中不断变化的噪声谱特性的限制，这种方法的效果并不显著。因此本章将着重介绍最小值跟踪算法和非平稳噪声估计算法，这两种算法在噪声估计中较为常用，即使是在语音段也能够连续地估计噪声。

2.1.1 最小值跟踪算法

最小值跟踪算法[52]基于单个频带的含噪语音信号功率普遍都会衰减到噪声信号的功率水平的理论前提，故而在每个频带内都对含噪语音信号功率谱密度的最小值进行跟踪，就能够获得对该频带噪声水平的一个近似估计。在此基础上提出了两种不同的噪声估计算法：最小值统计算法(Minimum Statistics, MS)和连续频谱最小值跟踪。前者是在一个有限长窗(分析段)之内跟踪含噪语音信号功率谱的最小值，后者虽然也是跟踪最小值，但是不需要加窗。下面具体介绍一下最小值统计算法(MS)。

令 $y(n)=x(n)+d(n)$ 为含噪语音，其中 $x(n)$ 为纯净语音，$d(n)$ 为噪声，$x(n)$ 与 $d(n)$ 是统计独立的并且均值为零。通过对 $y(n)$ 的 M 个样本加窗 $w(n)$，再进行 M 点 FFT，能够把带噪语音信号从时域变换到频域：

$$Y(\lambda,k)=\sum_{m=0}^{M-1}y(\lambda M+m)w(m)e^{-j2\pi mk/M} \tag{2-1}$$

其中 λ 为帧数标记，k 为频率分量，$Y(\lambda,k)$ 是 $y(n)$ 的短时傅里叶变换。在假设语音和噪

声相互独立的条件下，带噪语音信号的功率近似等于纯净语音信号与噪声信号的功率之和：

$$|Y(\lambda,k)|^2 \approx |X(\lambda,k)|^2 + |D(\lambda,k)|^2 \tag{2-2}$$

其中$|Y(\lambda,k)|^2$、$|X(\lambda,k)|^2$和$|D(\lambda,k)|^2$分别为含噪语音、纯净语音与噪声的功率。由前面的假设可知，该方法能够根据跟踪每一帧的$|Y(\lambda,k)|^2$的最小值进而估计出噪声功率谱。通过实验得出，窗长选取 0.8~1.4s 之内可得到良好的效果。其具体流程图如图 2-1 所示：

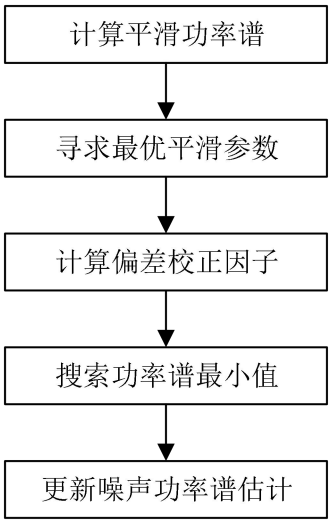


图 2-1 MS 的流程图

Fig.2-1 Flow chart of the MS

1. 计算平滑功率谱

因为含噪语音信号的功率谱$|Y(\lambda,k)|^2$随时间变化剧烈，所以通常使用一阶递归的功率谱：

$$P(\lambda,k) = \alpha P(\lambda-1,k) + (1-\alpha)|Y(\lambda,k)|^2 \tag{2-3}$$

其中α是平滑因子，取值范围是$0.7 \le \alpha \le 0.9$，λ是帧数，k是频率分量，计算的结果$P(\lambda,k)$是对含噪语音进行平滑以后的功率谱。但是仍然存在两个问题：对噪声信号因为估计偏差较大而产生的欠估计以及因为平滑因子的选择不当而产生的过估计。能够使用以下方法更加精确地进行噪声估计：(1)引入一个偏差因子来解决对噪声估计值偏低的问题；(2)引入一个随时间与频率变化而变化的平滑常数来代替一个固定值。

2. 寻求最优平滑参数

理想条件下，希望在语音活动过程中平滑参数较小，以至于更好地跟踪语音的非平

稳段。通过在语音间隙段最小化功率谱 $P(\lambda,k)$ 与噪声功率谱 $D(\lambda,k)$ 两者之间的均方误差从而得到一个最优的平滑参数，其均方误差的表示式为：

$$E\left\{P(\lambda,k)-D(\lambda,k)\middle|P(\lambda-1,k)\right\} \tag{2-4}$$

固定参数 α 可以利用一个与时频有关的平滑因子来替代，则 $P(\lambda,k)$ 的表示式为：

$$P(\lambda,k)=\alpha(\lambda,k)P(\lambda-1,k)+(1-\alpha(\lambda,k))\left|Y(\lambda,k)\right|^2 \tag{2-5}$$

令均方误差对 $\alpha(\lambda,k)$ 的一阶导数等于零，能够得到 $\alpha(\lambda,k)$ 的最优值：

$$\alpha_{opt}(\lambda,k)=\frac{1}{1+(P(\lambda-1,k)/D(\lambda,k)-1)^2} \tag{2-6}$$

3. 计算偏差校正因子

令最小值估计和在 $D(\lambda,k)=1$ 时计算出的均值估计的倒数相乘，能够对偏差进行校正，偏差校正因子的计算式如下：

$$B_{\min}=\frac{1}{E\left\{P_{\min}(\lambda,k)\middle|_{D(\lambda,k)-1}\right\}} \tag{2-7}$$

采取渐进求取的方法，通过下式对偏差因子计算近似值：

$$B_{\min}(\lambda,k)=1+(L-1)\frac{2}{Q_{eq}(\lambda,k)} \tag{2-8}$$

其中 L 为帧数，$Q_{eq}(\lambda,k)$ 与“等价自由度”成正比关系，表达式如下：

$$Q_{eq}(\lambda,k)=\frac{Q(\lambda,k)-M(L)}{1-M(L)} \tag{2-9}$$

其中 $M(L)$ 是关于帧数的函数，取值范围是 0 到 1，$Q(\lambda,k)$ 被定义是归一化方差，计算式为 $Q(\lambda,k)=4/N$，N 用来表示自由度的大小。

4. 搜索功率谱密度最小值

在长度为 L 帧的窗内搜索功率谱密度的最小值 $P_{\min}(\lambda,k)$。每当处理第 V $(V<D)$ 帧时就更新该最小值。

$$
\begin{aligned}
&\text{如果}\quad \mathrm{mod}(\lambda / M) = 0 \\
&\quad P_{\min}(\lambda,k) = \min\left\{P_{tmp}(\lambda-1,k), P(\lambda,k)\right\} \\
&\quad P_{tmp}(\lambda,k) = P(\lambda,k) \\
&\text{否则} \\
&\quad P_{\min}(\lambda,k) = \min\left\{P_{\min}(\lambda-1,k), P(\lambda,k)\right\} \\
&\quad P_{tmp}(\lambda,k) = \min\left\{P_{\min}(\lambda-1,k), P(\lambda,k)\right\}
\end{aligned}
\tag{2-10}
$$

其中 M 是连续的平滑功率谱估计值的数目，$P_{tmp}(\lambda,k)$ 是临时变量，mod 是取模运算。“如果”语句是检查帧数是否能够被 M 整除的，能整除的话就更新临时变量 $P_{tmp}(\lambda,k)$。该运算在每帧每个频点只进行一次运算即可。

5. 更新噪声功率谱估计

将公式(2-7)计算得出的偏差校正因子 $B_{\min}(\lambda,k)$ 与功率谱密度最小值 $P_{\min}(\lambda,k)$ 用于下式来更新噪声的功率谱密度

$$
D_{final}(\lambda,k) = B_{\min}(\lambda,k) * P_{\min}(\lambda,k) \tag{2-11}
$$

2.1.2 非平稳噪声估计算法

（2.1.1）中介绍的最小值跟踪算法因为不能对相应噪声功率谱的变化进行快速跟踪，并且在选取窗函数计算最小值时，与真实值存在一定延迟，对此简要介绍另外一种噪声估计算法，即非平稳噪声估计算法[53]，该算法不用在有限长的窗之内对噪声功率谱密度的最小值进行跟踪，其算法流程图如图 2-2 所示。

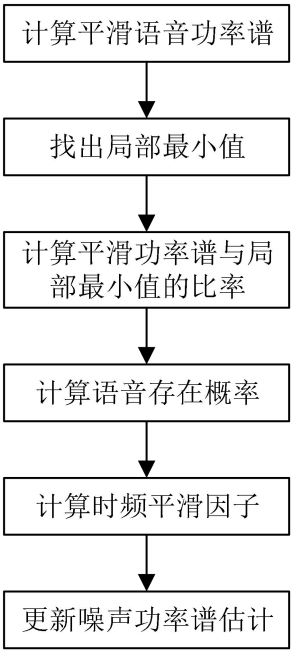


图 2-2 非平稳噪声估计算法流程图

Fig. 2.2 Flow chart of the non-stationary noise estimation algorithm

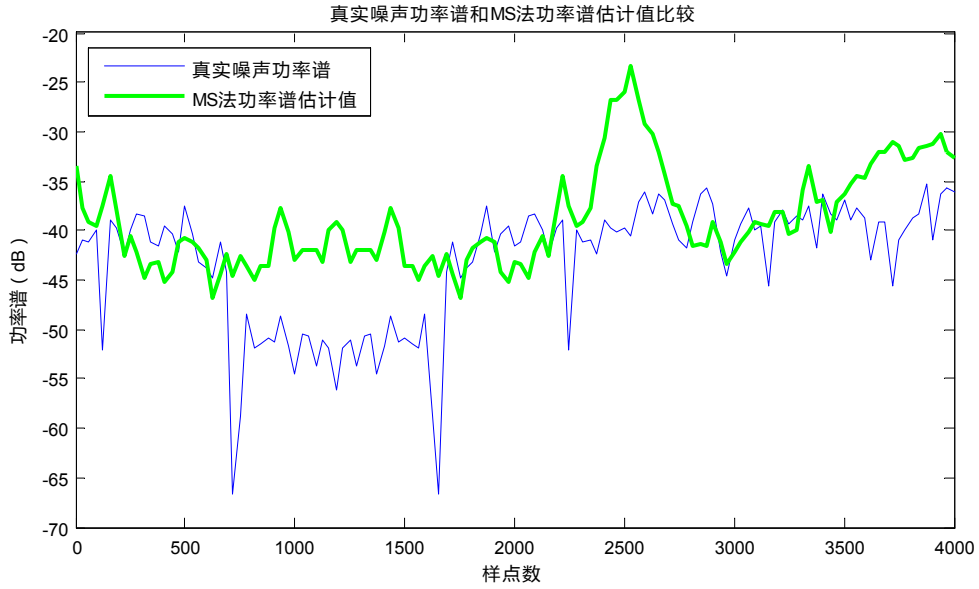
与 MS 法最大的区别在于，该方法在语音信号的各频点处都需要计算出语音存在概率，其计算式为：

$$S_r(\lambda,k) = P(\lambda,k) / P_{\min}(\lambda,k) \tag{2-12}$$

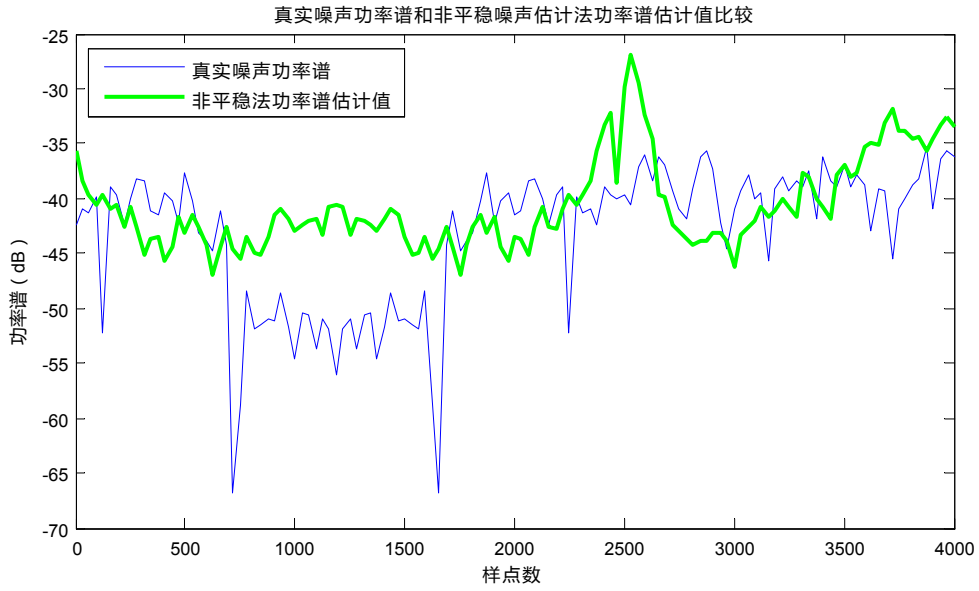
其中 $P_{\min}(\lambda,k)$ 代表噪声功率谱的局部最小值。$P(\lambda,k)$ 是对含噪语音进行平滑以后的功率谱。

将计算得到的 $S_r(\lambda,k)$ 与一个和频率有关的阈值进行比较从而得出频谱上的语音段和无语音段，在语音段和无语音段分别选取不同的阈值对噪声功率谱进行平滑。

为了证实上述两种方法的实验效果，本文采用 Noise-92 语音库的 F16 背景噪声，在信噪比为 0dB 的条件下，分别使用 MS 法和非平稳噪声估计法进行实验仿真，得到下图结果。



（a）MS法



(b) 非平稳噪声估计法

图2-3真实噪声功率谱与噪声功率谱估计值曲线图

Fig. 2-3 Estimation curve of real noise power spectrum and noise power spectrum

由上图可知，与话音活动检测算法相比而言，以上两种噪声估计算法在语音存在段都能够连续地跟踪噪声，然而非平稳噪声估计法的包络曲线与噪声的真实功率谱包络曲线更为接近。但是，在噪声信号发生急剧变化时，这两种算法都不能实时跟踪噪声。

2.2 传统的幅值估计方法

2.2.1 谱减法

谱减法是 Boll 等人最先提出的基于短时傅里叶变换的去噪方法，其算法原理是：将噪声假定为加性噪声，用含噪语音信号谱减去对噪声谱的估计值，进而计算出纯净语音信号的频谱，其中利用含噪语音的相位来近似代替纯净语音信号的相位信息。其具体原理如下所示：

假定带噪语音信号 $y(t)=s(t)+n(t)$，式子两边同时进行离散时间傅里叶变换：

$$Y(w)=X(w)+D(w) \tag{2-13}$$

将其转换为极坐标形式：

$$Y(w)=|Y(w)|e^{j\varphi_y(w)} \tag{2-14}$$

其中 $|Y(w)|$ 代表含噪语音信号的幅度谱，$\varphi_y(w)$ 代表含噪语音信号的相位谱，可以估计出纯净信号的谱为：

$$|\hat{X}(w)|=[|Y(w)|-|D(w)|]e^{j\varphi_y(w)} \tag{2-15}$$

该式说明了谱减法的基本原理，使用离散时间傅里叶变换得出含噪语音信号的幅度谱，并在语音不存在的情况下估计出噪声谱，然后用含噪语音信号的幅度谱减去所估计的噪声幅度谱，最后再通过对该差分谱求逆傅里叶变换进而得到增强语音信号。需要注意的是，由于对噪声谱的错误估计，导致增强语音信号的幅度谱 $|\hat{X}(w)|=[|Y(w)|-|D(w)|]$ 也许是负值，故需要对差分谱进行半波整流[54]，也就是把负的谱分量置零：

$$|\hat{X}(w)|=\begin{cases}|Y(w)|-|D(w)| & if\ |Y(w)|>|D(w)|\\ 0 & else\end{cases} \tag{2-16}$$

在某些情况下，使用功率谱做谱减的效果可能优于幅度谱，为了得到带噪语音信号的功率谱，将公式(2-13)中的 $Y(w)$ 和其共轭 $Y^*(w)$ 相乘，将公式(2-13)转化为以下形式：

$$\begin{aligned}|Y(w)|^2&=|X(w)|^2+|D(w)|^2+X(w)\cdot D^*(w)+X^*(w)D(w)\\&=|X(w)|^2+|D(w)|^2+2\,\mathrm{Re}\{X(w)D^*(w)\}\end{aligned} \tag{2-17}$$

谱减法的一般形式表示为：

$$|\hat{X}(w)|^p=|Y(w)|^p-|D(w)|^p \tag{2-18}$$

其中 p 是幂指数，其流程图如下所示：

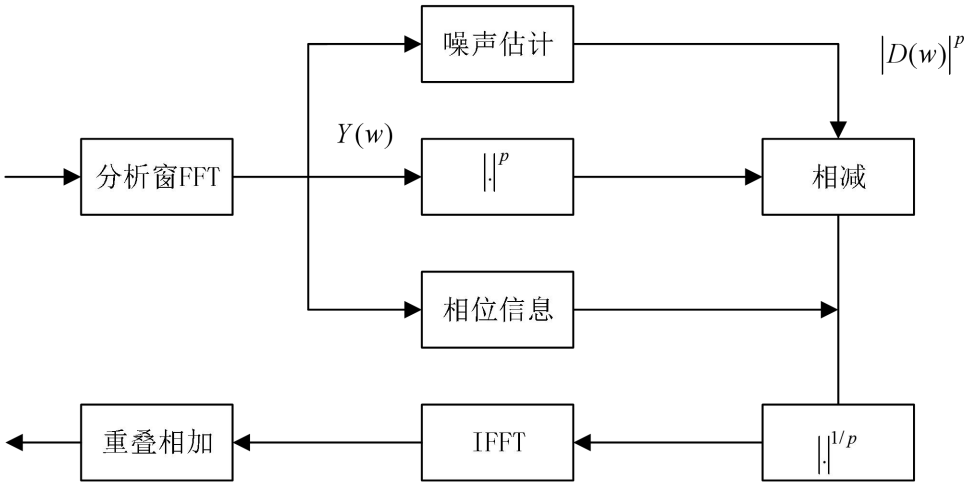


图 2-4 谱减法流程图

Fig. 2-4 Flow chart of the spectral subtraction algorithm

因为交叉项的原因，公式(2-18)只是一种近似值，而公式(2-17)中的交叉项只是在统计意义上为零，也就是假定拥有足够的数据以得到期望值且保证语音信号是平稳的。但是，语音信号通常不是平稳的，在多数应用中，语音信号是被逐帧处理的，且这些交叉项期望值不一定等于零。相比于带噪语音信号功率谱，至少在低频部分交叉项并没有小到能够被忽略的地步。谱减法的另一个缺点是使用带噪语音信号相位，在低信噪比 $(<0dB)$ 时会使语音信号变得粗糙，进而降低语音质量。针对传统的谱减法忽略交叉项和利用带噪语音信号相位近似代替纯净语音信号相位的问题，Islam M T 等人将语音谱和噪声谱之间的交叉项视为非零，通过一种改进的谱减法获得纯净语音谱的粗略估计，再通过概率方法计算出带噪语音信号相位谱的补偿量，最后结合修正后的幅度谱与相位谱进而得到增强语音信号谱。文献[55]中定义了一个修正谱减法的增益函数，具体表达式为：

$$|Z(w)|^2 = H_{MSS}(w)Y(w)^2 \tag{2-19}$$

其中 $|Z(w)|$ 是对 $X(w)$ 的估计，$H_{MSS}(w)$ 是修正谱减法的增益函数，其表达式如下

$$H_{MSS}(w) = \sqrt{\left|1 - \frac{|D(w)|^2}{|Y(w)|^2} - \chi\right|} \tag{2-20}$$

其中 χ 是 $D(w)$ 和 $X(w)$ 之间的交叉相关项，χ 能够粗略评估带噪语音信号谱相较于背景噪声的变化，其定义为：

$$\chi = \frac{(Y(w) - D(w)) \cdot D^*(w) + (Y(w) - D(w))^* \cdot D(w)}{|Y(w)|^2} \tag{2-21}$$

相比于传统谱减法，改进算法在低信噪比和中等信噪比情况下的性能显著。

2.2.2 维纳滤波法

维纳滤波器被普遍视为在均方意义上最佳的纯净语音信号谱的线性估计器。与上面的谱减法比较而言，其优点是使用维纳滤波方法处理后的背景噪声近似于白噪声而并非起伏的音乐噪声。这就说明，相比于谱减法，维纳滤波法能够更加有效抑制音乐噪声。其具体原理如下：

设含噪语音信号 $y(t) = s(t) + n(t)$，通过设计一个维纳滤波器 $h(t)$，使输入信号是 $y(t)$ 时，输出信号为：

$$\hat{s}(t) = y(t) * h(t) = \sum_{k=0}^{M-1} h(k) y(n-k) \tag{2-22}$$

对等式两边分别做DFT变换能够得到下式：

$$H(\omega) = \frac{P_{xy}(\omega)}{P_y(\omega)} \tag{2-23}$$

其中 $P_y(\omega)$、$P_{xy}(\omega)$ 分别是 $y(t)$、$s(t)$ 和 $y(t)$ 的功率谱密度与互功率谱密度，因为 $s(t)$ 和 $y(t)$ 是不相关的，能够得到：

$$P_{xy}(\omega) = P_s(\omega) \tag{2-24}$$

$$P_y(\omega) = P_s(\omega) + \lambda_n(k) \tag{2-25}$$

故最终得到维纳滤波器的增益函数表示如式(2-26)所示：

$$H(\omega) = \frac{P_s(\omega)}{P_s(\omega) + \lambda_n(k)} \tag{2-26}$$

其中 $P_s(\omega)$，$\lambda_n(k)$ 分别是语音信号与噪声信号的功率谱密度。那么增强语音信号可以表示成：

$$\hat{S}_k(\omega) = H(\omega) \cdot P_y(\omega) \tag{2-27}$$

结合先验信噪比的定义式，能够将式(2-26)改写成：

$$H(\omega_k)=\frac{\xi_k}{1+\xi_k} \tag{2-28}$$

上面所介绍的维纳滤波方法的流程图表示如下：

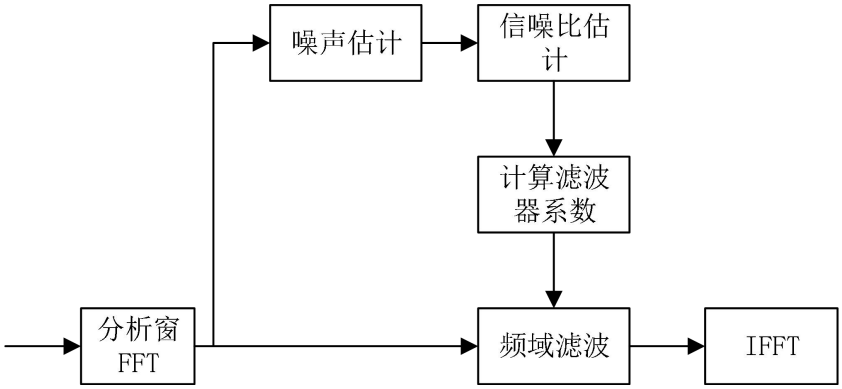


图 2-5 维纳滤波法流程图

Fig. 2-5 Flow chart of the wiener filter algorithm

由公式(2-28)可知，在$\xi_k \to 0$ 的时候，局部 SNR 极小，维纳滤波器的输出将会被严重衰减，此时$H(\omega_k)\approx 0$ ；在$\xi_k \to \infty$的时候，局部 SNR 极大，维纳滤波器不会产生语音失真，此时$H(\omega_k)\approx 1$，所以，维纳滤波器可以对频谱上 *SNR* 高的部分维持较高的权重，相反对 SNR 低的部分产生抑制。即维纳滤波器对信号所产生的抑制效果和每个频率分量处的 SNR(即ξ_k)成比例关系。

2.2.3 MMSE-LSA 法

在上一节中，维纳滤波算法增强的时域信号是通过卷积带噪语音信号和线性(维纳)滤波器得到的，相当于在频域把含噪语音频谱和维纳滤波器相乘计算出增强信号频谱。但是，线性估计器并非是对纯净信号谱的最优估计器，非线性估计反而可能产生更好的效果。非线性估计器的推导方法包括最大似然估计和贝叶斯估计，本文主要介绍贝叶斯估计中的 MMSE-LSA 算法。具体流程图如下所示：

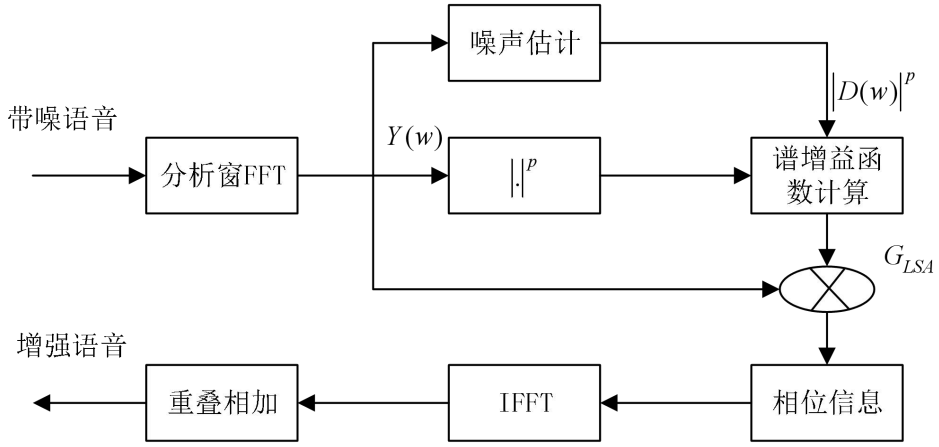


图 2-6 MMSE-LSA 算法流程图

Fig. 2-6 Flow chart of the MMSE-LSA algorithm

假设$x(t)$代表纯净语音，$v(t)$代表平稳加性高斯噪声，且$x(t)$与$v(t)$是相互独立的，计算出含噪语音信号$y(t)$的时域表示式为：

$$y(t)=x(t)+v(t) \tag{2-29}$$

令$Y_k=R_k\exp(j\theta_k)$，N_k，$A_k=S_k\exp(j\alpha_k)$分别代表对含噪语音信号，噪声和纯净语音信号做 FFT 变换后的第k个频谱分量，而 MMSE-LSA(Minimum Mean-Square Error of Log-Spectral Amplitude)幅值估计就是为了得到对S_k的估计值$\hat{S}_k$，$\hat{S}_k$的表示式为：

$$\hat{S}_k=\exp\left\{E[\ln S_k]\middle|Y_k\right\} \tag{2-30}$$

其中$E[\ln S_k]\middle|Y_k$是不可以直接计算得到的，但我们能够通过计算$\ln S_k$对条件Y_k的矩量生成函数对其进行化简。令$Z_k=\ln S_k$，则Z_k对条件Y_k的矩量生成函数是：

$$\Phi_{Z_k|Y_k}(\mu)=E\left\{\exp(\mu Z_k)\middle|Y_k\right\}=E\left\{X_k^{\mu}\middle|Y_k\right\} \tag{2-31}$$

$\ln S_k$的条件均值能够通过计算$\Phi_{Z_k|Y_k}(\mu)$对μ的导数而得到，表达式是：

$$E\left\{\ln S_k\middle|Y_k\right\}=\frac{d\Phi_{Z_k|Y_k}(\mu)\big|_{\mu=0}}{d\mu} \tag{2-32}$$

从式(2-32)能够看出，计算条件均值要先计算出矩量生成函数，计算式如下：

$$\begin{aligned}\Phi_{Z_k|Y_k}(\mu)&=E\left\{X_k^{\mu}\middle|Y_k\right\}\\&=\frac{\int_0^{2\pi}\int_0^{\infty}a_k^{\mu}p(Y_k|a_k,\theta_s)p(a_k,\theta_s)d\theta_s da_k}{\int_0^{2\pi}\int_0^{\infty}p(Y_k|a_k,\theta_s)p(a_k,\theta_s)d\theta_s da_k}\end{aligned} \tag{2-33}$$

假定噪声 $n(t)$ 是平稳高斯白噪声，那么 $p(Y_k|a_k,\theta_s)$ 和 $p(a_k,\theta_s)$ 表示为

$$\begin{aligned} p(Y_k|a_k,\theta_s) &= \frac{1}{\pi\lambda_n(k)}\exp\left\{-\frac{1}{\lambda_n(k)}|Y_k-S_k|^2\right\} \\ p(a_k,\theta_s) &= \frac{a_k}{\pi\lambda_s(k)}\exp\left\{-\frac{a_k^2}{\lambda_s(k)}\right\} \end{aligned} \tag{2-34}$$

将式(2-34)代入式(2-33)中得到：

$$\Phi_{Z_k|Y_k}(\mu)=\lambda_k^{\mu/2}\Phi(-\mu/2,1;-\nu_k)\Gamma(\mu/2+1) \tag{2-35}$$

其中 $\Gamma(\cdot)$ 是伽马函数[56]，$\Phi(a,b;c)$ 是超几何函数[57]，ν_k 的计算式是 $v_k=\frac{\xi_k}{1+\xi_k}\gamma_k$，其中 γ_k 是后验信噪比，将式子(2-35)代入式子(2-32)从而得到：

$$E\left\{\ln S_k|Y_k\right\}=\frac{\ln\lambda_k}{2}+\frac{\ln\nu_k}{2}+\frac{\int_{\nu_k}^{\infty}\frac{e^{-t}}{t}dt}{2} \tag{2-36}$$

计算出对数谱的 MMSE 增益函数的表达式是：

$$G=\frac{\xi_k}{1+\xi_k}\exp\left\{\frac{1}{2}\int_{\nu_k}^{\infty}\frac{e^{-t}}{t}dt\right\} \tag{2-37}$$

其中 ξ_k 是先验信噪比，通过下式计算得到：

$$\xi_k=\frac{\lambda_s(k)}{\lambda_n(k)} \tag{2-38}$$

其中 $\lambda_s(k)$ 是纯净信号的功率，$\lambda_n(k)$ 是噪声的功率谱。

2.3 基频估计

对于浊音而言，基音被称为声带振动的频率，高音调指的是快速振动，而低音调指的是缓慢振动。然而对于清音来说，是不会存在声带振动的。基音周期就是声带开启并闭合一次的时间，基音频率通过对基音周期计算倒数而得出。基频估计在语音信号处理中起着至关重要的作用。虽然目前已经提出了许多用于基音跟踪的算法，但当语音受到噪声的严重干扰时，这些算法并不能产生很好的效果。在噪声中进行基音跟踪存在时间的连续性和谐波模式被破坏的困难。

近年来，许多研究试图解决基音跟踪中的噪声鲁棒性问题，其中的大部分研究工作使用时间、谱或谱时间域信息估计每个语音时间帧的基音候选或基音概率。时间域方法根据基音的波形相似性原理提取基音周期；例如，Zhu,Changjia 等人提出对自相关函数法的修改，以提高基音估计的准确性[58]。谱域方法基于谐波建模，如 PEFAC(Pitch Estimation Filter with Amplitude Compression)[59]使用非线性振幅压缩和梳状滤波器抑制谱图中的噪声，并从谐波峰值中选择候选基频。谱时间法首先将信号分解成一系列子带，然后对每个频带进行时间分析。例如，Wang 和 Hansen 将语音信号分解为重叠的时频段，并推导出每个时频段的基频候选和似然得分[60]。在对基频候选和概率进行估计之后使用动态规划或隐马尔可夫模型(Hidden Markov Model, HMM)将局部基频轨迹整合到连续基频轨迹中。

本章重点介绍自相关函数法与倒谱法。

2.3.1 自相关函数法估计基频

基于自相关函数的基音估计算法(Auto Correlation Function, ACF)就是将语音信号的自相关函数的最大值的倒数作为基音频率，即

$$p = \arg\max_{f<f_{\max}} \int_0^{\infty} \left| X(f') \right|^2 \cos(2\pi f'/f)\,\mathrm{d}f' \tag{2-39}$$

其中引入参数 $f_{\max}$ 以避免积分在无穷大处有最大值。

基音频率一般在 $60-500Hz$ 范围以内，基音周期的最小值与最大值分别定义为：

$$P_{\min} = fs/500 \tag{2-40}$$

$$P_{\max} = fs/60 \tag{2-41}$$

fs 为采样率，ACF 法的原理是将初始信号与其延迟后的信号进行比较，以求出其基音频率。

2.3.2 倒谱法估计基频

语音信号 $x(t)$ 的倒谱 $c(t)$ 与它的自相关函数非常相似，唯一的区别在于它使用的是谱的对数而不是平方，即

$$c(t)=\int_{0}^{\infty}\log\left|X(f)\right|\cos(2\pi ft)df \tag{2-42}$$

倒谱估计将语音信号倒谱的最大值的倒数作为基音频率，即

$$p=\arg\max_{f<f_{\max}}\int_{0}^{\infty}\log\left|X(f^{'})\right|\cos(2\pi f^{'}/f)df^{'} \tag{2-43}$$

假设语音信号 $x(t)$、声门脉冲激励 $\mu(t)$ 与声道滤波器的倒谱分别为 $\hat{x}(t)$、$\hat{\mu}(t)$ 与 $\hat{v}(t)$，可以得到：

$$\hat{x}(t)=\hat{\mu}(t)+\hat{v}(t) \tag{2-44}$$

由上式可知在倒谱域里 $\hat{\mu}(t)$ 与 $\hat{v}(t)$ 是相对分离的，这也就意味着通过将包含基音信息的声脉冲倒谱和声道响应倒谱相互分离可以计算出基音周期。

为了证实自相关函数法的基频估计效果，本文采用 863 语音库中的纯净语音信号，以 SP01 为例，采样率是 8kHz，帧长是 256，帧移是 128，对纯净语音信号进行端点检测与基频估计，得到如图 2-7 结果。

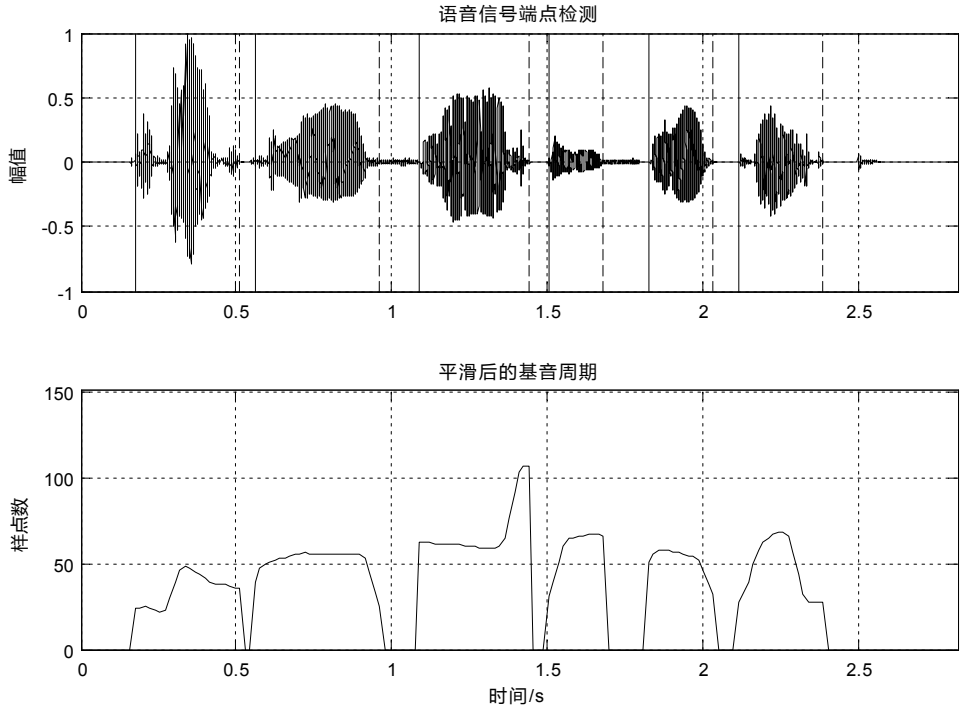


图 2-7 纯净语音基音周期图

Fig. 2-7 Pitch period diagram with clean speech

图 2-7 中垂直实线表示的是浊音段的开始，虚线表示的是浊音段的结束，由上图能够看出，自相关函数法（ACF）可以比较准确的估计出语音信号浊音段与清音段的开始和结束。

2.4 小结

由于在语音增强中噪声估计的重要性，本章先详细介绍了两种噪声估计算法的原理，包括最小值跟踪方法与非平稳噪声估计方法，并对其进行了实验验证；其次，分析了传统的基于幅值信息的估计方法如维纳滤波法和 MMSE-LSA 法的基本原理；最后，介绍了在相位重构算法中需要使用的基频估计方法，并对其进行相关的实验仿真，为后文中的研究打好基础。

第三章 改进相位谱信息的语音增强算法

3.1 引言

传统的单通道语音增强方法主要集中在从原始的带噪语音信号中减去噪声的估计量或功率，同时保持带噪语音相位信息的完整性。这是因为之前的研究显示相位谱在信噪比较高的情况下感知效果不明显。最近研究发现相位谱中也包含了很多与语音可懂度相关的信息，这些信息对于语音增强起到一定的作用[61,62] 。针对传统相位谱补偿(PSC)方法[29]中补偿因子固定，无法对含噪语音相位谱进行灵活补偿的问题，本章提出了一种改进相位谱信息的语音增强方法。首先，提出了基于每帧语音输入信噪比的 *Sigmoid* 型相位谱补偿函数；接着，结合改进 DD 的先验信噪比估计与语音存在概率算法(Probability of Speech Presence, SPP)来估计噪声功率谱；最后，在维纳滤波中结合噪声功率谱估计与相位谱补偿来获得增强语音。相比传统相位谱补偿(PSC)算法而言，改进算法可以有效抑制音频信号中的各类噪声，同时增强语音信号感知质量，提升语音的可懂度。

3.2 传统相位谱补偿算法

假定 $x(t)$ 是纯净语音信号， $v(t)$ 是平稳加性高斯噪声，且 $x(t)$ 和 $v(t)$ 两者相互独立，那么含噪语音 $y(t)$ 的时域表达式是：

$$y(t)=x(t)+v(t) \tag{3-1}$$

对式(3-1)做短时傅里叶变换，其频域表示式为

$$Y(n,k)=\sum_{m=-\infty}^{\infty} y(m)w(n-m)\exp(-j2\pi km/N) \tag{3-2}$$

其中，k 为频率索引，n 为帧索引，N 为 DFT 帧大小，$w(n)$ 为窗函数。含噪语音 $Y(n,k)$ 的极坐标形式为

$$Y(n,k)=\left|Y(n,k)\right|\exp(j\angle Y(n,k)) \tag{3-3}$$

其中，$\left|Y(n,k)\right|$ 为带噪语音信号第 n 帧第 k 个频率点的幅度谱，$\angle Y(n,k)$ 为带噪语音信号第 n 帧第 k 个频率点的相位谱。传统 PSC 方法中的相位谱补偿函数的表示式为：

$$\wedge(n,k)=\lambda\psi(k)\left|\hat{D}(n,k)\right| \tag{3-4}$$

其中，λ 是补偿因子，文献[63]中得出其最优值 $\lambda=3.14$；$\psi(k)$ 为判决函数，其表达式

为

$$\psi(k)=\begin{cases}1, & 若0.0<k/N<0.5\\ -1, & 若0.5<k/N<1.0\\ 0, & else\end{cases} \tag{3-5}$$

式(3-4)中$\left|\hat{D}(n,k)\right|$是估算出的第$n$帧第$k$个频率点的噪声幅度值。补偿后的频谱表达式为

$$Y_{\wedge}(n,k)=Y(n,k)+\wedge(n,k) \tag{3-6}$$

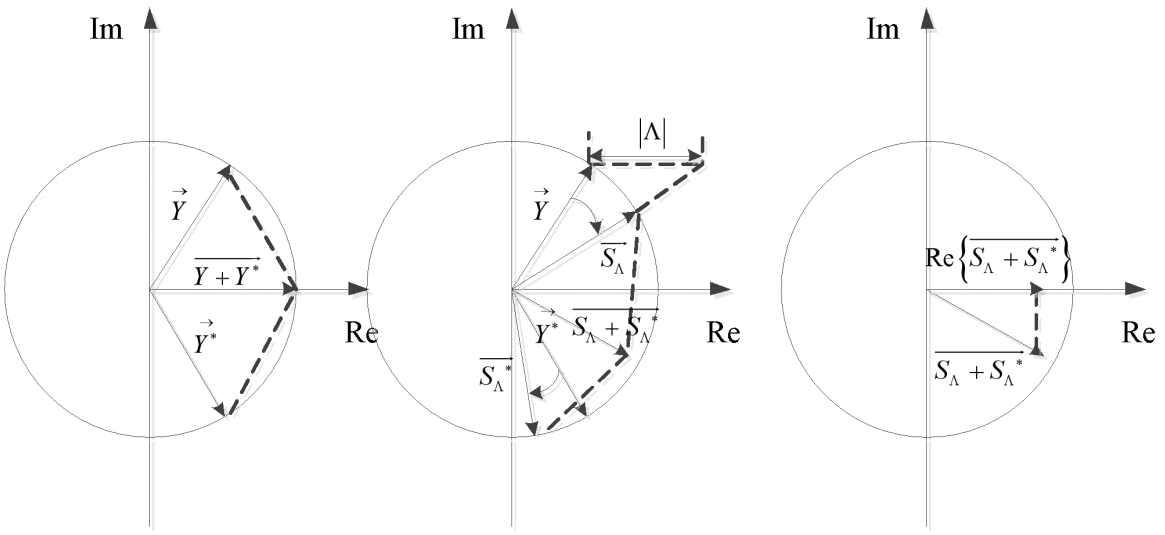
其中，$Y(n,k)$为带噪语音信号第n帧第k个频率点的频谱，$\wedge(n,k)$为相位谱补偿函数。对补偿后的频谱取相位得到相位谱

$$\angle Y_{\wedge}(n,k)=\arg\left[Y_{\wedge}(n,k)\right] \tag{3-7}$$

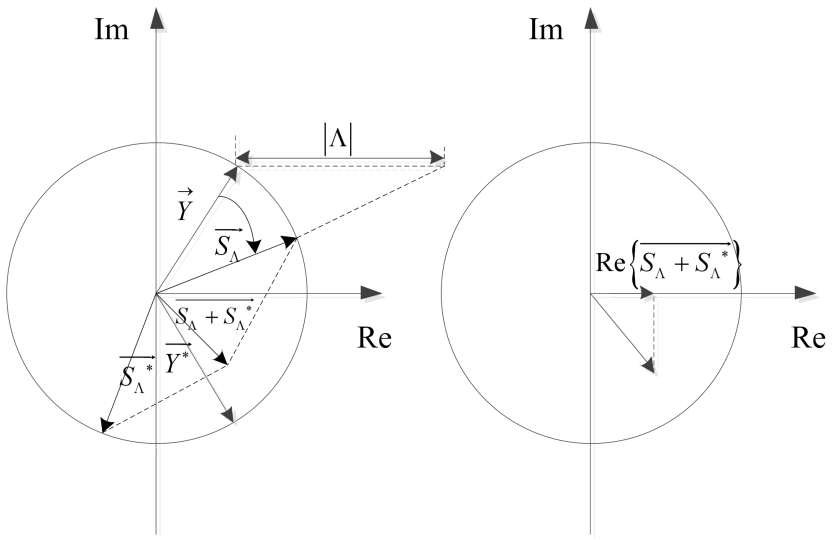
其中，arg(•)表示取复数幅角函数。结合补偿后的相位谱与经过短时傅里叶变换的幅度谱，得到语音增强后的频谱表达式为

$$S_{\wedge}(n,k)=\left|Y(n,k)\right|\exp(j\angle Y_{\wedge}(n,k)) \tag{3-8}$$

传统相位谱补偿算法(PSC)的去噪原理能够用图(3-1)来具体说明。



(a) 语音信号幅值>相位谱补偿函数



(b) 语音信号幅值<相位谱补偿函数

图3-1不同情况下去噪原理

Fig. 3-1 The denoising principle in different cases

将图(3-1)分成下面两种情况讨论：

在$\left|\vec{Y}\right|>|\Lambda|$条件下，由图可知，对两个共轭向量的矢量和进行实部运算时，其幅值并未有突出的改变。

在$\left|\vec{Y}\right|<|\Lambda|$条件下，这两个共轭向量的夹角产生了明显改变，且和向量的幅值缩小到接近于零。

由图（3-1）可知，在噪声信号幅值明显低于语音信号幅值的情况下，传统相位谱补偿算法引入相位谱补偿函数$|\Lambda|$之后的噪声去除效果不是很明显，但是在噪声幅值明显高于语音信号幅值的情况下，去噪效果尤为显著。PSC 中相位谱补偿函数的噪声去除原理正是如此。

3.3 改进相位谱信息算法

本章主要提出新的*Sigmoid* 型相位谱补偿函数、新的改进 DD 先验信噪比估计、并将新的改进 DD 先验信噪比估计应用在基于语音存在概率(SPP)的噪声估计算法中，具体算法框架如图 3-2 所示。

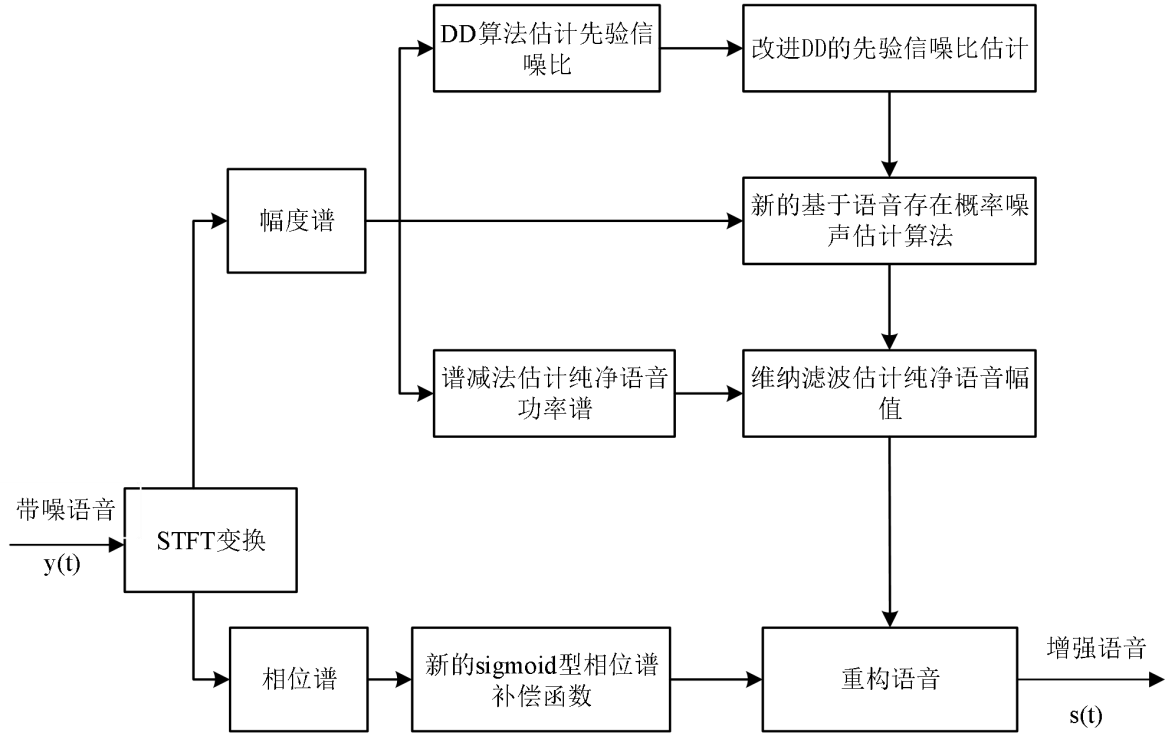


图 3-2 改进相位谱信息算法框图

Fig. 3-2 The block diagram of improving phase spectrum information

3.3.1 新的相位谱补偿函数

传统相位谱语音增强算法增强语音主要依赖于补偿函数 $\wedge(n,k)$，而补偿因子 λ 是一个固定经验值，不能随噪声变化而灵活变化，因此在实际中 PSC 算法无法获得最优值。本节对传统算法所定义的补偿函数 $\wedge(n,k)$ 中采用固定补偿因子 λ 存在的缺陷进行改进，提出一种新的相位谱补偿函数，将其设置为一个随含噪语音信噪比变化而相应变化的 *Sigmoid* 型函数，该函数的表达式为

$$\lambda_{new}=c\left[\frac{-1}{1+\exp^{-\sqrt{\frac{(|Y(n,k)|-|D(n,k)|)^2}{|D(n,k)|^2}}}}\right] \tag{3-9}$$

其中，c 为为固定经验值，取值 3.5；$|D(n,k)|$ 为噪声信号第 n 帧第 k 个频率点处的幅度谱。

将 λ_{new} 代入式子(3-4)中，得到新的相位谱补偿函数的计算式如下：

$$\wedge_{new}(n,k)=c\left[\frac{-1}{1+e^{-\sqrt{\frac{(|Y(n,k)|-|D(n,k)|)^2}{|D(n,k)|^2}}}}\right]\psi(k)\left|\hat{D}(n,k)\right| \tag{3-10}$$

带噪语音的能量基本集中于低频和中值频点处，特别是当信号为浊音帧时，因此在这些区域信噪比非常高。检测到这些区域，为了快速跟踪语音的变化，权值因子可以适当设置为较小的值。由式(3-10)可知，由于 *Sigmoid* 函数随自变量单调递增的性质，在语音存在的区域信噪比很高补偿因子会相对较小，因此可以跟踪突然的信噪比变化，对含噪语音的频谱进行补偿；反之类似。对结合式(3-9)与式(3-6)得到的频谱进行取相位运算可得新的相位谱为

$$\angle Y_{new}(n,k)=\arg\left[Y_{new}(n,k)\right]=\arg\left[Y(n,k)+\wedge_{new}(n,k)\right] \tag{3-11}$$

3.3.2 新的改进 DD 先验信噪比估计结合基于语音存在概率的噪声估计算法

为实现纯净语音信号的恢复，需要从带噪语音信号中估计出谱幅值和相位信息[64]。上述相位谱补偿算法中，式(3-8)用含噪语音幅度谱直接代替纯净语音幅度谱与补偿后的相位谱叠加，且式(3-4)中用带噪语音的幅度谱替代噪声的幅度谱[65]，并没有估计噪声幅度谱，这样会使语音增强效果不佳。因此，就需要从带噪语音信号中估计出噪声功率谱从而获得纯净语音信号的幅度谱。本章提出了一种新的改进决策导向(Decision-Directed，DD)结合基于语音存在概率的噪声估计算法，估计噪声功率谱用一个后验子带信噪比计算，基于该子带的后验信噪比的一个 *Sigmoid* 型权值用于增强听觉域中的语音频谱。具体步骤如下：

（1）根据贝叶斯公式计算出语音存在下的后验概率 $P(H_1|Y)$。用 H_1 代表语音存在，用 H_0 代表语音缺失，进而得到 $P(H_1|Y)$：

$$P(H_1|Y)=P(H_1)P(Y|H_1)/(P(H_1)P(Y|H_1)+P(H_0)P(Y|H_0)) \tag{3-12}$$

式(3-12)中假设语音存在和语音缺失的概率相等，即 $P(H_1)=P(H_0)=0.5$。由于 STFT 系数服从复高斯分布，概率 $P(Y|H_1)$ 可以采用下式计算：

$$P(Y|H_1)=\frac{1}{\pi\left|\hat{D}(n,k)\right|^2(1+\xi_{H_1})}\exp(-\frac{|Y(n,k)|^2}{\left|\hat{D}(n,k)\right|^2(1+\xi_{H_1})}) \tag{3-13}$$

$P(Y|H_0)$ 计算式同理。其中 ξ 是先验信噪比，定义为

$$\xi=\frac{P_S(n,k)}{P_N(n,k)} \tag{3-14}$$

其中，$P_S(n,k)$ 和 $P_n(n,k)$ 分别为语音的 DFT 幅度和噪声方差，n 表示帧索引，k 表示频率索引。

当语音缺失时，噪声功率根据当前含噪语音第 n 帧第 k 个频率点功率 $|Y(n,k)|^2$ 和之前噪声第 $n-1$ 帧第 k 个频率点功率的估计值 $\left|\hat{D}(n-1,k)\right|^2$ 更新，而当语音存在时则使用之前第 $n-1$ 帧第 k 个频率点的估计。文献[32]中先验信噪比是一个经验常数 $10\lg\xi_{H_1}=15dB$ 而不是随噪声变化灵活变化的值，不能跟踪突然的噪声起伏。故引入新的改进决策导向(DD)的先验信噪比估计方法，即

$$\hat{\xi}_{H1}=\frac{\mu G|Y(n,k)|^2}{|D(n,k)|^2}+(1-\mu)\left[\frac{|Y(n,k)|^2}{|D(n,k)|^2}-1\right] \tag{3-15}$$

其中 μ 为一个基于后验信噪比的 *Sigmoid* 型权值，其表达式为

$$\mu=\frac{1}{1+e^{-b\left[\frac{|Y(n,k)|^2}{|D(n,k)|^2}-1\right]}} \tag{3-16}$$

G 为增益函数，根据先验信噪比计算得出[66]，其计算式为

$$G=\frac{\xi_{H_1}}{1+\xi_{H_1}} \tag{3-17}$$

其中 ξ_{H_1} 为 DD 算法估计的先验信噪比。DD 算法是由 Ephraim 与 Malah 提出的基于前一帧语音先验信噪比和当前帧语音后验信噪比估计的加权和的决策导向(Decision-Directed)方法。其表达式为：

$$\xi_{H_1}=\alpha\frac{|Y(n-1,k)|^2}{|\hat{D}(n,k)|^2}+(1-\alpha)\max\left[\frac{|Y(n,k)|^2}{|\hat{D}(n,k)|^2}-1,0\right] \tag{3-18}$$

其中α为时频相关平滑因子，取$\alpha=0.5$为其经验值。结合式(3-12)、式(3-13)与式(3-15)，计算得出一种新的语音存在下的后验概率，即

$$P(H_1|Y)=\left[1+(1+\hat{\xi}_{H_1})\exp\left[-\frac{|Y(n,k)|^2}{|\hat{D}(n,k)|^2}\frac{\hat{\xi}_{H_1}}{1+\hat{\xi}_{H_1}}\right]\right]^{-1} \tag{3-19}$$

(2)对噪声功率谱进行初步估计

$$|\hat{N}|^2=P(H_0|Y)|Y(n,k)|^2+P(H_1|Y)|\hat{D}(n,k)|^2 \tag{3-20}$$

当$PH_{1_{mean}}>0.9$时

$$P(H_1|Y)=PH_{1_{mean}} \tag{3-21}$$

其中$PH_{1_{mean}}=(1-I)*PH_{1_{mean}}+I*P(H_1|Y)$

I为语音存在决策，其表达式为

$$I=\begin{cases}1, P(H_1|Y)>mean(P(H_1|Y))\\ 0, P(H_1|Y)<mean(P(H_1|Y))\end{cases} \tag{3-22}$$

(3)最后，更新噪声功率谱估计

$$|\hat{D}(n,k)|^2=\beta|\hat{D}(n-1,k)|^2+(1-\beta)|N(n,k)|^2 \tag{3-23}$$

其中，β是平滑系数，选取 0.9 为其经验常数。通过以上步骤计算出每一帧噪声信号的功率谱估计值$|\hat{D}(n,k)|^2$。

为了评估该基于语音存在概率噪声估计方法的性能，将其和传统方法进行比较。实验中，语音分帧后的第 1 帧语音的后验概率初始值为$10\lg\xi_{H_1}=15dB$，之后结合式(3-19)和式(3-22)对噪声功率谱进行初步估计，再由式(3-23)对噪声功率谱进行计算并更新。选自 863 语音库的纯净语音，Noise-92 语音库的 F16 噪声用于测试对比，以 SP01 为例，含噪语音输入信噪比设置为 5dB，用噪声功率谱估计值曲线图来对噪声估计方法的性能

进行评价如图 3-3 所示。

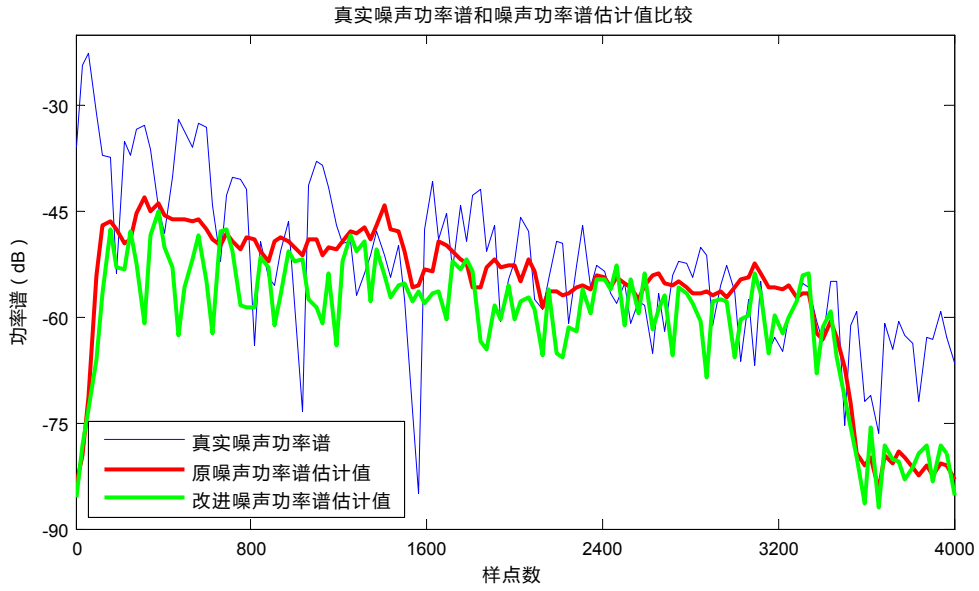


图3-3真实噪声功率谱与估计噪声功率谱比较图

Fig. 3-3 Comparison of the really noise power spectrum and noise-estimated power spectrum

由图可知，不同于传统的语音存在概率噪声估计方法，新的算法在每个频率点根据语音输入信噪比计算语音存在的先验概率而不是使用一个固定值。在噪声急剧变化时仍能够对其进行实时跟踪，相比传统的 SPP 估计方法而言，整体谱包络与原噪声功率谱更为接近。

3.3.3 改进相位谱信息算法

在维纳滤波方法中结合新的 *Sigmoid* 型相位谱补偿算法与新的基于语音存在概率(SPP)的噪声估计算法得到改进相位谱信息算法。本文结合文献[67]中估计出的纯净语音信号幅度谱$\left|\hat{S}_k(n,k)\right|$的方法和式(3-11)中新提出的补偿相位谱，得出在频域上增强后的语音信号为

$$S(n,k)=\left|\hat{S}_k(n,k)\right|\exp(j\angle Y_{new}(n,k)) \tag{3-24}$$

对式子(3-24)作傅里叶逆变换，得到最终增强后的时域信号为

$$s(t)=T_{IFFT}(S(n,k)) \tag{3-25}$$

3.4 仿真与实验

实验中从Noise-92语音库选择的三种不同的背景噪声用于测试，其中包括：白噪声、

F16噪声和Pink噪声。带噪语音信号的输入信噪比设置成0dB、5dB、10dB。以SP15为例，采样率取8000Hz，选用汉宁窗为其窗函数，帧长取256，帧移取64。使用信噪比和感知语音质量评估测度来对增强语音进行评价。以白噪声为例，将信噪比设置成5dB进行测试，实验结果如图3-4所示。

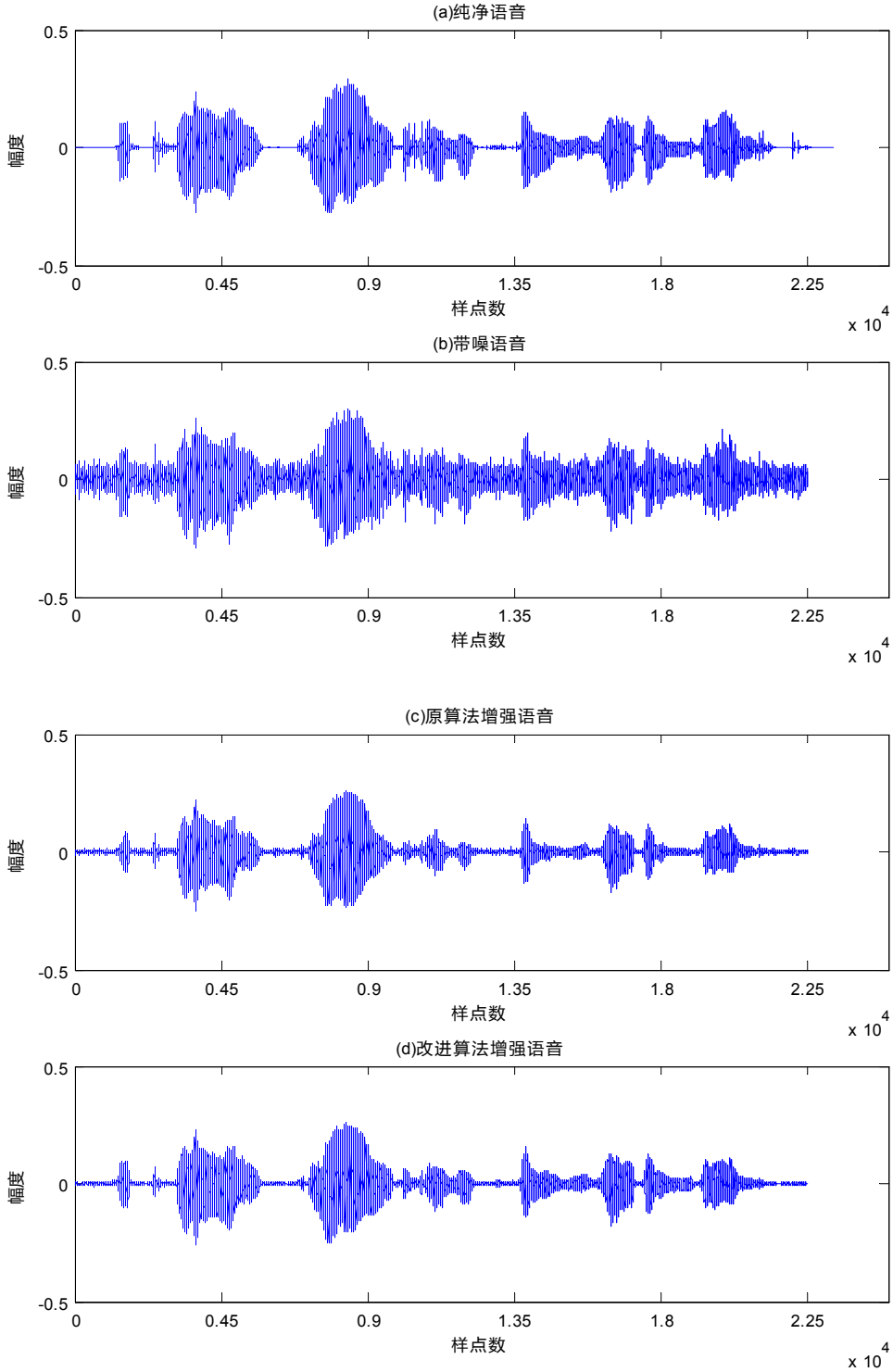


图3-4 两种语音增强算法的去噪结果波形对比图

Fig.3-4 Waveform comparison of denoising results two speech enhancement algorithms

从上述两种语音增强算法的结果对比图能够看出，使用改进方法相比传统PSC方法而言，在有效抑制音频信号中噪声的同时减少了失真，尤其是在非语音段效果更加显著。表3-1、表3-2、表3-3和表3-4分别列出了SP15和SP01语音在信噪比是0dB、5dB、10dB时的SNR和PESQ结果。

表 3-1 两种语音增强算法的信噪比结果（SP15）

Tab3-1 Signal to noise ratio results of two speech enhancement algorithms（SP15）

噪声类型	输入信噪比	PSC	改进算法
White噪声	0	7.1989	8.7572
	5	9.8398	12.2676
	10	11.4855	14.4070
Pink噪声	0	6.9569	8.5326
	5	9.6667	12.1591
	10	11.3984	15.8556
F16噪声	0	7.2382	8.7839
	5	9.7113	12.3283
	10	11.2432	16.3669

表 3-2 两种语音增强算法的信噪比结果（SP01）

Tab3-2 Signal to noise ratio results of two speech enhancement algorithms（SP01）

噪声类型	输入信噪比	PSC	改进算法
White噪声	0	8.2020	8.5529
	5	11.7512	12.6924
	10	14.7867	15.7301
Pink噪声	0	7.0305	8.1734
	5	11.0823	12.1801
	10	14.5539	15.9639
F16噪声	0	7.9758	8.6334
	5	11.6561	12.7130
	10	14.7254	16.7218

表 3-3 两种语音增强算法的 PESQ 结果（SP15）

Tab3-3 PESQ results of two speech enhancement algorithms（SP15）

噪声类型	输入信噪比	纯净SP15语音	含噪语音	PSC	改进方法
White噪声	0	4.5000	1.2958	1.7163	1.9205
	5	4.5000	1.6080	2.0634	2.4073
	10	4.5000	1.9649	2.4570	2.6583
Pink噪声	0	4.5000	1.3941	1.8265	1.9473
	5	4.5000	1.7445	2.2471	2.3743
	10	4.5000	2.1169	2.6690	2.7528
F16噪声	0	4.5000	1.4204	1.9260	2.0024
	5	4.5000	1.7833	2.3722	2.4593
	10	4.5000	2.1669	2.7863	2.8881

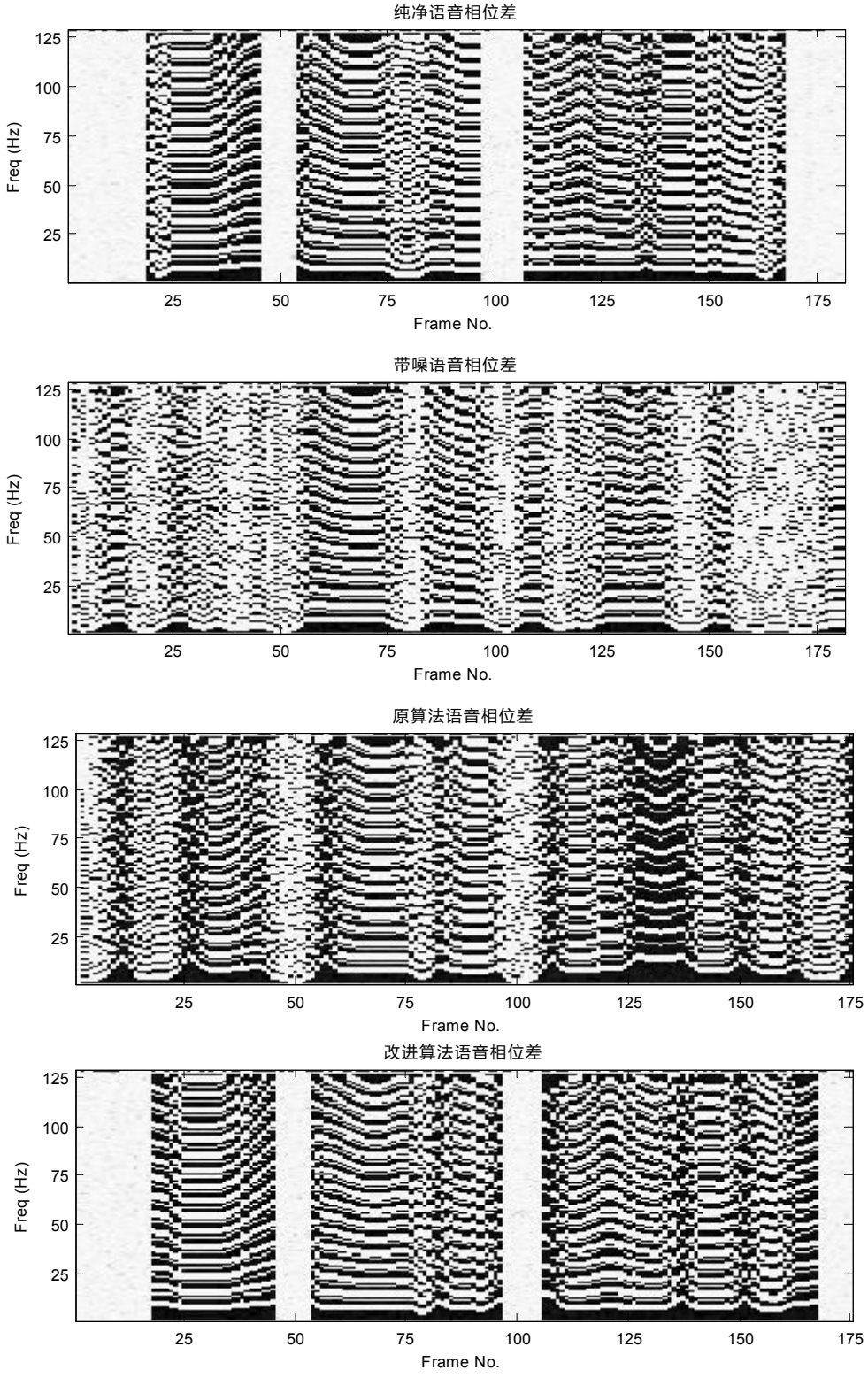
表 3-4 两种语音增强算法的 PESQ 结果（SP01）

Tab3-4 PESQ results of two speech enhancement algorithms（SP01）

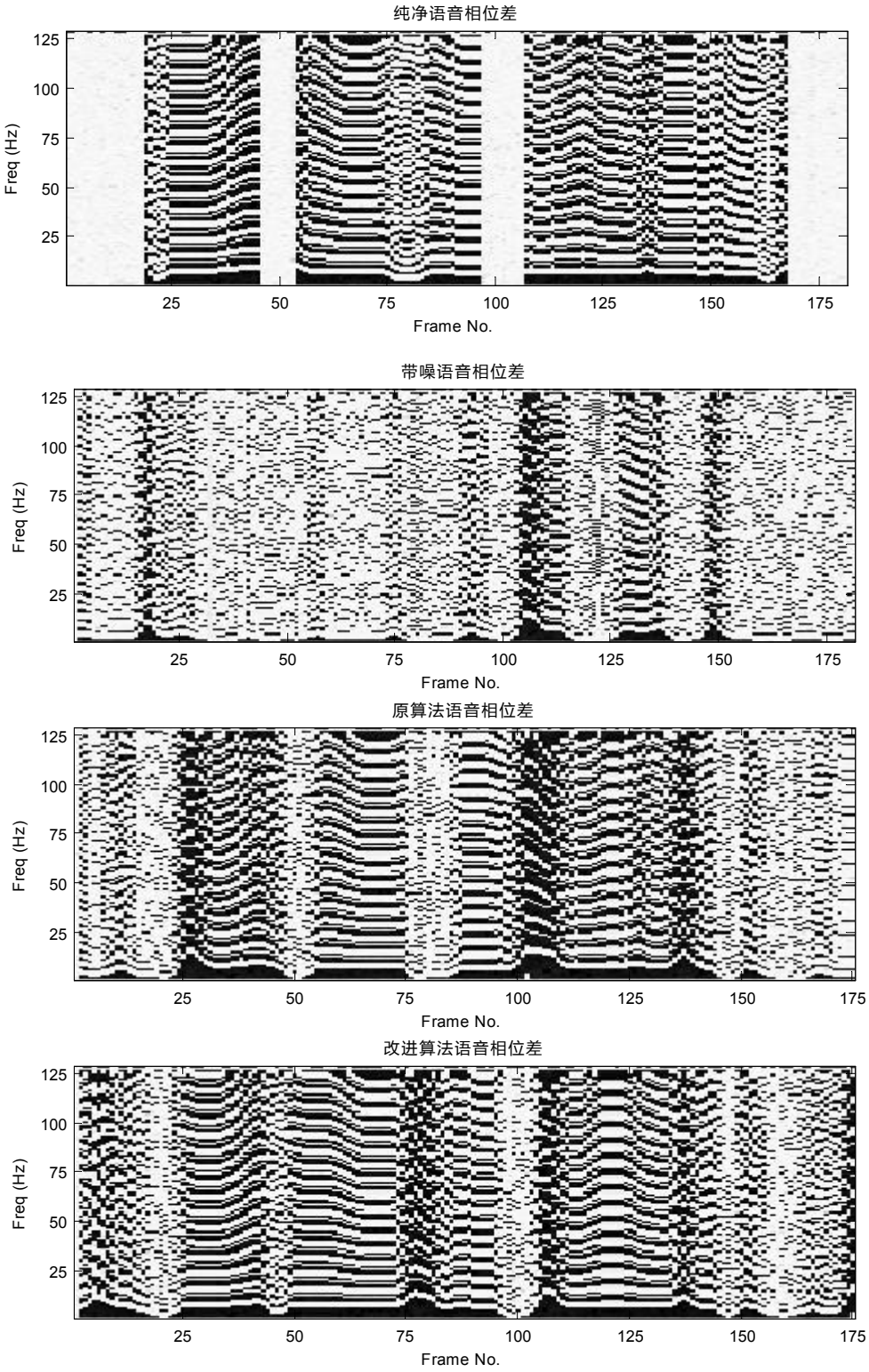
噪声类型	输入信噪比	纯净SP01语音	含噪语音	PSC	改进方法
White噪声	0	4.5000	1.4564	1.7446	1.8119
	5	4.5000	1.7604	2.0865	2.3211
	10	4.5000	2.0754	2.3696	2.6178
Pink噪声	0	4.5000	1.6199	2.0151	2.0376
	5	4.5000	1.9546	2.2937	2.3368
	10	4.5000	2.2874	2.6026	2.6394
F16噪声	0	4.5000	1.7457	2.0201	2.0486
	5	4.5000	2.0393	2.3597	2.3501
	10	4.5000	2.3537	2.6759	2.6847

结果表明，通过本文提出的方法，增强后语音的感知质量和可懂度在各种 SNR 不同的背景噪声下明显提高。其中 SP01 语音的信噪比平均提高了 1.07dB，PESQ 平均提高了 0.08，SP15 语音的信噪比平均提高了 2.74dB，PESQ 平均提高了 0.15。

为了更准确地比较本章改进方法与传统相位谱补偿方法，实验以SP15语音为例，选取白噪声与Pink噪声为背景噪声，在信噪比设置成5dB的条件下，得到相位图的对比结果如图3-5所示。如图3-6是白噪声、F16噪声和Pink噪声语谱图的对比结果。



(a) 白噪声



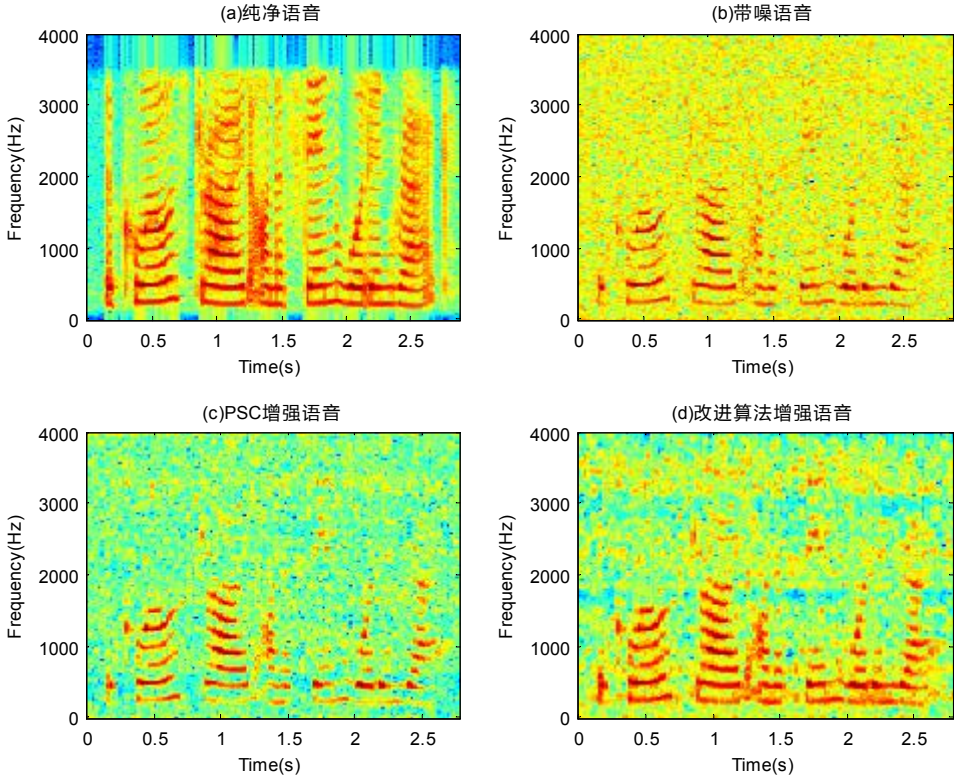
(b) pink噪声

图3-5 两种语音增强算法的相位对比图

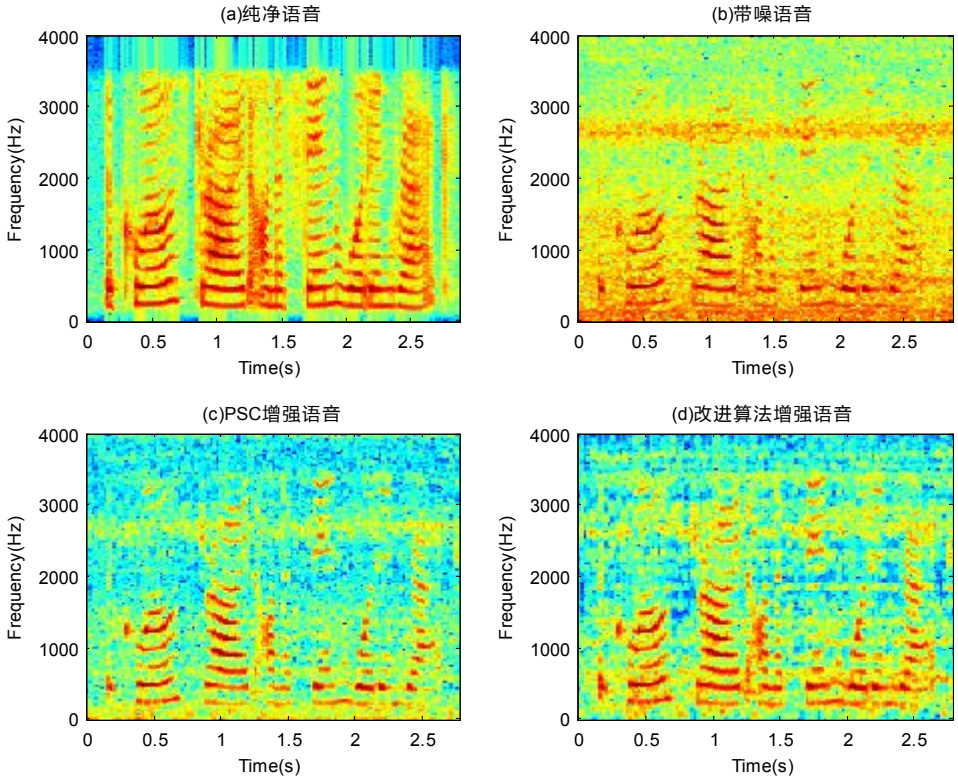
Fig. 3-5 Phase comparison of denoising results of two speech enhancement algorithms

从相位对比图可以看出，纯净语音的相位差结构清晰显著，然而引入背景噪声之后的相位差结构变得模糊不清，与原算法相比而言，改进算法的语音相位差结构更为清晰

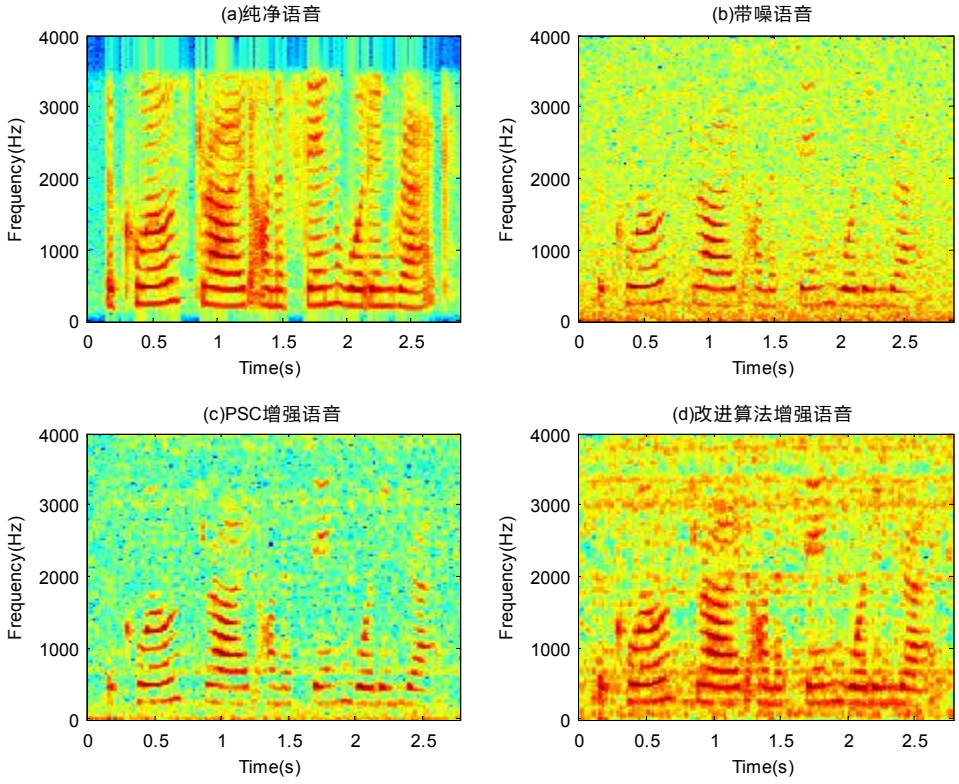
明显，且更加接近于纯净语音的相位差结构，进一步验证了本章所提出的改进相位谱信息算法在改善语音谐波结构方面的有效性。



(a)白噪声



(b)F16噪声



(c)Pink噪声

图3-6 两种语音增强算法的去噪结果语谱对比图

Fig.3-6 Spectrum comparison of denoising results of two speech enhancement algorithms

根据语音信号的语谱图能够看出纯净语音信号所对应的谱线非常规则且清楚，但是在引入背景噪声之后，纯净语音本来的谐波结构因为噪声的干扰而变得不清晰。传统的相位谱补偿方法能够在某种程度上增强语音信号，但是与最初纯净语音信号的频谱结构相较而言，却缺失了不少重要信息，在使用本章所改进的相位谱信息算法之后语音部分的谱线更为完整清晰，增强后的相位谱恢复了带噪语音信号中丢失的纯净相位的谐波结构，对语音信号的增强效果也更加明显。

3.5 本章总结

针对在语音增强中传统相位谱补偿方法使语音主观感知质量改善受限问题，本文提出了一种新的随噪声急剧变化而灵活变化的 *Sigmoid* 型相位谱补偿因子来改进传统 PSC 算法。同时，提出了一种新的改进先验信噪比的语音存在概率噪声估计方法，将其应用在维纳滤波算法中获得纯净语音幅度谱并与新的相位谱结合。大量实验表明，该算法在不同信噪比条件下，相比原算法语音质量显著改善，同时语音可懂度也有了明显提升。

第四章 改进基于谐波模型的相位重构算法

4.1 引言

上一章里介绍了一种改进相位谱信息的语音增强方法，该方法是通过对语音信号频谱进行补偿，再使用取相位运算计算出增强后的相位信息。然而随着纯净语音相位估计的准确性对抑制噪声的影响越来越重要，在语音增强中单独使用相位估计重建信号改进性能的方法也在不断被提出，在尽可能接近带噪语音信号感知质量的情况下增强语音，改进了常规幅值估计方法的性能[68,69]。Krawczyk M 等人提出基于谐波模型的方法[33]，仅从基频和噪声观测中重建语音频谱相位，在一定信噪比范围内提高了语音质量，但同时引入了语音失真，导致语音信号的可懂度较低。针对该问题，本章考虑了相位失真对语音失真和可懂度的影响，提出利用信噪比信息和时频特征重构谐波相位模型，用改进二元假设模型的 MMSE-LSA 估计纯净语音幅值谱，并将重构的相位与估计的幅值相结合进行语音增强，得到信噪比高，失真较小的增强语音信号。在介绍改进算法之前，先具体介绍一下传统基于谐波模型的相位重构算法。

4.2 传统基于谐波模型的相位重构算法

如图 4-1 是纯净信号的相位谱信息图，由图中能够看出，相位信息与频谱信息不同，它没有明显的谐波结构，因此要研究语音信号的相位信息，必须要将该信息可视化，也就是把每一帧信号的相位差值利用频谱图显示出来，如图 4-2 所示，从图中能够看出相位信息能够反映出语音的谐波结构，即相位信息有助于提高语音增强的效果。

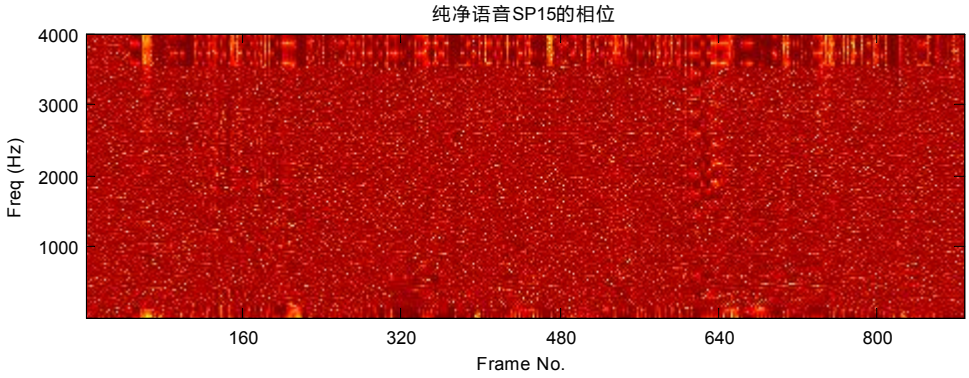


图 4-1 纯净语音 SP15 的相位图

Fig. 4-1 Phasegram of the clean speech SP15

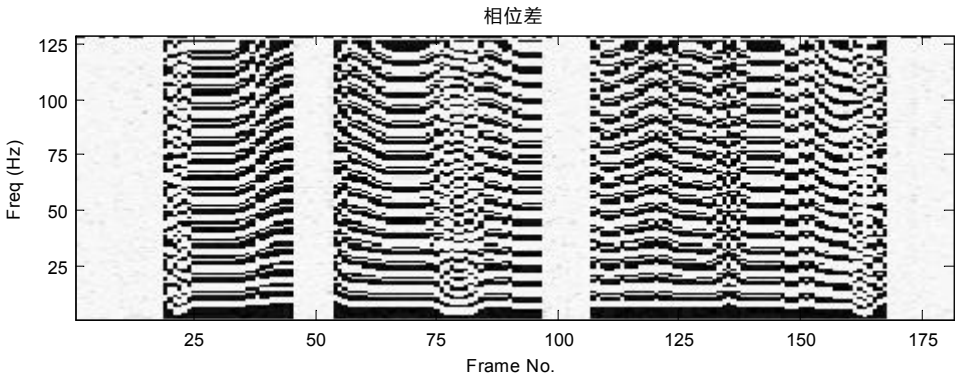


图4-2相位差

Fig.4-2 Diagram of the phase shift

语音信号 $y(n)$ 由纯净语音 $s(n)$ 和平稳加性高斯噪声 $d(n)$ 组成，且 $s(n)$ 与 $d(n)$ 相互独立，则 $y(n)$ 的时域表达式为：

$$y(n)=s(n)+d(n) \tag{4-1}$$

对上式进行 STFT 变换得到频域的表达式：

$$Y(k,l)=\sum_{n=0}^{N-1}y(lL+n)w(n)e^{-j\Omega_k n}=\left|Y(k,l)\right|e^{j\phi_Y(k,l)} \tag{4-2}$$

其中 k 为频带数，l 为帧数，N 为离散傅里叶变换长度，L 为帧移数目，$\Omega_k=2\pi k/N$ 为归一化角频率，$w(n)$ 为分析所需的窗函数。$\left|Y(k,l)\right|$ 表示带噪语音幅值，$\phi_Y(k,l)$ 表示带噪语音相位。采用 PEFAC 算法[59]进行清浊音分段，得到带噪语音浊音段的近似谐波模型，对其做 STFT 变换得到

$$Y_{k,l}\approx\ A_{h,l}^{k}\left|W_{k-\kappa_{h,l}^{k}}\right|\exp(j(\psi_{h,l}^{k}+\Omega_{h,l}^{k}lL+\phi_{k-\upsilon_{h,l}^{k}}^{W})) \tag{4-3}$$

其中 $\Delta\phi_{k,l}^{S}=princ\left\{\phi_{k,l}^{S}-\phi_{k,(l-1)}^{S}\right\}\approx princ\left\{\Omega_{h,l}^{k}L\right\}$ ，$\phi_{k-\upsilon_{h,l}^{k}}^{W}$ 为窗函数的相位，$A_{h,l}^{k}$ 为谐波幅值，$\psi_{h,l}^{k}$ 为谐波相位，。$\upsilon_{h,l}^{k}=\dfrac{N}{2\pi}\Omega_{h,l}^{k}\in[0,\mathrm{N})$ 将谐波频率 $\Omega_{h,l}^{k}$ 映射到指数形式。

对浊音段每帧频带根据能量的不同分类进行讨论：

如果经过短时傅里叶变换后信号的能量大致集中在谐波谱上，那么能够使用时域的方法计算相位：

$$\phi_{k,l}^{S}=\ princ\left\{\phi_{k,(l-1)}^{S}+\Omega_{h,l}^{k}L\right\} \tag{4-4}$$

其中 $\phi_{k,(l-1)}^{S}$ 是前一帧的相位信息，初始值能够使用带噪语音信号的相位值近似代替。

如果经过短时傅里叶变换后信号的能量很小，那么就不能把带噪语音信号的相位信息作为初始值，此时使用频域的方法计算相位：

$$\phi_{k',l}^{S}=princ\left\{\phi_{k'}^{S}-\phi_{k'-\kappa_{h}^{k}}^{W}+\phi_{k'-\upsilon_{h}^{k}+i}^{W}\right\} \tag{4-5}$$

在式（4-5）中，必须先求取窗函数的相位。窗函数在频域中的计算式为：

$$W(\Omega)=\sin(\frac{M}{2}\Omega)e^{-j\frac{M-1}{2}\Omega}\left[a\frac{1}{\sin(\frac{1}{2}\Omega)}-\frac{1-a}{2}(\frac{\exp(-j\frac{\pi}{M})}{\sin(\frac{1}{2}(\Omega-\frac{2\pi}{M}))}+\frac{\exp(j\frac{\pi}{M})}{\sin(\frac{1}{2}(\Omega+\frac{2\pi}{M}))})\right] \quad (4\text{-}6)$$

其中$W(\Omega)$为窗函数的频域式。本文采用汉宁窗，其中$a=0.5$，对上式计算相位就能够得到窗函数的相位值，将其代入公式(4-5)得到相位信息在频域中的计算式。其原理图如下所示。

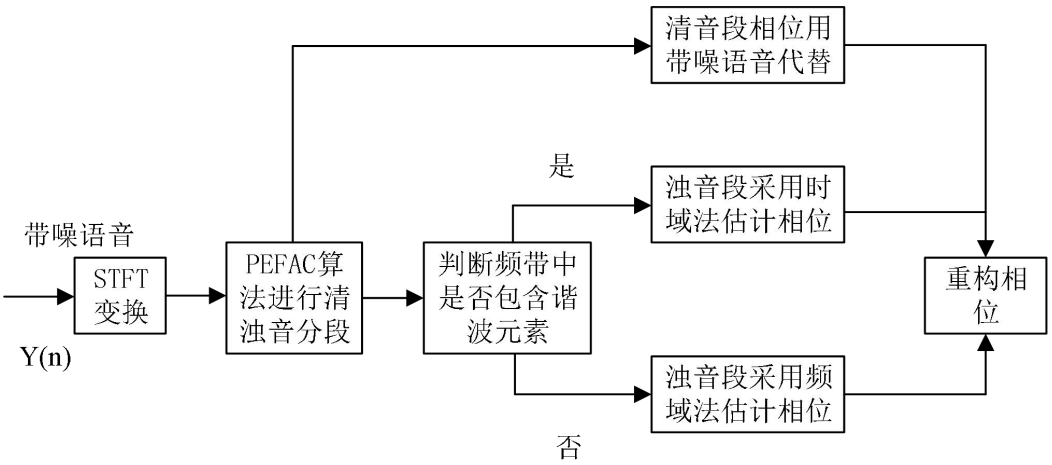


图4-3基于谐波模型的相位重构原理图

Fig.4-3 Flow chart of the Harmonic Model-based phase reconstruction

4.3 用信噪比信息与时频特征改进基于谐波模型的相位重构算法

传统的基于谐波模型的相位重构算法中，相位重构是只对浊音段的相位信息进行重构，而清音段用带噪语音信号的相位来近似，虽提高了语音质量，但因为没有考虑清浊音过度段的问题，所以导致语音不连贯(同时引入了语音失真)，针对此，本文提出了用信噪比信息与时频特征改进基于谐波模型的相位谱重构方法，并把它应用在语音增强中。具体为：首先，在对带噪语音用 PEFAC 算法[59]进行基频估计和清浊音分段的基础上，引入与相位失真相关的时频特征计算决策阈值、同时使用信噪比信息计算相位偏差；然后，将相位偏差与决策阈值进行比较用于估计清音段和浊音段的语音相位；最后，结合重构的语音相位与改进二元假设模型的 MMSE-LSA 语音幅值估计，得到增强语音。用信噪比信息与时频特征改进基于谐波模型的相位信息重构算法原理框图如图 4-4 所示。

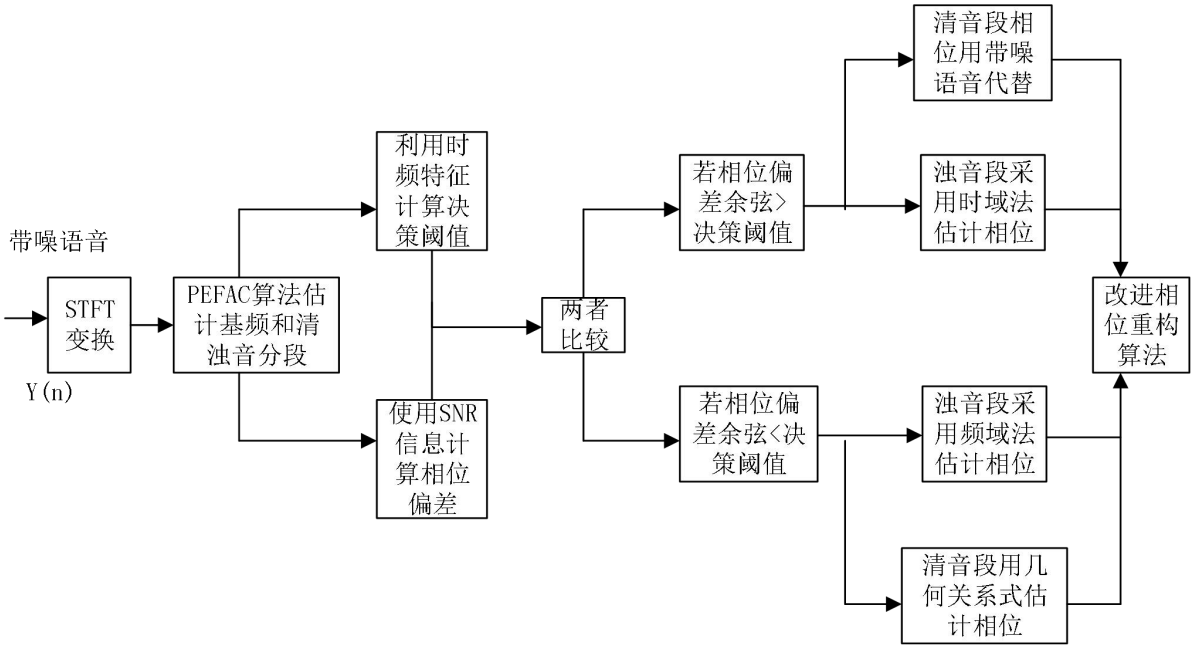


图4-4改进算法原理图

Fig.4-4 Flow chart of the proposed method

4.3.1 PEFAC 法区分清浊音

改进基于谐波模型的相位信息重构算法的第一步是对含噪语音信号进行清浊音分段，在此基础上才能够进行接下来的相位重构，本章选取在噪声条件下依然有着良好鲁棒性的基频估计算法 PEFAC 进行清浊音的分段。对语音信号清浊音的分段采用 2 元素特征矢量，具体包括以下参数：

（1）归一化时帧谱的对数平均功率：

$$L_t = \lg \mathrm{E}_t \tag{4-7}$$

其中 $\mathrm{E}_t = \frac{1}{K}\sum_{i=1}^{K} Y_t'(q_i)$

式中 K 表示谐波个数，$Y_t'(q)$ 为归一化对数平均功率。

由于浊音帧中包含了大部分的语音能量，所以其平均功率通常高于清音帧。使用 PEFAC 基频估计算法计算得到的 L_t，在削弱噪声信号能量的同时突显了谐波能量，进而使得语音信号清浊音的能量差异更为明显。

（2）$Z_t(q)$ 中最高的 3 个峰值和与 E_t 的比值：

$$r_t = \frac{\sum_{n=1}^{3} a_{t,n}}{E_t + \varepsilon} \tag{4-8}$$

其中 ε 是一个小的归一化常数，$Z_t(q)$ 的计算式为

$$Z_t(q) = Y_t^{'}(q) * h(-q) \tag{4-9}$$

其中 $h(q)$ 为匹配滤波器，计算式为：

$$h(q) = \begin{cases} 1/[\gamma - \cos(2\pi e^q)] - \beta, & \lg(0.5) < q < \lg(K + 0.5) \\ 0, & \text{其他} \end{cases} \tag{4-10}$$

其中 γ 是用来控制峰值宽度的参数，β 是修正参数，需满足条件 $\int h(q)dq = 0$，K 表示的是谐波峰值的个数。

这个比值主要取决于和谐波相关的分子，选择 $Z_t(q)$ 中最高的 3 个峰值的目的在于有效提高对噪声信号的鲁棒性。

4.3.2 利用时频特征计算决策阈值

将语音信号 $y(n)$ 进行基音-同步信号分割[59]为 $t(l)$ 段，表达式为：

$$t(l) = t(l-1) + \frac{1}{4 f_0(l-1)} \tag{4-11}$$

其中 $f_0(l)$ 是第 l 帧的基频。

将谐波相位 $\psi(h,l)$ 分解为三个部分：

$$\psi(h,l) = \angle V(h,l) + \psi_d(h,l) + h\sum_{l^{'}=0}^{l} \omega_0(l^{'})(t(l^{'}) - t(l^{'}-1)) \tag{4-12}$$

其中，$\angle V(h,l)$ 表示声道滤波器的最小相位响应，可以从实倒谱中估计。$\psi_d(h,l)$ 是分散相位，由激励信号的附加相位信息和基础谐波相位的随机特征组成。$h\sum_{l^{'}=0}^{l} \omega_0(l^{'})(t(l^{'}) - t(l^{'}-1))$ 是线性相位分量，只依赖于第 l 帧的基频 $f_0(l)$。$\omega_0(l^{'})$ 为归一化基本频率。

令 $\Psi(h,l) = \angle V(h,l) + \psi_d(h,l)$，$\psi_{l\ln}(h,l) = h\sum_{l^{'}=0}^{l} \omega_0(l^{'})(t(l^{'}) - t(l^{'}-1))$

即 $\Psi(h,l)$ 和 $\psi_{l\ln}(h,l)$ 分别为展开相位和线性相位，也就是说公式(4-12)把谐波相位分解成展开相位与线性相位。其中展开相位通常被用于基于相位信息的语音增强，表达式如下：

$$\Psi(h,l)=\psi(\mathrm{h,l})-\psi_{l\ln}(h,l) \tag{4-13}$$

由式(4-13)可知，用带噪语音谐波相位减去用 PEFAC 算法估计出的谐波相位的线性相位部分，可计算出展开相位 $\Psi(h,l)$ 。

声道滤波器的频率响应可以假设在一个音素内沿时间固定，因此，最小相位分量显示的是缓慢变化的统计量。在被噪声破坏的语音信号中，噪声的加入会污染声道滤波器的相位信息，而线性相位部分只取决于估计基频的精度。去掉了线性相位部分，通过增强带噪语音的展开相位，以减少噪声污染[30]。因此，本文将展开相位的时频特征及信噪比信息应用于传统基于模型的相位重构算法中增强带噪语音。

定义 $\Psi_x(h,l)$ 和 $\Psi_y(h,l)$ 为纯净和带噪语音信号的展开相位分量。在浊音或无浊音的假设下，二元假设检验能够表述成：

$$H_0:\hat{\Psi}_x(h,l)=\Psi_y(h,l) \tag{4-14}$$

$$H_1:\hat{\Psi}_x(h,l)=\Psi_y(h,l)+e(h,l) \tag{4-15}$$

其中 $e(h,l)$ 为误差项。

假设 H_0 表示在谱相位中无谐波结构的情形，因此，可以假定谱相位均匀地分布在相位变量的范围 $[-\pi,\pi]$ 内，即：

$$p(\psi_y(h,l)|H_0)\sim \mathrm{U}[-\pi,\pi]=\frac{1}{2\pi} \tag{4-16}$$

均匀分布没有关于相位谐波结构(最大不确定性和随机性)的信息。因此，可以假设带噪语音相位 $\psi_y(h,l)$ 是谱相位 $\hat{\psi}_x(h,l)$ 的最佳估计。对未知纯净语音谱相位的 MMSE 估计用带噪语音的相位表示。

定义 H_1 假设为谱相位中存在谐波的情况，假设谱相位服从 von 米塞斯分布 $\psi_y(h,l)\sim \mathrm{VM}(\psi_\mu(h,l),\kappa(h,l))$ ，则：

$$p(\psi_y(h,l)|H_1)=\frac{e^{\kappa(h,l)\cos(\psi_y(h,l)-\psi_\mu(h,l))}}{2\pi I_0(\kappa(h,l))} \tag{4-17}$$

式子(4-17)中$I_v(\bullet)$是第一类顺序v的修正贝塞尔函数，由误差项$e(h,l)$表示相位估计中的不确定性，并通过均值$\psi_\mu(h,l)$和浓度参数$\kappa(h,l)$进行统计建模。给定两个假设H_1和H_2，接受其中任何一个的决定是由$R(H_0:H_1)=\frac{I_0(\kappa(h,l))}{e^{\kappa(h,l)\cos(\psi_y(h,l)-\psi_\mu(h,l))}}$决定的。

应用决策规则作为二元假设：对$R(H_0:H_1)\underset{H_0}{\overset{H_1}{\gtrless}}1$取对数，最终得到：

$$\cos(\psi_y(h,l)-\psi_\mu(h,l))\underset{H_0}{\overset{H_1}{\gtrless}}\theta_{th}(h,l) \tag{4-18}$$

其中定义$\theta_{th}(h,l)=\frac{\ln I_0(\kappa(h,l))}{\kappa(h,l)}$为决策阈值。平均值$\psi_\mu(h,l)$和浓度参数$\kappa(h,l)$所构建的von 米塞斯分布模型表示的是相位的先验分布[70]。均值$\psi_\mu(h,l)=\hat{\psi}_x(h,l)$，使用清音段和浊音段分别估计出的语音相位来表示。浓度参数$\kappa(h,l)=\frac{2|Y(k,l)||S(k,l)|}{\sigma_N^2}$。$\sigma_N^2$表示噪声方差，$|S(k,l)|$是纯净语音幅值。如图 4-5 是冯·米塞斯分布图。在文献[70]中，相位的浓度参数是依赖于 SNR 的。因此，接下来我们提出如何利用信噪比信息计算相位偏差。

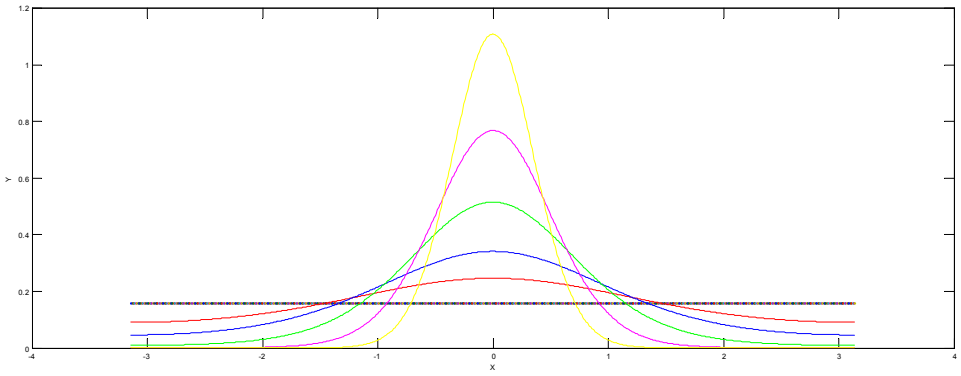


图4-5 冯·米塞斯分布图

Fig.4-5 Map of theVon Mises Distribution

图中横轴表示的是角度自变量，纵轴表示的是概率密度。其中黑色，红色，蓝色，绿色，粉色，黄色线条分别表示的是$\kappa=0$，$\kappa=1/2$，$\kappa=1$，$\kappa=2$，$\kappa=4$，$\kappa=8$。

4.3.3 利用信噪比信息计算相位偏差

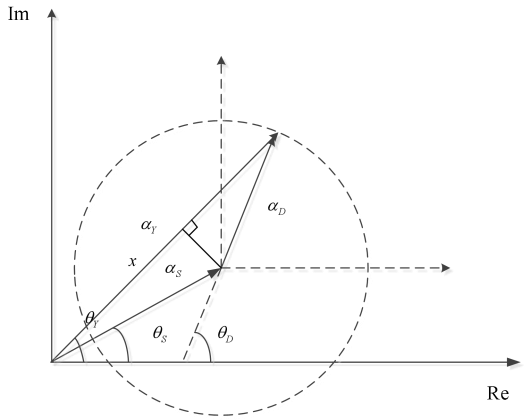


图4-6计算相位偏差原理图

Fig.4-6 Flow chart of the computation of phase deviation

由图 4-6 可知

$$\alpha_S^2 = \alpha_Y^2 + \alpha_D^2 - 2\alpha_Y\alpha_D\cos(\theta_D - \theta_Y) \tag{4-19}$$

根据先验 SNRξ和后验 SNRγ的定义式，有：

$$\cos\left(\theta_D - \theta_Y\right) = \frac{\alpha_Y^2 + \alpha_D^2 - \alpha_S^2}{2\alpha_Y\alpha_D} = \frac{\gamma + 1 - \xi}{2\sqrt{\gamma}} \tag{4-20}$$

定义$\phi_{dev} = \theta_Y - \theta_S$作为带噪语音相位$\theta_Y$与纯净语音相位$\theta_S$间的相位偏离。相位偏差概念最初是由 Vary(1985)定义，他研究了增强带噪语音相位信息对语音增强的影响，得出了相位偏差大于 0.679 弧度的情况下增强相位对提高语音的感知质量效果不明显的结论。

由图 4-6 计算出几何关系

$$\cos(\theta_D - \theta_Y) = \frac{\alpha_Y - x}{\alpha_D} \tag{4-21}$$

$$\cos\phi_{dev} = \frac{x}{\alpha_S} \tag{4-22}$$

因此

$$\cos\phi_{dev} = \frac{\alpha_Y}{\alpha_S} - \frac{\alpha_D}{\alpha_S} * \cos(\theta_D - \theta_Y) = \frac{\gamma + \xi - 1}{2\sqrt{\xi\gamma}} \tag{4-23}$$

4.3.4 估计清音段和浊音段的语音相位

给定先验和后验 SNR 的估计，本文使用(4-23)得到$\cos\phi_{dev}(h,l)$的估计。然后将估计的相位偏差余弦与使用时频信息计算的决策阈值$\theta_{th}(h,l)$进行比较。

$if\ \cos\phi_{dev}(h,l)>\theta_{th}(h,l)$　　浊音段语音相位$\psi(h,l)=princ\left\{\phi^{S}_{k,(l-1)}+\Omega^{k}_{h,l}L\right\}$

清音段语音相位$\psi(h,l)=\phi_{Y}(k,l)$

由上式可知，当相位偏差余弦超过由浓度参数控制的阈值时，由于相位偏差与人的感知无关，因此使用带噪语音相位来估计清音段语音相位，同时使用时域方法估计浊音段语音相位：

$if\ \cos\phi_{dev}(h,l)<\theta_{th}(h,l)$　　浊音段语音相位$\psi(h,l)=princ\left\{\phi^{S}_{k'}-\phi^{W}_{k'-\kappa^{k}_{h}}+\phi^{W}_{k'-\upsilon^{k}_{h}+i}\right\}$

清音段语音相位$\psi(h,l)=\phi_{Y}(k,l)-\arccos\phi_{dev}$

当相位偏差余弦低于由浓度参数控制的阈值时，使用频域方法估计浊音段语音相位，同时使用相位的几何关系式来估计清音段语音相位。将时频特征和信噪比信息应用在相位重构中估计语音相位，既增强了感知语音质量又提升了语音连贯性。

4.3.5 改进二元假设模型的 MMSE 对数谱幅度估计(MMSE-LSA)

根据二元假设模型用 MMSE-LSA 语音增强算法得到纯净语音信号S_k的估计值为：

$$S_k=\exp\left\{[E(\ln S_k)\left|Y_k,H_1^k\right]P(H_1^k\left|Y_k\right)+[E(\ln S_k)\left|Y_k,H_0^k\right]P(H_0^k\left|Y_k\right)\right\} \tag{4-24}$$

其中H_0^k和H_1^k分别假设的是在频点k处语音信号缺失和存在两种条件。$P(H_1^k\left|Y_k\right)$和$P(H_0^k\left|Y_k\right)$分别代表的是语音存在条件下的后验概率和语音缺失条件下的后验概率，在条件先验信噪比ξ_k'取值较大的情况下，语音存在条件下的后验概率$P(H_1^k\left|Y_k\right)\approx 1$，当$\xi_k'$取值较小时，$P(H_1^k\left|Y_k\right)\approx 1-q_k$，其中$q_k=0.5$即近似等于语音存在条件下的先验概率$P(H_1^k)$。

该 MMSE-LSA 算法虽然能够显著提高语音质量，但因为采用恒定的加权因子G_{DD}，使谱估计不够准确，导致产生音乐噪声。为了解决这一问题，本文提出一种用改进的 TSNR(Two Step Noise Reduction, TSNR)的增益联合当前帧的先验信息来代替原算法中恒定的加权因子G_{DD}，以提高先验信噪比估计的精确度，从而高度消除音乐噪声。其中

TSNR 算法[71]采用基于决策导向(DD)方法的结果联合系统的增益因子从而修正对当前帧语音信号的先验信噪比估计，公式如下：

$$\xi_{TSNR}(n,k)_{new}=\frac{G_{TSNR}(n,k)\left|Y(n,k)G_{TSNR}\right|^2+(1-G_{TSNR})\left|\hat{S}(n-1,k)\right|^2}{\lambda_d(k,n)} \tag{4-25}$$

实验中 $G_{TSNR}(n,k)=\dfrac{\xi_{TSNR}}{1+\xi_{TSNR}}$， $\lambda_d(k,n)$ 是对噪声的功率谱估计。

$$\xi_{TSNR}(n,k)=\frac{\left|G_{DD}Y(n,k)\right|^2}{\lambda_d(k,n)} \tag{4-26}$$

其中 $G_{DD}=\xi_{DD}(n,k)/(1+\xi_{DD}(n,k))$

在公式(4-24)中 $E[(\ln S_k)\left|Y_k,H_0^k\right]=0$，由此计算出 MMSE 对数幅度谱的估计式为：

$$\hat{S}_k=(\exp[E(\ln S_k)\left|Y_k,H_1^k\right])^{P(H_1^k|Y_k)}=[G_{LSA}(\xi_k,\nu_k)Y_k]^{P(H_1^k|Y_k)} \tag{4-27}$$

4.3.6 结合幅度与相位估计

将上述幅度与相位的改进方法结合得到最终增强后的信号频域表示式如下：

$$S(n,k)=\left|\hat{S}(n,k)\right|e^{jn\omega_0(l)}e^{j\angle\psi(h,l)} \tag{4-28}$$

其中 $\hat{S}(n,k)$ 为结合语音存在概率的对数 MMSE 估计，$\psi(h,l)$ 为重构后的语音谐波相位。将式子(4-28)进行逆短时傅里叶变换得到增强后的语音信号。

4.4 实验仿真

要实现本章所提出的方法需要预先进行基频估计，为了从带噪语音中提取出基频 f_0，采用稳健的基频估计算法 PEFAC。为了验证该算法的实验效果，选自 863 语音库中的纯净语音 SP01，在带噪语音信号信噪比是 5dB 的条件下进行测试，得到基频估计图如图 4-7 所示。为了估计先验和后验 SNR，采用改进两步噪声消除算法(TSNR)连同 MMSE 无偏噪声估计器。分别以 SP01 和 SP15 纯净语音为例，采样率取 8000Hz，选取汉宁窗为其窗函数，帧长设置为 256 ，帧移设置为 64。在信噪比为 5dB 的条件下进行测试，将 F16 噪声作为背景噪声，原算法与本文改进算法实验结果对比的波形图如图 4-8 所示，将白噪声作为背景噪声，语谱图如图 4-9 所示。将白噪声和 F16 噪声分别选作背景噪声，将信噪比设置成 5dB 所得相位信息的对比结果图如图 4-10 所示。

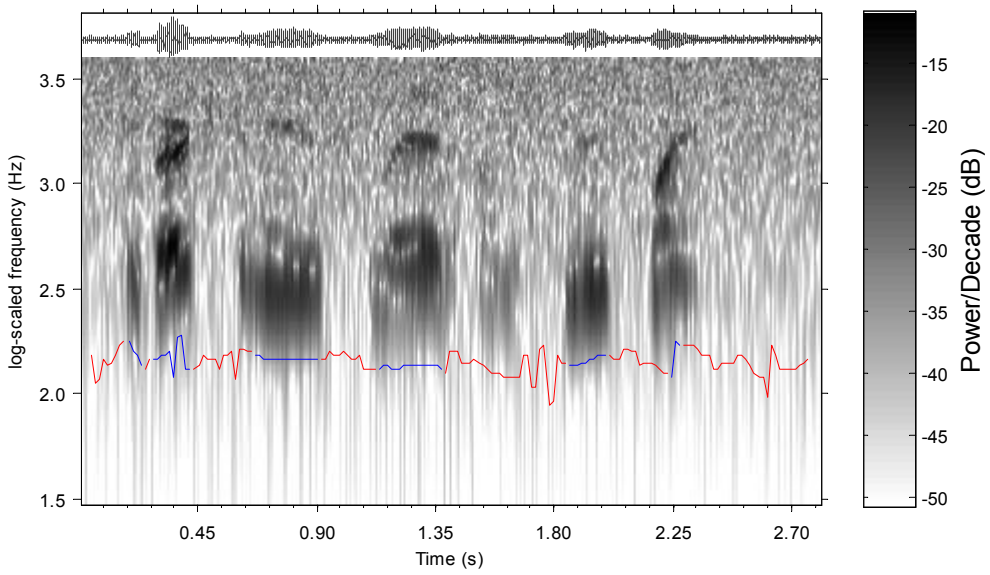
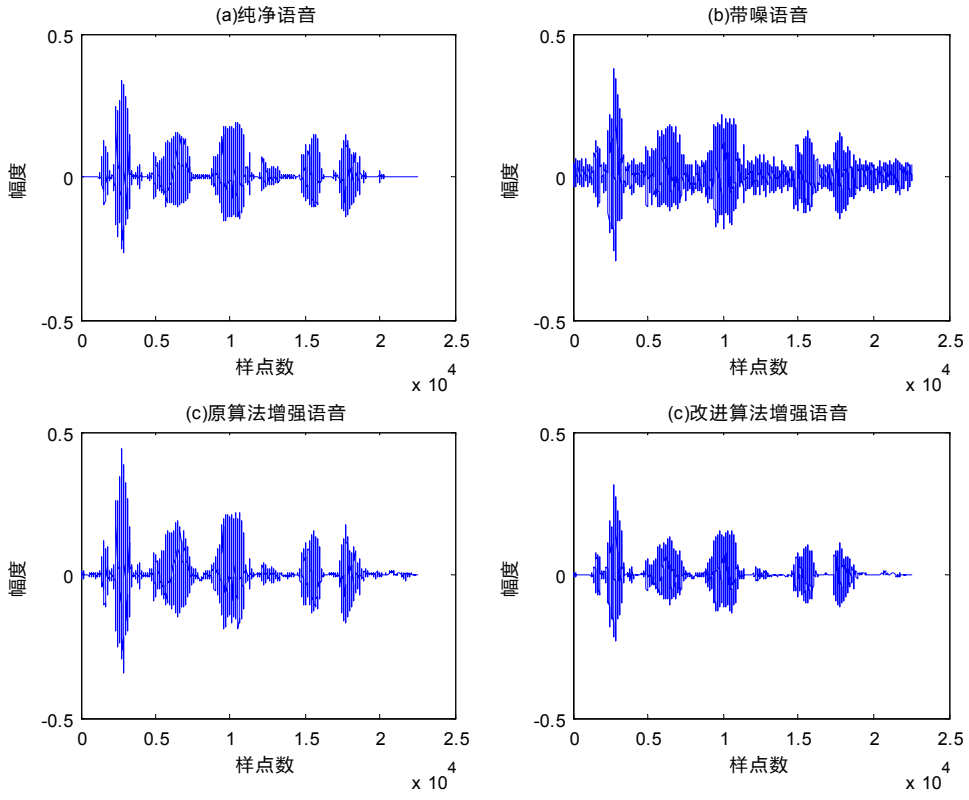
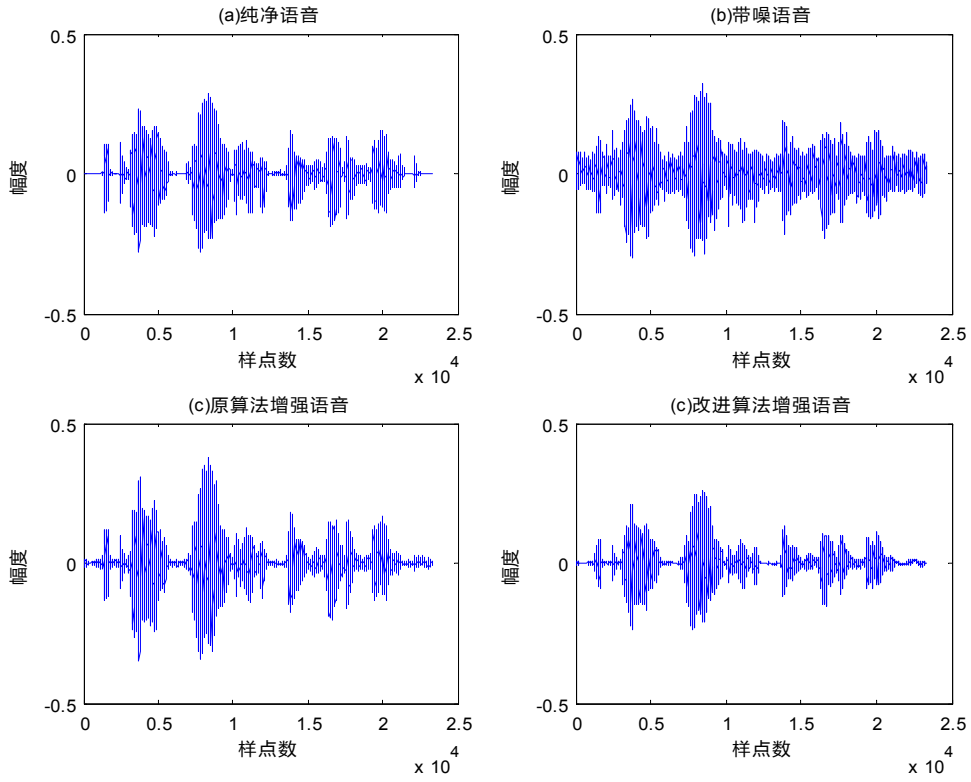


图4-7基频估计图

Fig.4-7 The graph of fundamental frequency estimation



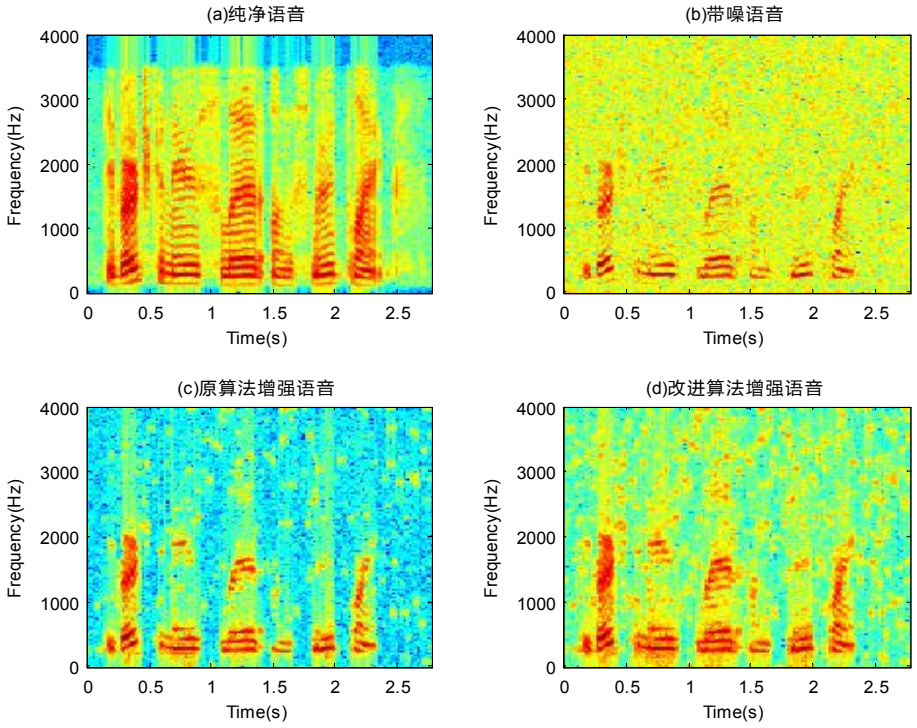
(a)SP01 语音



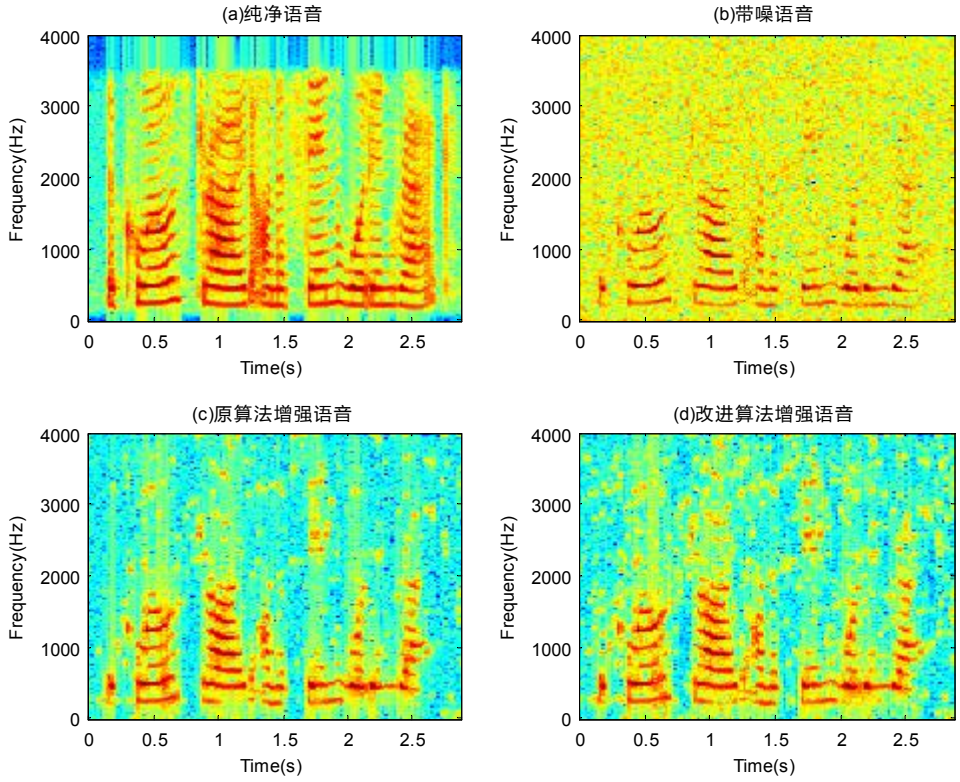
(b)SP15 语音

图 4-8 两种相位重构算法的去噪结果波形对比图

Fig.4-8 Waveform comparison of denoising results of two phase reconstruction algorithms



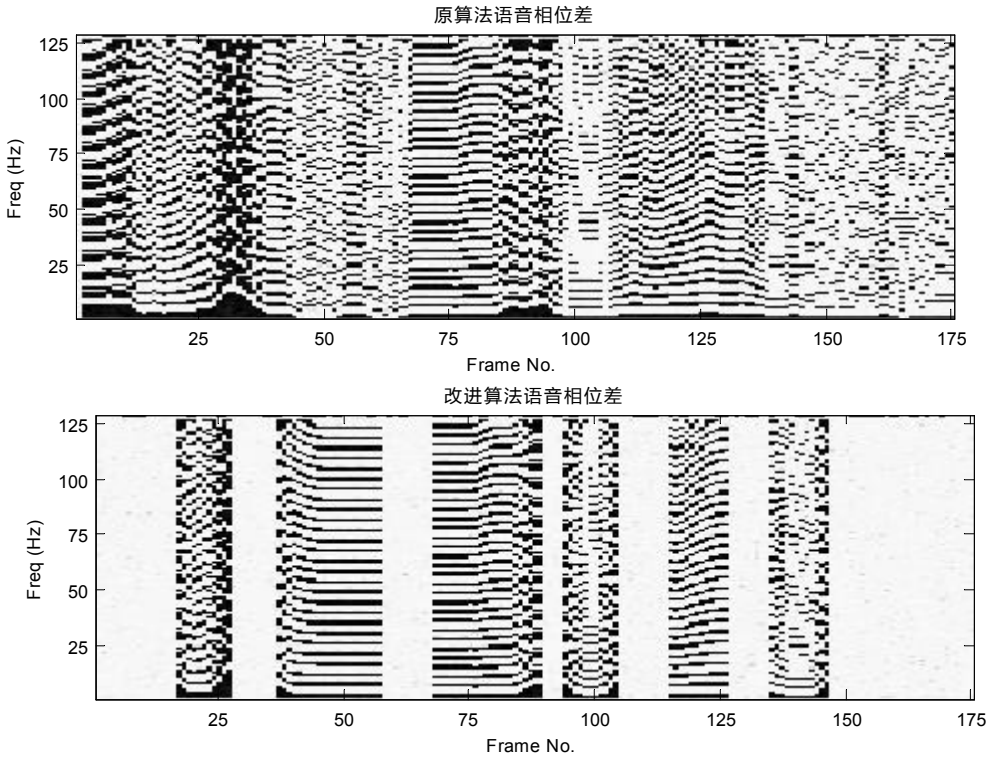
(a)SP01 语音



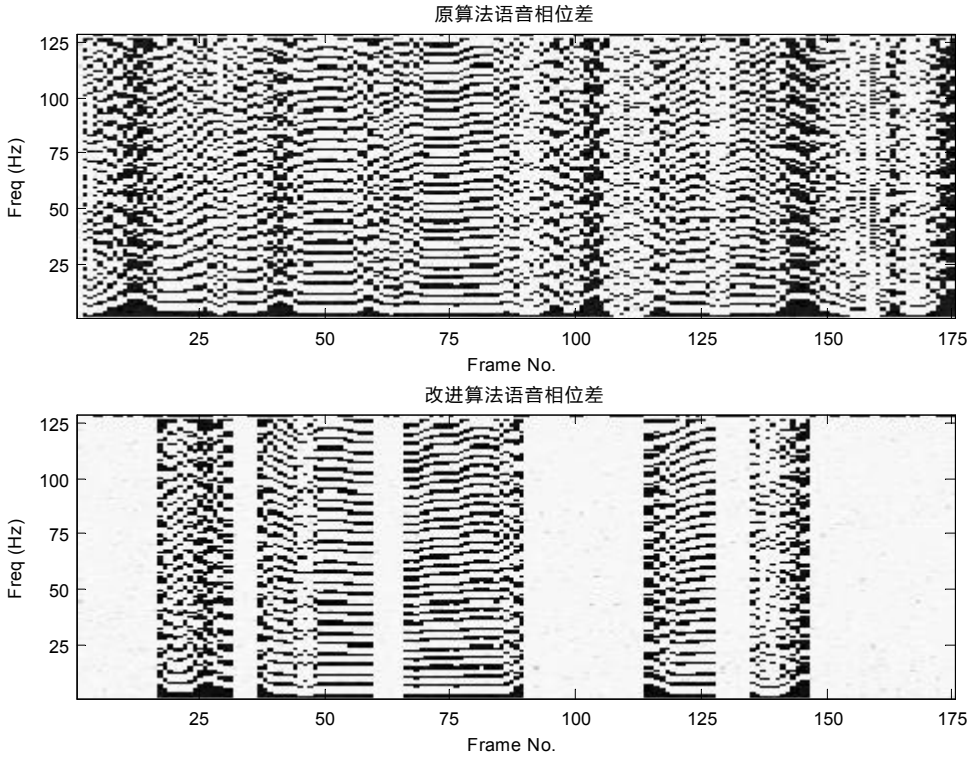
(b)SP15 语音

图4-9两种相位重构算法的去噪结果语谱对比图

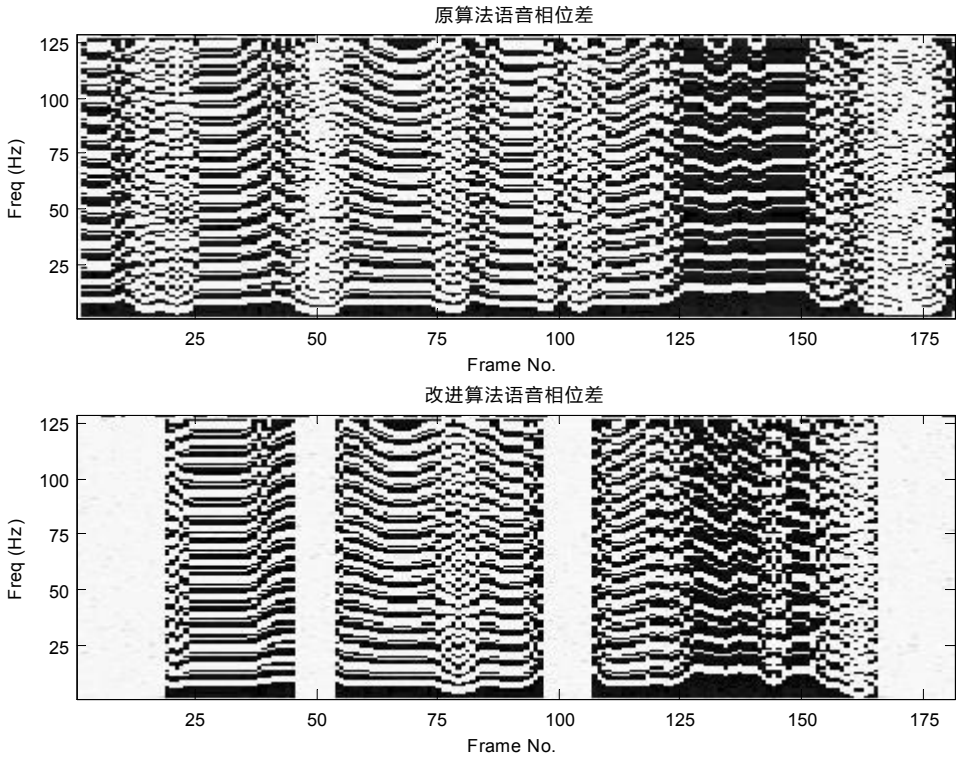
Fig.4-9 Spectrum comparison of denoising results of two phase reconstruction algorithms



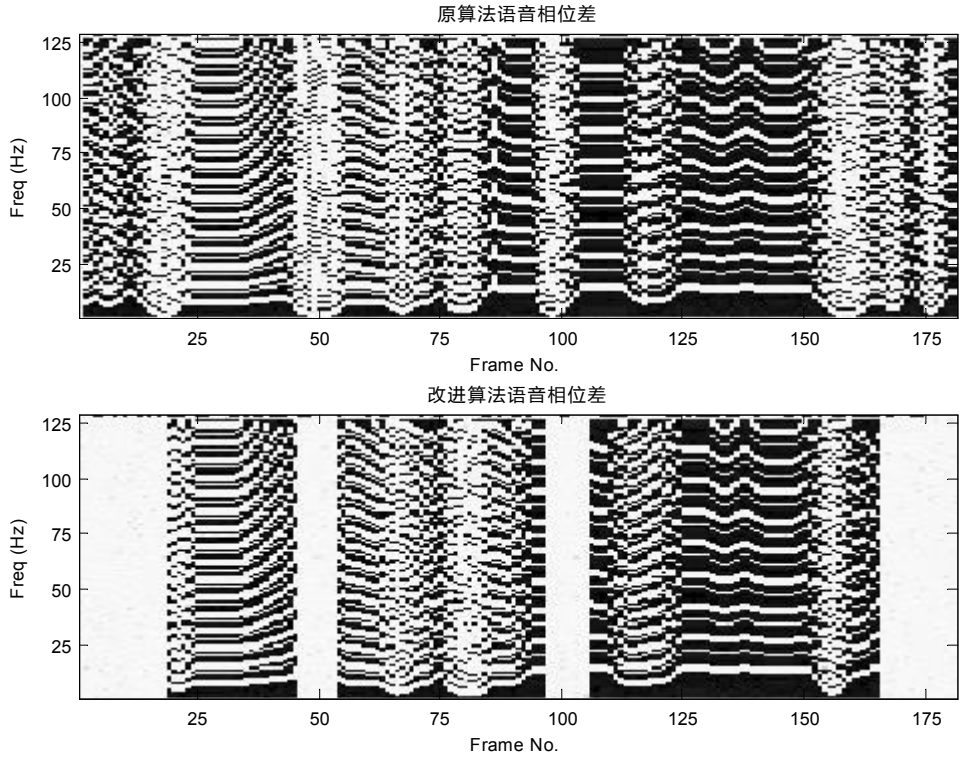
(a) SP01白噪声相位效果图



(b) SP01F16噪声相位效果图



(c) SP15白噪声相位效果图



(d)SP15F16噪声相位效果图

图4-10两种相位重构算法的相位对比图

Fig.4-10 Phase comparison of denoising results of two phase reconstruction algorithms

从上述两种相位重构算法的去噪结果波形对比图能够看出，原算法仅仅是在语音信号的浊音段有效抑制了噪声，减少了信号失真，而改进的相位重构算法在语音信号的清音段也实现了去噪效果。

从上述两种相位重构算法的语谱对比图能够看出，纯净语音信号所对应的谱线非常规则且清楚，但是在引入背景噪声之后，纯净语音本来的谐波结构因为噪声的干扰而变得不清晰。传统基于谐波模型的相位重构方法能够在某种程度上增强语音信号，但是与最初纯净语音信号的频谱结构相较而言，却缺失了不少重要信息，在使用本章所改进的相位重构算法之后语音部分的谱线更为完整清晰，增强后的相位谱恢复了带噪语音信号中丢失的纯净相位的谐波结构，对语音信号的增强效果也更加明显。

从上述两种相位重构算法的相位对比图能够看出，原算法的语音相位差结构较为模糊，而改进算法的语音相位差结构更为清晰明显，能够清楚地显现出信号的谐波结构，进一步验证了本章所提出的改进基于谐波模型的相位重构算法在改善语音谐波结构方面的有效性。

表 4-1 两种语音增强算法的信噪比结果(SP01)

Tab4-1 Signal-to-noise ratio results of two speech enhancement algorithms(SP01)

噪声类型	输入信噪比(dB)	原算法(dB)	改进算法(dB)
White 噪声	0	7.7692	8.2813
	5	10.9203	11.7749
	10	12.9470	15.3428
	15	14.8041	18.4642
Pink 噪声	0	7.7294	8.0674
	5	10.5005	11.6328
	10	13.3217	14.7645
	15	15.0630	18.1276
F16 噪声	0	7.5699	7.6268
	5	10.6411	11.4105
	10	13.0429	14.5357
	15	15.0049	18.2353

表 4-2 两种语音增强算法的 PESQ 结果(SP01)

Tab4-2 PESQ results of two speech enhancement algorithms(SP01)

噪声类型	输入信噪比(dB)	纯净 SP01 语音	含噪语音	原算法	改进算法
White 噪声	0	4.5000	1.4564	1.9079	1.9416
	5	4.5000	1.7481	2.2883	2.3019
	10	4.5000	2.0649	2.6325	2.6758
	15	4.5000	2.4078	2.7530	2.8714
Pink 噪声	0	4.5000	1.6199	2.0038	2.0873
	5	4.5000	1.9546	2.3198	2.4240
	10	4.5000	2.2874	2.6320	2.7088
	15	4.5000	2.5874	2.7664	2.9639
F16 噪声	0	4.5000	1.8803	2.2299	2.2433
	5	4.5000	2.1562	2.5168	2.5855
	10	4.5000	2.4621	2.6921	2.8431
	15	4.5000	2.6569	2.8486	3.0166

表 4-3 两种语音增强算法的信噪比结果(SP15)

Tab4-3 Signal-to-noise ratio results of two speech enhancement algorithms(SP15)

噪声类型	输入信噪比(dB)	原算法(dB)	改进算法(dB)
White 噪声	0	4.7413	6.6161
	5	5.2112	8.0194
	10	10.4186	14.2552
	15	15.3014	18.2615
Pink 噪声	0	4.1475	6.3913
	5	5.9592	8.7592
	10	10.0196	13.7947
	15	15.8768	19.3986
F16 噪声	0	2.2990	6.0169
	5	5.4097	8.2092
	10	10.6806	14.4447
	15	15.5405	19.9922

表 4-4 两种语音增强算法的 PESQ 结果(SP15)

Tab4-4 PESQ results of two speech enhancement algorithms(SP15)

噪声类型	输入信噪比(dB)	纯净 SP15 语音	含噪语音	原算法	改进算法
White 噪声	0	4.5000	1.3067	1.7587	1.7923
	5	4.5000	1.6190	2.0897	2.1314
	10	4.5000	1.9805	2.2853	2.4064
	15	4.5000	2.3711	2.6541	2.9032
Pink 噪声	0	4.5000	1.4019	1.8053	1.8621
	5	4.5000	1.7537	2.1393	2.2804
	10	4.5000	2.1331	2.3673	2.5182
	15	4.5000	2.5629	2.5020	2.7418
F16 噪声	0	4.5000	1.4299	1.7101	1.9344
	5	4.5000	1.7943	2.2528	2.3589
	10	4.5000	2.1843	2.3984	2.5667
	15	4.5000	2.6002	2.5174	2.7250

将基于谐波模型相位谱重构算法结合传统 MMSE-LSA 估计与本文改进算法进行对

比实验，使用信噪比和感知语音质量评估测度来对增强语音进行评价，结果如表 4-1、表 4-2、表 4-3 和表 4-4 所示。

结果表明，通过本文提出的方法，增强后语音的感知质量和可懂度在各种 SNR 不同的背景噪声下明显提高。其中 SP01 语音的信噪比平均提高了 1.57dB，PESQ 平均提高了 0.25，SP15 语音的信噪比平均提高了 3.21dB，PESQ 平均提高了 0.11。

4.5 本章总结

本文提出了一种利用从带噪语音中估计的基频和信噪比信息改进基于谐波模型的相位重构算法，该方法使用时频特征结合 SNR 信息重构清浊音段语音信号相位，并结合改进的相位估计方法和 MMSE-LSA 幅值估计方法。实验结果表明，相比传统的基于谐波模型的相位重构算法而言，本方法在语音质量和可懂度方面具有更明显的提高。

第五章 总结与展望

5.1 工作总结

当前绝大多数的单通道语音增强算法只是对语音信号的幅值信息做处理，而保持相位信息不变，这样会导致语音质量和可懂度的降低。针对该问题，本文提出了一种改进相位谱信息和用信噪比信息与时频特征改进基于谐波模型的相位重构语音增强算法。

第三章中提出的改进相位谱信息算法的具体流程如下：首先，针对传统相位谱补偿算法中补偿因子是一个固定值从而导致语音增强效果受限的问题，提出一种新的基于每帧语音信号输入信噪比的*Sigmoid*型相位谱补偿函数，能够根据噪声的变化来灵活地对含噪语音信号的相位谱做出补偿。并且结合新的改进语音存在概率方法来对噪声信号的功率谱密度进行估计，将其和传统的MMSE-LSA幅值估计方法相结合用于语音增强。

第四章中为了进一步深入研究基于信号相位谱信息的语音增强方法，在本章提出了结合信噪比信息与时频特征改进基于谐波模型的相位重构语音增强算法。具体处理过程如下：首先，该方法引入与相位失真有关的时频特征并计算决策阈值；然后，利用信噪比信息计算带噪语音与纯净语音的相位偏差，将其与决策阈值相比较并采用时域和频域相结合的方法进一步估计出清音段和浊音段的语音相位，能够有效增强语音可懂度；最后，将重构的相位信息与改进二元假设模型的幅值估计结合以进行语音增强。

在不同噪声背景下，通过对不同带噪语音信号进行实验仿真，本文提出的改进相位谱信息和用信噪比信息与时频特征改进基于谐波模型的相位重构语音增强算法可以有效抑制音频信号中的各类噪声，同时增强语音信号感知质量，并且提升其可懂度。

5.2 工作展望

在本文提出的算法上，仍然有许多地方值得去进一步深入研究：

(1)在本文提出的改进相位谱信息和用信噪比信息与时频特征改进基于谐波模型的相位重构语音增强算法中，对概率密度函数的分布模型推导分别采用复高斯模型和冯·米塞斯分布模型，实际上一些复杂的超高斯模型如拉普拉斯分布模型也可以有效地表示语音与噪声的分布特性，然而这些模型在实验仿真中往往较难构建，在今后的工作中，对于使用这些统计分布模型拟合语音与噪声的分布是一个非常重要的研究方向。

(2)本文提出的用信噪比信息与时频特征改进基于谐波模型的相位重构语音增强算法中，首先使用 PEFAC 算法进行清浊音段的划分，因此 PEFAC 算法的精确性对后续算法的实施有着重要的影响，在以后的实验中仍然要继续研究在不同的噪声条件下，性能更显著的基频估计算法。

(3)本文提出的两种改进相位信息的语音增强方法都需要进行噪声功率谱估计，在第三章提出的噪声估计方法中，估计噪声功率谱用一个后验子带信噪比计算，基于该子带的后验信噪比的一个 *Sigmoid* 型权值用于增强听觉域中的语音频谱。同时在改进 MMSE-LSA 的幅值估计中对两步噪声消除算法的加权因子做出新的改进，然而在整个 MMSE-LSA 幅值估计中，本文提出的改进形式产生的效果存在很大局限，如何能够对先验信噪比进行更优估计以改进整体的语音增强效果，将是以后研究的重点。

参考文献

[1] 李伟林, 文剑, 马文凯. 基于深度神经网络的语音识别系统研究[J]. 计算机科学, 2016, 43(2):45-49.

[2] 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(5).

[3] M. Berouti, R. Schwartz, J. Makhoul, Enhancement of speech corrupted by acoustic noise, in IEEE International Conference on Acoustics, Speech, and Signal Processing, [J].ICASSP, IEEE ,1979,79(4):208–211.

[4] Lockwood P, Boudy J. Experiments with a nonlinear spectral subtractor (NSS), Hidden Markov models and the projection, for robust speech recognition in cars[J]. Speech Communication, 1992, 11(2-3):215-228.

[5] Saldanha J C , Shruthi O R . Reduction of noise for speech signal enhancement using Spectral Subtraction method[C]// International Conference on Information Science. IEEE, 2017.

[6] Seok J W , Bae K S . Reduction of musical noise in spectral subtraction method using subframe phase randomisation[J]. Electronics Letters, 1999, 35(2):123.

[7] Hansen P C , Jensen S H . FIR filter representations of reduced-rank noise reduction[J]. IEEE Transactions on Signal Processing, 1998, 46(6):1737-1741.

[8] Doclo S , Moonen M . On the output SNR of the speech-distortion weighted multichannel Wiener filter[J]. IEEE Signal Processing Letters, 2005, 12(12):809-811.

[9] Modhave N , Karuna Y , Tonde S . Design of matrix wiener filter for noise reduction and speech enhancement in hearing aids[C]// IEEE International Conference on Recent Trends in Electronics. IEEE, 2017.

[10] Yelwande A , Kansal S , Dixit A . Adaptive wiener filter for speech enhancement[C]// International Conference on Information. IEEE, 2018.

[11] Wei J , Ou S , Shen S , et al. Laplacian-Gaussian mixture based dual-gain wiener filter for speech enhancement[C]// IEEE International Conference on Signal & Image Processing. IEEE, 2017.

[12] Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator[J]. IEEE Transactions on Acoustics Speech & Signal Processing, 2003, 32(6):1109-1121.

[13] Borgstrom B J, Alwan A. A Unified Framework for Designing Optimal STSA Estimators Assuming Maximum Likelihood Phase Equivalence of Speech and Noise[J]. Audio Speech & Language Processing IEEE Transactions on, 2011, 19(8):2579-2590.

[14] Berkun R, Cohen I. Microphone array power ratio for quality assessment of reverberated speech[J]. EURASIP Journal on Advances in Signal Processing, 2015, 2015(1):1-11.

[15] Talbi M. Electrocardiogram de-noising based on forward wavelet transform translation invariant application in bionic wavelet domain[J]. Sadhana, 2014, 39(4): 921-937.

[16] Yao J , Zhang Y T . Bionic wavelet transform: a new time-frequency method based on an auditory model[J]. IEEE transactions on bio-medical engineering, 2001, 48(8):856-863.

[17] Johnson M T , Yuan X , Ren Y . Speech signal enhancement through adaptive wavelet thresholding[J]. Speech Communication, 2007, 49(2):123-133.

[18] Jia H R, Zhang X Y, Bai J. A continuous differentiable wavelet threshold function for speech enhancement[J].Journal of Central South University, 2013,20(8):2219-2225.

[19] Mourad T . Speech enhancement based on stationary bionic wavelet transform and maximum a posterior estimator of magnitude-squared spectrum[J]. International Journal of Speech Technology, 2017, 20(1):1-14.

[20] Ephraim Y, Trees H L V. A signal subspace approach for speech enhancement[J]. IEEE Transactions on Speech & Audio Processing, 1993, 3(4):251-266.

[21] Jabloun F , Champagne B . Incorporating the human hearing properties in the signal subspace approach for speech enhancement[J]. IEEE Transactions on Speech and Audio Processing, 2004, 11(6):700-708.

[22] You C H , Koh S N , Rahardja S . An invertible frequency eigendomain transformation for masking-based subspace speech enhancement[J]. IEEE Signal Processing Letters, 2005, 12(6):461-464.

[23] Surendran S , Kumar T K . Perceptual Subspace Speech Enhancement with Variance Normalization[C]// International Conference on Microelectronics. IEEE, 2016.

[24] Wang D L , Lim J S . The unimportance of phase in speech enhancement[J]. IEEE Transactions on Acoustics Speech and Signal Processing, 1982, 30(4):679-681.

[25] Vary P , Eurasip M . Noise suppression by spectral magnitude estimation —mechanism and theoretical limits[J]. Signal Processing, 1985, 8(4):387-400.

[26] Lotter T, Vary P. Speech enhancement by map spectral amplitude estimation using a super-Gaussian speech model[J]. Eurasip Journal on Advances in Signal Processing, 2005, 2005(7):1-17.

[27] Erkelens J S , Hendriks R C , Heusdens R . On the Estimation of Complex Speech DFT

Coefficients Without Assuming Independent Real and Imaginary Parts[J]. IEEE Signal Processing Letters, 2008, 15:213-216.

[28] Alsteris L D , Paliwal K K . Iterative reconstruction of speech from short-time Fourier transform phase and magnitude spectra[J]. Computer Speech and Language, 2007, 21(1):174-186.

[29] Paliwal K , Kamil Wójcicki, Shannon B . The importance of phase in speech enhancement[J]. Speech Communication, 2011, 53(4):465-494.

[30] Mowlaee P , Kulmer J . Harmonic Phase Estimation in Single-Channel Speech Enhancement Using Phase Decomposition and SNR Information[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2015, 23(9):1521-1532.

[31] Mowlaee P , Kulmer J . Phase Estimation in Single-Channel Speech Enhancement: Limits-Potential[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2015, 23(8):1283-1294.

[32] Gerkmann T, Krawczyk M, Rehr R. Phase estimation in speech enhancement — Unimportant, important, or impossible?[C]// Electrical & Electronics Engineers in Israel. IEEE, 2012:1-5.

[33] Krawczyk M, Gerkmann T. STFT phase reconstruction in voiced speech for an improved single-channel speech enhancement[J]. IEEE/ACM Transactions on AudioSpeechand Language Processing,2014, 22(12): 1931-1940.

[34] Dang X , Khan M I A , Nakai T . Noise reduction of speech signal based on phase spectrum estimation[C]// International Conference on Informatics. IEEE, 2013.

[35] Barysenka S Y , Vorobiov V I , Mowlaee P . Single-Channel Speech Enhancement Using Inter-Component Phase Relations[J]. Speech Communication,2018:S0167639317303540.

[36] Jia H, Wang W, Dong W, et al. Speech Enhancement using Modified MMSE-LSA and Phase Reconstruction in Voiced and Unvoiced Speech[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2018:S0218001419580023-.

[37] Kulmer J, Mowlaee P. Phase Estimation in Single Channel Speech Enhancement Using Phase Decomposition[J]. IEEE Signal Processing Letters, 2015, 22(5):598-602.

[38] Mowlaee P, Saeidi R. Time-frequency constraints for phase estimation in single-channel speech enhancement[C]// International Workshop on Acoustic Signal Enhancement. IEEE, 2014:337-341.

[39] Wakabayashi Y, Fukumori T, Nakayama M, et al. Single-Channel Speech Enhancement With Phase Reconstruction Based on Phase Distortion Averaging[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2018, 26(9):1559-1569.

[40] Mowlaee P , Saeidi R . Iterative Closed-Loop Phase-Aware Single-Channel Speech Enhancement[J]. IEEE Signal Processing Letters, 2013, 20(12):1235-1239.

[41] Gunawan D, Sen D. Iterative Phase Estimation for the Synthesis of Separated Sources From Single-Channel Mixtures[J]. IEEE Signal Processing Letters, 2010, 17(5):421-424.

[42] Gerkmann T, Krawczyk M. MMSE-Optimal Spectral Amplitude Estimation Given the STFT-Phase[J]. IEEE Signal Processing Letters, 2013, 20(2):129-132.

[43] Islam M T , Asaduzzaman, Shahnaz C , et al. Speech Enhancement in Adverse Environments Based on Non-stationary Noise-driven Spectral Subtraction and SNR-dependent Phase Compensation[J]. 2018.

[44] Maly A, Mowlaee P. On the importance of harmonic phase modification for improved speech signal reconstruction[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2016:584-588.

[45] 阎福智. 语音信号处理中特征提取方法研究[J]. 中国新通信, 2013(21):127-128.

[46] Mclaughlin P D, Liang T, Homiedan M, et al. High pitch, low voltage dual source CT pulmonary angiography: assessment of image quality and diagnostic acceptability with hybrid iterative reconstruction.[J]. Emergency Radiology, 2015, 22(2):117-123.

[47] 孙锦华, 韩会梅. 低信噪比下时频联合的载波同步算法[J]. 西安交通大学学报, 2015, 49(2):62-68.

[48] 徐宇卓, 马建芬, 张雪英. 基于语音起始段检测语音可懂度客观评价方法[J]. 电子技术应用, 2015, 41(6):150-153.

[49] Pennock S. Accuracy of the perceptual evaluation of speech quality (PESQ) algorithm[J]. Measurement of Speech & Audio Quality in Networks Line Workshop Mesaqin ', 2002.

[50] 邬鑫锋, 曾以成, 刘伯权. 新型几何谱减语音增强方法[J]. 计算机工程与应用, 2010, 46(23):144-147.

[51] Ma Y, Nishihara A. Erratum to: Efficient voice activity detection algorithm using long-term spectral flatness measure[J]. EURASIP Journal on Audio, Speech, and Music Processing, 2015, 2013(1):87.

[52] 熊晶. 语音增强中噪声估计的研究[D]. 兰州交通大学, 2015.

[53] 曹龙涛, 李如玮, 鲍长春,等. 基于噪声估计的二值掩蔽语音增强算法[J]. 计算机工程与应用, 2015, 51(17):222-227.

[54] 严思伟, 屈晓旭, 娄景艺. 基于连续噪声谱估计的谱减法语音增强算法[J]. 通信技术, 2018, 318(6):64-69.

[55]Islam M T , Shahnaz C , Zhu W P , et al. A Divide and Conquer Strategy for Musical Noise-free Speech Enhancement in Adverse Environments[J]. 2018.

[56]刘颖. 不完全伽玛函数计算算法改进与应用[D]. 电子科技大学, 2015.

[57]彭志刚, 潘文君, 熊松林. 与超几何函数相关的几类解析函数族的性质[J]. 数学物理学报：A辑, 2018, 38(2): 215-221.

[58]Zhu C , Yu L , Yan Z , et al. Frequency Estimation of the Plenoptic Function Using the Autocorrelation Theorem[J]. IEEE Transactions on Computational Imaging, 2017:1-1.

[59]Gonzalez S , Brookes M . PEFAC - A Pitch Estimation Algorithm Robust to High Levels of Noise[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 22(2):518-530.

[60]Wang D , Hansen J H L . F0 estimation for noisy speech by exploring temporal harmonic structures in local time frequency spectrum segment[C]// The 41st IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2016.

[61]袁文浩, 梁春燕, 夏斌,等. 一种融合相位估计的深度卷积神经网络语音增强方法[J]. 电子学报, 2018, 46(10):2359-2366.

[62] Mowlaee P , Stahl J , Kulmer J . Iterative joint MAP single-channel speech enhancement given non-uniform phase prior[J]. Speech Communication, 2017, 86:85-96.

[63]Li Z , Wu W , Zhang Q , et al. Speech enhancement using magnitude and phase spectrum compensation[C]// IEEE/ACIS International Conference on Computer & Information Science. IEEE, 2016.

[64]王栋, 贾海蓉. 改进相位谱补偿的语音增强算法[J]. 西安电子科技大学学报(自然科学版), 2017,44(3):83-88.

[65]王虎, 李晶, 赵恒淼, et al. 稀疏低秩模型及相位谱补偿的语音增强算法[J]. 计算机工程与应用, 2018.

[66]Chinaev A, Haeb-Umbach R. A generalized log-spectral amplitude estimator for single-channel speech enhancement[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2017:4980-4984.

[67]容强, 肖汉. 基于MMSE维纳滤波语音增强方法研究与Matlab实现[J]. 计算机应用与软件, 2015, 32(1):153-156.

[68]Lee J, Skoglund J, Shabestary T, et al. Phase-Sensitive Joint Learning Algorithms for Deep Learning-Based Speech Enhancement[J]. IEEE Signal Processing Letters, 2018, 25(8):1276-1280.

[69] Mowlaee P , Saeidi R , Stylianou Y . Advances in phase-aware signal processing in speech communication[J]. Speech Communication, 2016:S0167639316300784.

[70] Gerkmann T. Bayesian Estimation of Clean Speech Spectral Coefficients Given a Priori Knowledge of the Phase[J]. IEEE Transactions on Signal Processing, 2014, 62(16):4199-4208.

[71] Saldanha J C. Speech Enhancement Using Filtering Techniques[C]// National Conference on Emerging Trends in Electronics and Communication. 2016.

致谢

时光荏苒，三年的研究生生涯即将结束，在这里我由衷的感谢三年里在学习和生活上给予我帮助的所有人。

首先感谢我尊敬的导师贾海蓉老师三年以来给我创造的有利学习条件以及在课题研究上的指导和信任，让我对所研究的方向产生了浓厚的兴趣。在学习中，老师不仅给予我专业上的指导，而且给了我更多学习的自由空间；在生活中，老师对我关怀备至，在我迷茫时耐心地开导我鼓励我。在老师的认真指导下，我才能顺利完成研究生阶段的学习和科研任务。

同时，感谢数字音视频技术研究中心的老师们。两年多的时间里，老师们不仅在专业知识上给予我帮助，也在论文撰写的过程中给我提了好多的宝贵意见。

另外我还要感谢数字音视频技术研究中心的同学们和我的舍友。在学习和生活上他们同样给了我极大的帮助，感谢他们给予我一段丰富多彩的研究生生涯。在我学习遇到困难的时候，在我心情低落的时候他们给了我安慰与鼓励，使我能够踏踏实实的学习生活。

最后特别感谢一直给予我支持和帮助的父母，感谢他们给我最无私的关怀与爱，他们的关爱是我前进的动力与源泉，让我能顺利的完成我的研究生学业，愿他们在今后的生活中还是一样的健健康康、开开心心。

最后，对在百忙之中能抽出时间审阅本论文和参与答辩的诸位老师致以真诚的感谢。

攻读学位期间发表的学术论文目录

[1] 吉慧芳, 贾海蓉，王雁.改进相位谱补偿的语音增强方法[J]. 计算机工程与应用,2018.（已录用）

[2] 王雁，贾海蓉，吉慧芳，王卫梅. 特征联合优化深度信念网络的语音增强算法[J].计算机工程与应用,2018.（已录用）