

基于多窗谱估计和几何谱减的低信噪比语音增强方法

李 湑^{1,2,3}, 胡 俊^{1,2}, 刘 新^{2,3,4}, 黄石磊²

(1 重庆邮电大学 光通信与网络重点实验室, 重庆 400065;

2 深港产学研基地 深圳市智能媒体和语音重点实验室 广东 深圳 518057;

3 北京大学 深圳研究院, 广东 深圳 518057; 4 安徽大学 计算智能与信号处理教育部重点实验室, 安徽 合肥 230601)

摘 要: 在低信噪比条件下, 传统谱减算法对语音和噪声的独立性假设会产生一定的“音乐噪声”, 从而降低了语音的可懂度。对此, 本文提出了一种结合多窗谱估计和几何谱减的语音增强算法。本算法利用多窗谱估计带噪语音信号的功率谱, 采用改进最小值控制递归平均算法实时跟踪估计噪声谱, 应用几何谱减算法求解增益函数, 从而恢复出精确的语音信号。通过在 IEEE 数据集上进行实验, 并以 PESQ 和 LSD 作为评价指标, 结果表明本算法在低信噪比条件下, 缩短了带噪语音与纯净语音之间的频谱距离, 并有效的抑制了增强语音的背景噪声, 提高了语音可懂度。

关键词: 语音增强; 多窗谱; 改进的最小控制递归平均; 几何谱减

中图分类号: TN912.35

文献标识码: A

文章编号: 1000-7180(2018)11-0062-05

The Low SNR Speech Enhancement Method Based on Multi-Taper Spectrum Estimation and Spectral Subtraction of Geometric

LI Xu^{1,2,3}, HU Jun^{1,2}, LIU Xin^{2,3,4}, HUANG Shi-lei²

(1 Chongqing Key Laboratory of Optical Communication and Networks, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;

2 IMSL Shenzhen Key Lab, PKU-HKUST Shenzhen Hong Kong Institution, Shenzhen 518057, China;

3 Peking University Shenzhen institute, Shenzhen 518057, China; 4 Key Laboratory of

Intelligent Computing & Signal Processing, Ministry of Education, Anhui University, Hefei 230601, China;)

Abstract: The traditional spectral subtraction methods tend to produce “music noise” under low signal-to-noise ratio (SNR) due to the independence assumption of speech and noise. The intelligibility of speech is reduced greatly. In this paper, we proposed a speech enhancement method via integrating multi-taper spectrum (MTM) estimation and spectral subtraction of geometric (GA). Our method employs the MTM to estimate the power spectrum of the noisy speech, and the improved minima controlled recursive average method to track the estimated noise spectrum in real time. Furthermore, the GA is used to calculate the gain function and speech signal is recovered accurately. Our method can reduce the spectral distance between the noisy speech and the clean speech under the low signal-to-noise ratio prominently, and restrain the background noise of the enhanced speech effectively. Experiments on the IEEE dataset, PESQ and LSD as the evaluation metrics, show that our method improves the speech intelligibility significantly.

Key words: speech enhancement; multi-taper spectrum; improved minima controlled recursive average; spectral subtraction of geometric;

1 引言

语音通信是人类社会重要的信息交流手段之一,然而人类却生活在一个极其复杂的声学环境当中,语音通信通常受到外界各种各样背景噪声的干扰,这无疑严重污染了人类赖以交流的语音信号.因此,噪声环境下的语音信号前端处理逐渐成为一个十分重要的研究课题,语音增强(Speech Enhancement, SE)技术也随之应运而生.

语音增强就是指从噪声背景中提取有用的纯净语音信号,抑制、降低噪声干扰的技术,目的在于提高语音的质量和可懂度,提高智能设备的自动语音识别(Automatic Speech Recognition, ASR)准确率,目前已经广泛应用于移动通信、军事通信、助听设备、自动语音识别等领域^[1].

谱减法(Spectral Subtraction, SS)作为语音增强算法中简单易行且普适性高的方法,一直以来都备受青睐.然而,该方法却存在着语音和噪声相关性不大甚至相互独立等限制性假设,以及在低信噪比环境下会使增强后的语音残存大量“音乐噪声”等缺点,所以其一般只适用于高信噪比条件下的平稳噪声环境.针对传统谱减法的缺点,本文提出了改进的语音增强算法.

2 多窗谱估计

多窗谱(Multi-taper Spectrum)估计是由Thomson^[2]提出的,该方法对同一数据序列利用多个正交的数据窗分别求直接谱,然后求平均得到谱估计值,因此可以得到较小的估计方差,是一种比传统语音增强方法使用的周期图法更为准确的谱估计方法^[3-4].多窗谱定义如下:

$$S_{m\ell}(\omega) = \frac{1}{L} \sum_{k=0}^{L-1} S_{mk}(\omega) \quad (1)$$

式(1)中, L 表示数据窗个数; S_{mk} 表示第 k 个数据窗的谱:

$$S_{mk}(\omega) = \left| \sum_{n=0}^{N-1} a_k(n)x(n)e^{-j\omega n} \right|^2 \quad (2)$$

式(2)中, $x(n)$ 为数据序列; N 为序列长度; $a_k(n)$ 为第 k 个数据窗,它满足多个数据窗之间相互正交,即:

$$\begin{cases} \sum a_k(n)a_j(n) = 0 & k \neq j \\ \sum a_k(n)a_j(n) = 1 & k = j \end{cases} \quad (3)$$

数据窗是一组相互正交的离散椭圆序列(Dis-

crete Prolate Spheroidal Sequences, DPSS),也叫做Slepian窗.

3 噪声估计

噪声估计是语音增强算法极其重要的部分,噪声估计的准确性会直接影响最终的增强效果.如果对于噪声估计过高,则相对微弱的语音将会被去掉,增强语音就会产生较大的语音失真;而估计过低,增强语音中会残留过多的背景噪声^[5].针对噪声估计过高或过低的问题,文献^[6]提出了一种改进的最小值控制递归平均(Improved Minima Controlled Recursive Average, IMCRA)噪声动态估计算法,该算法包括两次平滑和最小值跟踪过程,第一次是对每一个频带的语音出现情况做粗略的估计,然后在第二次反复过程中,通过最小值跟踪剔除掉强语音分量,对噪声谱进行平滑.相较于传统的静态噪声估计算法、最小统计(Minima Statistics, MS)算法、Hirsch算法以及最小值控制递归平均(Minima Controlled Recursive Average, MCRA)算法等,改进的最小值控制递归平均算法计算效率更高,它能够快速跟踪噪声谱的突变,而且在信噪比相对较低时能防止弱语音段噪声估计值的偏大.因此,本文的噪声估计算法采用动态的改进最小值控制递归平均算法.

4 几何谱减

4.1 增益函数

传统谱减法通常假设语音和噪声的相关性不大甚至相互独立,因此其一般忽略了交叉项对语音增强效果的影响,然而在一定程度上这会导致增强后的语音产生大量的“音乐噪声”.针对此问题,文献^[7]中Yang Lu和Loizou提出了几何谱减算法,其利用带噪语音信号的短时相位关系推导出三个向量之间的几何关系,从而进一步计算出增益函数.将带噪语音信号、纯净语音信号和噪声信号映射到极坐标下,三者之间的几何关系为:

$$a_Y e^{j\theta_Y} = a_X e^{j\theta_X} + a_D e^{j\theta_D} \quad (4)$$

式(4)中, $\{a_Y, a_X, a_D\}$, $\{\theta_Y, \theta_X, \theta_D\}$ 分别表示带噪语音信号、纯净语音信号、噪声信号的幅度谱值和相位值,相应的几何关系图如下图1所示.

由图1的几何关系我们可以推导出 $\bar{AB} = a_Y \sin(\theta_D - \theta_Y) = a_X \sin(\theta_D - \theta_X)$,其中,令 $C_{YD} \triangleq \sin(\theta_D - \theta_Y)$, $C_{XD} \triangleq \sin(\theta_D - \theta_X)$,则可以进一步得

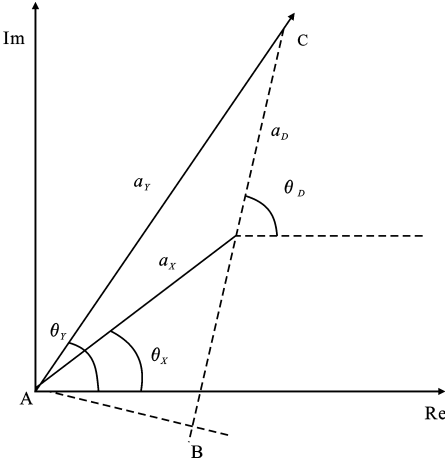


图1 几何关系图

到增益函数 H_{GA} 为:

$$H_{GA} = \frac{a_X}{a_Y} = \sqrt{\frac{1 - C_{YD}^2}{1 - C_{XD}^2}} \quad (5)$$

故估计得到纯净语音功率谱为:

$$|\hat{X}(i, k)|^2 = H_{GA}^2(i, k) \cdot |\hat{Y}(i, k)|^2 \quad (6)$$

公式(5)所表示的增益函数 H_{GA} 主要依赖于带噪语音信号(或纯净语音信号)与噪声信号之间的相位差值信息(即 C_{YD} 和 C_{XD})的估计,然而,这本身就是一项艰巨的任务,目前准确有效测定相位值的方法较少且极其复杂,同时还存在相位模糊问题.因此,可以利用三者的几何关系,由三角函数的余弦定理得:

$$C_{YD} = \frac{a_Y^2 + a_D^2 - a_X^2}{2 a_Y a_D} \quad (7)$$

$$C_{XD} = \frac{a_Y^2 - a_D^2 - a_X^2}{2 a_X a_D} \quad (8)$$

令 $\xi \triangleq \frac{a_X^2}{a_Y^2}$ 、 $\gamma \triangleq \frac{a_Y^2}{a_D^2}$ 分别表示先验信噪比和后验信噪比,则增益函数 H_{GA} 又可以写为:

$$H_{GA}(\xi, \gamma) = \sqrt{\frac{1 - (\gamma + 1 - \xi)^2}{4\gamma}} \quad (9)$$

当纯净语音信号和噪声信号的向量正交($C_{XD} = 0$)时,几何谱减法将会还原为传统的谱减算法,即表示纯净语音信号和噪声信号两者不相关的情况;而其他情况下,由于传统谱减法对于交叉项的忽略,增强语音会产生“音乐噪声”,几何谱减则恰巧能够弥补这个缺点^[7].

4.2 后验信噪比

基于当前帧的后验信噪比 $\hat{\gamma}(i, k)$ 可以通过对

带噪语音功率谱和噪声功率谱比值 $\frac{a_Y^2}{a_D^2}$ 的计算来获得,

得 $\hat{\gamma}(i, k)$ 的瞬时估计值为:

$$\hat{\gamma}_I(i, k) = \left[\frac{\hat{a}_Y(i, k)}{\hat{a}_D(i, k)} \right]^2 \quad (10)$$

式中, I 表示瞬时测量; $\hat{a}_Y^2(i, k)$ 是由多窗谱估计算法估计得到的带噪语音信号功率谱; $\hat{a}_D(i, k)$ 是由改进的最小值控制递归平均算法跟踪估计得到的噪声信号幅度谱.(i 表示第 i 帧; k 表示第 k 条谱线).而采用平滑的方法可以防止后验信噪比 $\hat{\gamma}(i, k)$ 快速变化,同时也可以限制信号的过减和 $\hat{\gamma}(i, k)$ 的过大,文献[7-9]给出了带有平滑系数 β 的后验信噪比 $\hat{\gamma}(i, k)$ 的估计方程:

$$\hat{\gamma}(i, k) = \beta \hat{\gamma}(i-1, k) + (1 - \beta) \min[\hat{\gamma}_I(i, k), 20] \quad (11)$$

其中 β 为平滑系数,而 $\min[\cdot]$ 用来限制 $\hat{\gamma}(i, k)$ 的范围,使其最大值不超过13 dB(即为 $10 \lg 20$),这就避免了信号的过大衰减.

4.3 先验信噪比

基于当前帧的先验信噪比 $\hat{\xi}(i, k)$ 可以利用增强了的前一帧幅度谱来估计计算 $\frac{a_X^2}{a_D^2}$ 获得, $\hat{\xi}(i, k)$ 的瞬时估计值为:

$$\hat{\xi}_I(i, k) = \frac{\hat{a}_X^2(i-1, k)}{\hat{a}_D^2(i-1, k)} \quad (12)$$

式(12)中, $\hat{\xi}_I(i, k)$ 表示语音信号第 i 帧第 k 条谱线的先验信噪比瞬时值,但该瞬时先验信噪比仅利用了前一帧谱信息.由式(7)可知: $\xi = \gamma + 1 - 2\sqrt{\gamma} C_{YD}$,因此可以利用该式中 ξ 和 γ 的真值关系及当前帧的谱信息来计算 ξ 的估计值.因为 $-1 \leq C_{YD} \leq 1$,可知 ξ 的范围为 $(\sqrt{\gamma} - 1)^2 \leq \xi \leq (\sqrt{\gamma} + 1)^2$,将 ξ 的下限值作为估计值,即为 $\hat{\xi}(i, k) = (\sqrt{\hat{\gamma}(i, k)} - 1)^2$,综合基于前一帧和当前帧 ξ 的两种估计值得到平滑的当前帧的先验信噪比为:

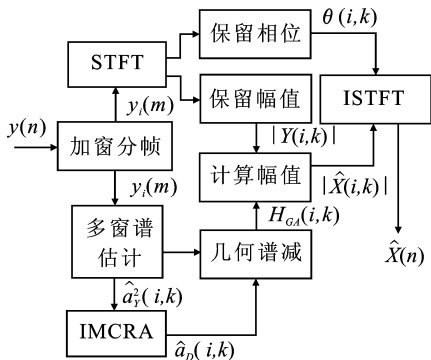
$$\hat{\xi}(i, k) = \alpha \cdot \left[\frac{\hat{a}_X^2(i-1, k)}{\hat{a}_D^2(i-1, k)} \right]^2 + (1 - \alpha) \cdot (\sqrt{\hat{\gamma}(i, k)} - 1)^2 \quad (13)$$

式中, α 是平滑系数,取值一般介于0.95~1.0之

间,式(13)是结合前一帧和当前帧的加权平均的平滑先验信噪比瞬时测量值, α 控制着前一帧和当前帧频谱信息。

5 算法系统框图及具体实现

本文提出了基于多窗谱估计和几何谱减的语音增强算法,区别于传统的谱减算法,其系统框图如图2所示。



算法具体实现步骤如下：

Step1 假定带噪语音信号为 $y(n)$,加窗分帧后为 $y_i(m)$,相邻帧之间保持百分之五十的重叠；

Step2 对分帧后的带噪语音信号 $y_i(m)$ 进行短时傅里叶变换,求出第 i 帧第 k 条谱线的幅度谱 $|Y(i,k)|$ 和相位谱 $\theta(i,k)$,并在相邻 M 帧之间做平滑处理,得到该条谱线的平均幅度谱 $|\bar{Y}(i,k)|$.将第 i 帧作为中心帧,在中心帧前后各取 M 帧(M 为正整数),平滑 $2M+1$ 帧得到:

$$|\bar{Y}(i,k)| = \frac{1}{2M+1} \sum_{j=-M}^M |Y(i+j,k)| \quad (14)$$

Step3 将分帧后的带噪语音信号 $y_i(m)$ 进行多窗谱估计,得到其第 i 帧第 k 条谱线的功率谱为 $\hat{a}_Y^2(i,k)$;

Step4 对Step3得到的功率谱估计值同样也进行相邻帧之间的平滑处理,计算得到平均功率谱 $\bar{a}_Y^2(i,k)$.将第 i 帧作为中心帧,在中心帧前后各取 M 帧(M 为正整数),平滑 $2M+1$ 帧得到:

$$\bar{a}_Y^2(k,i) = \frac{1}{2M+1} \sum_{j=-M}^M \hat{a}_Y^2(i+j,k) \quad (15)$$

Step5 利用改进的最小值控制递归平均算法跟踪估计得到第 i 帧第 k 条谱线的噪声谱 $\hat{a}_D(i,k)$;

Step6 将Step4求得的带噪语音的功率谱 $\bar{a}_Y^2(i,k)$ 和Step5求得的噪声谱 $\hat{a}_D(i,k)$ 依次代入

式(10)~式(13)求得第 i 帧第 k 条谱线的后验信噪比 $\hat{\gamma}(i,k)$ 和先验信噪比 $\hat{\xi}(i,k)$,在此基础上可以进一步求得第 i 帧第 k 条谱线的增益函数 $H_{GA}(i,k)$;

Step7 将Step2求得的 $|\bar{Y}(i,k)|$ 和Step6求得的 $H_{GA}(i,k)$ 代入公式(6)求得估计的纯净语音功率谱,在此基础上求出其幅度谱估计值 $|\hat{X}(i,k)|$;

Step8 将Step2求得的 $\theta(i,k)$ 和Step7求得的 $|\hat{X}(i,k)|$ 相结合,利用重叠相加原理和短时傅里叶逆变换恢复得到增强后的语音信号 $\hat{x}(n)$ 。

6 仿真结果及实验分析

实验所用纯净语音信号来源于IEEE语音数据库,噪声选自于Noisex-92噪声数据库.实验通过人工加噪方式,将语音数据库选取的任意一句纯净语音信号 $x(n)$ 与噪声数据库随机选取的某种噪声信号 $d(n)$ 按照某个指定的信噪比进行随机混合叠加(实验结果展示了一5 dB、0 dB和5 dB的相关性能指标),进一步得到大量实验所需带噪语音信号 $y(n)$,并将与其对应的纯净语音信号 $x(n)$ 作为评价增强效果的对比标准语音信号(实验结果展示了babble噪声、restaurant噪声和car噪声的相关性能指标在大量实验下的平均值)。

本文选择语音质量感知评价(Perceptual Evaluation of Speech Quality, PESQ)和对数谱距离(Log-Spectral Distance, LSD)^[10]作为语音增强效果的评价标准.实验对比分析了传统语音增强算法和本文提出的改进语音增强算法在不同噪声环境背景下的语音增强效果,所得实验结果分别如表1和表2所示.其中表1反映了不同噪声环境下不同语音增强算法所得增强语音的PESQ值,表2反映了其LSD值(为方便表示,表1、2中用SS代表传统谱减法、MMSE代表最小均方误差估计方法、GA代表几何谱减方法、MGS代表本文所提基于多窗谱估计和几何谱减的方法)。

由表1可以看出,当输入信噪比分别为-5 dB、0 dB和5 dB时,在babble和restaurant背景噪声环境下,本算法的PESQ值要明显高于传统谱减算法和最小均方误差估计等方法,尤其在低信噪比(-5 dB)的restaurant背景噪声情况下,本算法的PESQ值高出最小均方误差估计方法0.5.而在car背景噪声环境下,本算法在信噪比较低(-5 dB和0 dB)时,性能虽优于谱减法和几何谱减方法,但略差

于最小均方误差估计方法;在信噪比较高(5 dB)时,增强效果不及最小均方误差估计和几何谱减方法.

表 1 不同噪声环境下的 PESQ

SNR	噪声	SS MMSE	GA	MGS
-5dB	babble	1.309 6	1.183 4	1.497 6
	restaurant	1.368 6	1.082 3	1.427 2
	car	1.289 4	1.590 8	1.256 9
0dB	babble	1.664 2	1.678 7	1.765 9
	restaurant	1.703 6	1.631 8	1.758 2
	car	1.561 3	2.137 1	1.795 1
5dB	babble	2.019 3	2.104 9	2.156 7
	restaurant	1.991 9	2.020 9	2.167 0
	car	2.085 1	2.301 3	2.246 6

表 2 不同噪声环境下的 LSD

SNR	噪声	SS MMSE	GA	MGS
-5dB	babble	2.666 7	2.291 6	2.509 1
	restaurant	2.595 7	2.106 7	2.213 5
	car	2.846 8	2.103 0	2.627 7
0dB	babble	2.514 1	1.974 4	2.097 1
	restaurant	2.397 2	1.902 1	1.943 3
	car	2.606 7	1.888 0	2.139 7
5dB	babble	2.199 9	1.726 1	1.767 9
	restaurant	2.096 8	1.677 0	1.685 7
	car	2.282 3	1.691 0	1.852 9

由表 2 可以看出,当输入信噪比较低时(-5 dB),无论在 babble、restaurant 还是 car 背景噪声环境下,本算法的 LSD 值好于传统谱减方法,和几何谱减法的性能相当.这主要由于传统谱减方法增强的结果中残存大量的“音乐噪声”,可能导致它的 LSD 值相对较高,这也说明本算法缩短了带噪语音信号与纯净语音信号之间的频谱距离.但相比基于谱距离测度设计的最小均方误差方法相比,性能略差.随着信噪比的逐渐增大(0 dB 和 5 dB),除最小均方误差方法以外,其余几种方法的 LSD 值相对比较接近,本算法的 LSD 值虽略微优于传统谱减方法,但相较于其他两种方法,其性能略差.

7 结束语

为提高低信噪比条件下带噪语音的可懂度,本文提出了基于多窗谱估计和几何谱减的语音增强算法,本算法利用多窗谱估计带噪语音功率谱,利用 IMCRA 动态跟踪估计噪声谱,利用几何谱减法求得精确的增益函数,并进一步得到增强语音的幅度谱,最后采用重叠相加原理和短时傅里叶逆变换得到增强的语音信号.通过大量的仿真验证,表明在信噪比较低的情况下,本算法相比于绝大多数传统的语

音增强方法取得了明显的增强效果,尤其在提高语音可懂度方面.然而和大多数语音增强算法一样,实验在利用重叠相加原理恢复语音信号时,相位选取的依然是原始带噪语音信号的相位,这无疑在一定程度上限制了该方法性能的上限.为了得到更精确的纯净语音信号估计,未来将进一步在纯净语音信号相位估计方面开展大量工作.

参考文献:

[1] 赵力,梁瑞宁,魏昕,等. 语音信号处理 [M]. 3 版. 北京:机械工业出版社,2016.

[2] Thomson D J. Spectrum estimation and harmonic analysis[J]. IEEE,1982, 70(9):1055-1096.

[3] 张青,吴进. 基于多窗谱估计的改进维纳滤波语音增强[J]. 计算机应用与软件,2017,34(3):67-71.

[4] 张正文,周航麒. 基于多窗谱估计的维纳滤波语音增强算法[J]. 河南理工大学学报(自然科学版),2015, 34(5):686-690.

[5] 张建伟,陶亮,周健,等. 噪声谱估计算法对语音可懂度的影响[J]. 声学技术,2015,34(5):424-430.

[6] Cohen L. Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging [J]. IEEE Transactions on Speech and Audio Processing, 2003,11(5): 466-475.

[7] Yang Lu, Loizou C. A geometric approach to spectral subtraction[J]. Speech Communication, 2008, 50(6): 453-466.

[8] 孟祥彩,王中训,刘伟,等. 基于相位影响的单声道语音增强改进算法[J]. 电声技术,2016, 40(11):59-61.

[9] 万义龙,张天骐,王金朝,等. 一种基于几何谱减法和听觉掩蔽效应的语音增强方法[J]. 微电子学与计算机,2014,31(2):80-84.

[10] 张学英. 数字语音处理及 MATLAB 仿真 [M]. 2 版. 北京:电子工业出版社,2016.

作者简介:

李 涓 男,(1993-),硕士研究生. 研究方向为语音增强、语音端点检测. E-mail:loveli5991529@163.com.

胡 俊 男,(1990-),硕士研究生. 研究方向为音频检索、语音识别.

刘 新 男,(1992-),硕士研究生. 研究方向为语音标点预测、语音识别.

黄石磊 男,(1979-),博士,副研究员,高级工程师. 研究方向为语音信号处理.