应用谱减法及其改进型算法进行语音增强

中国电子科技集团公司第54研究所　杨秋成　　空军工程大学　蔡　斌　陈娅冰

摘　要　该文主要研究了谱减法及其改进型算法在语音增强中的应用。在对谱减法及其几种改进型算法的理论分析基础上，进行了对比仿真实验研究。客观测试和非正式听音测试表明，增强后的语音的信噪比及音质均有明显改善，其中基于先验信噪比估计的改进型谱减法效果最好。

关键词　语音信号处理　语音增强　谱减法及其改进型算法

1　引言

语音增强是语音信号处理的一个重要分支，一直是非常活跃的研究领域。语音增强的目的是改善音质，提高清晰度和可懂度，减少听觉疲劳。由于信号的相位对语音的感知并不重要，所以基于短时谱幅度(STSA)的增强方法应用最为广泛，谱减法及其改进型算法是其中的典型代表。

本文主要研究了基本谱减法、基于先验信噪比估计的谱减法、利用人耳听觉掩蔽效应的谱减法等几种算法在语音增强中的应用。在对以上几种算法的理论分析基础之上，进行了对比仿真试验研究，并给出了实验和分析结果。

2　基本谱减法

带噪语音 $y(n)$一般可表示为：

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

$s(n)$为纯净语音，$d(n)$为加性平稳高斯噪声，二者不相关。谱减法中，带噪语音的短时谱幅度(STSA)$|Y(w)|$减去噪声谱的估计值$|\hat{D}(w)|$后可以得增强后语音的短时谱幅度 STSA$|\hat{S}(w)|$，即：

$$|\hat{S}(w)|^2=\begin{cases}|Y(w)|^2-|\hat{D}(w)|^2, & \text{如果}|Y(w)|^2>|\hat{D}(w)|^2\\ 0 & \text{其他}\end{cases} \tag{2}$$

由于人耳对语音信号的相位不敏感，$|\hat{D}(w)|^2$可在无音段估计得到。带噪语音的相位直接与$|\hat{S}(w)|$相乘恢复出增强后的语音，即：

$$\hat{S}(n)=IFFT[\,|\hat{S}(w)|\cdot\exp(j\arg Y(w))] \tag{3}$$

谱减法也可用线性时变滤波器形式来表示，即对$|Y(w)|$乘以增益函数 $G(w)$将(2)式变为乘积形式：

$$|\hat{S}(w)|=G(w)\cdot|Y(w)| \qquad 0\le G(w)\le 1 \tag{4}$$

对应于式(2)，则：

$$G(w)=\sqrt{1-\frac{|\hat{D}(w)|^2}{|Y(w)|^2}} \tag{5}$$

如果$|\hat{D}(w)|^2>|Y(w)|^2$，则 $G(w)=0$，这样可保证 $G(w)$为实函数。从式(4)、式(5)中可以清楚地看出谱减法的物理意义：它相当于对带噪语音的每一个频谱分量乘以一个系数 $G(w)$。当该段只含语音时，没有任何衰减，$G(w)=1$；而当该段只含噪声时，衰减最大，$G(w)=0$。当介于两者之间时，$G(w)$由后验信噪比决定，即：

$$SNR_{post}^{(w)}=\frac{|Y(w)|^2}{|\hat{D}(w)|^2} \tag{6}$$

在实际的增强过程中，更多地使用的是谱减法的推广形式：

$$G(w)=G[SNR_{post}(w)]=\begin{cases}\left(1-\alpha\cdot\left[\dfrac{|\hat{D}(w)|}{|Y(w)|}\right]^{\gamma_1}\right)^{\gamma_2}, & \text{如果}\left[\dfrac{|\hat{D}(w)|}{|Y(w)|}\right]^{\gamma_1}<\dfrac{1}{\alpha+\beta}\\ \left(\beta\cdot\left[\dfrac{|\hat{D}(w)|}{|Y(w)|}\right]^{\gamma_1}\right)^{\gamma_2}, & \text{其他}\end{cases} \tag{7}$$

式(7)是谱减法最为灵活的一种形式，它包含谱减法的基本思想，而且给出了三个调节系数，以在噪声抑制、剩余噪声衰减和语音失真之间达到最好的折衷。其中：

1)过减系数 $\alpha(\alpha>1)$：α 值越大，剩余噪声衰减越大，同时语音失真也会越大。

2)谱平滑系数 $\beta(0<\beta<1)$：β 值增大可降低剩余的音乐噪声，但会增加增强后语音的背景噪声。

3)指数 $\gamma=\gamma_1=1/\gamma_2$：这个参数决定了增益函数从 $G(w)=0$ 到 $G(w)=1$ 的平滑程度。

谱减参数 α、β 和 γ 的选择是谱减法的核心问题。实际上，在低信噪比条件下，减小语音失真和降低剩余噪声不可兼得，只能在二者之间达到最好的折衷，提高可懂度。

语音信号中，说话人由于呼吸会不断产生语音间歇，我们可以利用这些间歇估计背景噪声，其中一种方法就是利用端点检测(VAD)来判定有/无语音。在无音段利用下式对噪声估计进行更新。

$$|\hat{D}_i(w)|=\eta|\hat{D}_{i-1}(w)|+(1-\eta)|Y(w)| \tag{8}$$

其中，i 为当前帧数，$i-1$ 为前一帧。

谱减法中，由于是利用无音期间加权平均值来代替当前分析帧各频率点的噪声频谱分量，且噪声

收稿日期：2003 年 5 月 15 日

频谱服从高斯分布，及其幅度随机变化范围很宽。因此在相减时，若该帧某频率点噪声分量较大，就会产生很大一部分的残留噪声，在频谱上呈现随机出现的尖峰，在听觉上形成节奏性起伏的类似音乐的噪声，通常称之为“音乐噪声”，这是在谱减法中常出现的，也是较难解决的问题。为了有效抑制“音乐噪声”，我们可以对原始的基本谱减法进行改进，得到改进型的谱减法。

3 改进型谱减法

3.1 基于听觉掩蔽效应的改进型谱减法

增强语音在很多情况下是直接为听觉服务的，所以应该结合人耳听觉特性来提高增强语音的听觉效果，其中将听觉掩蔽效应与基本谱减法相结合可获得较好的增强效果。掩蔽效应是指一个声音的存在会对另一个声音的感知产生掩蔽效应，主要发生在同时进入听觉系统的不同频率的两个声音之间，即同时掩蔽。其主要算法流程见图 1 所示。

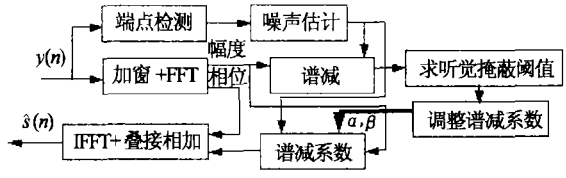


图 1 基于听觉掩蔽效应的改进型谱减法主要算法流程

主要算法步骤：

1)加窗分帧、进行 N 点 FFT 变换得到带噪语音频谱；

2)端点检测，在无音段进行噪声估计；

3)利用基本谱减法得到语音频谱的粗估计，由此计算听觉掩蔽阈 $T(w)$；

4)根据 $T(w)$来调节谱减系数 α,β；

5)利用调整后的 α,β 的进行系数谱减；

6)IFFT，用叠接相加法得到 $\hat{s}(n)$。

其中，计算听觉掩蔽阈 $T(w)$是最为关键的一步，它可由以下计算步骤得到：

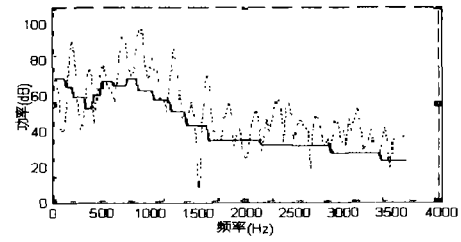
1)计算 Bark 谱：将临界带内信号的功率谱相加，即可得到 Bark 谱；

2)计算扩展谱：Bark 谱反映的只是临界带内被激励的情况，反映各临界带间响应情况的谱称之为扩展谱，扩展谱由 Bark 谱与扩展函数卷积得到，即任一临界带处的扩展谱应是各临界带 Bark 谱贡献的总和；

3)减去偏移量：将扩展谱减去一个偏移量即可得到扩展谱意义下的掩蔽阈值；

4)归一化并计入绝对听阈。

其中，临界带的划分、扩展函数和偏移量具体的计算方法可见参考文献[3]。图 2 为一帧听觉掩蔽阈的计算结果。



实线：听觉掩蔽阈值　虚线：语音粗估计

图 2 一帧听觉掩蔽阈的计算结果

在根据 $T(w)$来调节谱减系数 α,β 时，听觉掩蔽阈值较大说明此 Bark 关键频率段中人耳对其他相近频率段的语音信号和噪声信号的抗干扰能力较强，所以应采用较小的谱减阈值系数；反之对于听觉掩蔽阈值较小的 Bark 关键频率段，应采用较大的谱减阈值系数。α,β 的值应由下式确定：

$$\begin{aligned}\alpha(w) &= F_\alpha[\alpha_{\min},\alpha_{\max},T(w)]\\ \beta(w) &= F_\beta[\beta_{\min},\beta_{\max},T(w)]\end{aligned} \tag{9}$$

$\alpha_{\min},\beta_{\min}$和 $\alpha_{\max},\beta_{\max}$分别为 α,β 的最小值和最大值。F_α 和 F_β 应满足：当 $T(w)=T(w)_{\min}$时，$F_\alpha=\alpha_{\max}$；当 $T(w)=T(w)_{\max}$时，$F_\alpha=\alpha_{\min}$。$T(w)_{\max}$和 $T(w)_{\min}$是每一帧语音的听觉掩蔽阈值的最大值和最小值。通过大量的实验，兼顾提高信噪比和保证语音信号可懂度和清晰度，减小音乐噪声，选择 $\alpha_{\min}=1$ 和 $\alpha_{\max}=6$；$\beta_{\min}=0$和 $\beta_{\max}=0.02$；$\gamma=\gamma_1=2$ 和 $\gamma_2=1/\gamma_1=0.5$。

则 F_α 和 F_β 的值可通过所给出的参数拟合得到，即：

$$\begin{aligned}\frac{T(w)_{\max}-T(w)}{\alpha-\alpha_{\min}} &= \frac{T(w)-T(w)_{\min}}{\alpha_{\max}-\alpha}\\ \frac{T(w)_{\max}-T(w)}{\beta-\beta_{\min}} &= \frac{T(w)-T(w)_{\min}}{\beta_{\max}-\beta}\end{aligned} \tag{10}$$

3.2 基于先验信噪比估计的改进型谱减法

Ephraim 和 Malah 提出的最小均方误差估计(MMSE)增强方法可以有效地抑制“音乐噪声”，cap$\grave{e}$在文献[2]中证明：由于其在计算增益函数时引入了先验信噪比，并采用了“Decision-Directed”(直接判决)法，简称为“D－D”法，进行先验信噪比的估计，所以取得了较好的增强效果。我们同样可以将这种方法引入到谱减法中，得到基于先验信噪比估计的改进型谱减法。先验信噪比 SNR_{prio}定义为：

$$SNR_{prio}(w)=\frac{|S(w)^2|}{|\hat{D}(w)^2|} \quad (11)$$

首先,将增益函数表示成先验信噪比的形式,即利用 $SNR_{post}(w,i)=1+SNR_{prio}(w,i)$。其中,i 为帧数。则式(5)可表示为:

$$G(w,i)=\sqrt{\frac{SNR_{prio}(w,i)}{1+SNR_{prio}(w,i)}}$$

其中,$SNR_{prio}(w,i)$用"D-D"法进行估计,即:

$$SNR_{prio}(w,i)=\eta\cdot\frac{|\hat{S}(w,i-1)|^2}{|\hat{D}(w)|^2}+(1-\eta)\cdot\max[SNR_{post}(w,i)-1,0] \quad (12)$$

其中,i 为当前帧,$i-1$ 为前一帧;$\hat{S}(w,i-1)$为前一帧语音的估计结果;η 为调节系数,一般在0.8—1之间;max(•,•)为两者之中取较大的值。从式(12)可以看出,先验信噪比 SNR_{prio} 是通过非线性的递推估计得到的。

4 实验结果及分析

对录制得到的纯净语音,添加取自 NOISE92x 的平稳高斯白噪声,输入信号的信噪比为 -5dB。对带噪语音进行 8kHz 采样,16 位线性量化,采用汉宁窗分帧,每帧 256 个采样点,帧间叠加 192 点。增强得到的语音利用加权叠加相加法进行恢复。比较主观听觉效果和实验结果的时域波形和语谱图可以发现:基本谱减法算法思想简单,运算量低,易于实时实现,但在信噪比较低的条件下剩余噪声和"音乐噪声"均较大;基于听觉掩蔽效应的谱减法运算量较大,增强效果比基本谱减法有一定的提高;基于先验信噪比估计的谱减法,性能最好,剩余噪声和"音乐噪声"均大大降低,且运算量与基本谱减法相当。

参考文献

[1]Pascal Scalart, Jozue Vieira Filho.: Speech Enhancement based on a priori signal to noise estimation. ICASSP, 1996

[2]Oliver Cappè.: Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise suppressor. IEEE Vol.2 Transactions on Speech and Audio Processing. 1994

[3]Virag. N.: Single channel speech enhancement based on masking properties of human auditory system. IEEE Trans Speech Audio process. 1999; 7,(2):126-137.

[4]杨行峻,迟惠生.语音信号数字处理[M].北京:电子工业出版社,1995

作者简介

杨秋成 男,中国电子科技集团公司第54研究所高级工程师,现从事无线通信工作。

蔡 斌 男,空军工程大学硕士研究生在读,主要从事语言信息处理和自控信号处理。

陈娅冰 女,空军工程大学硕士研究生在读,主要从事光纤通信。

(上接第17页)

3 几种合并方式的性能比较

目前我们经常采用的合并方式有选择式合并、等增益合并和最大比值合并,这几种方式各有其特点。确定使用何种合并方式一般在要设备的复杂程度和性能之间进行折中。最大比值合并性能最好,但设备较复杂;选择式合并则最简单,但比较而言,性能降低。在上述系统中,为了有效的抗多径衰落,采用了4重分集。但由于是机载系统,为了易于实现和简化设备,利用一付天线采用极化分集和频率分集的方式完成。而合并方式也采用了混合分集方式,兼顾了设备的复杂性和性能。其性能应优于选择式合并,但劣于等增益合并。因此,这里计算三种合并方式的性能,进行定量的比较。从分集的效果而言,一般极化分集信号之间有一定的相关性,频率分集较极化分集效果好,因此,从实际情况出发,我们重点比较分集信号间具有相关性的情况,并假设频率分集的相关性小于极化分集。

4 结论

根据计算数据和模拟结果,绘出了误码性能曲线,其中典型的数据列于表1。

由表1可以得到,在各分集支路都是独立的情况下,选择性合并较等增益合并损失了2.4dB,而混合合并方式的性能比单纯采用等增益合并性能低1.5dB,但比全部采用选择性合并好0.9dB。当存在相关性的时候,分集接收的性能有损失,对于表1中所给出的情况,各种分集合并方式都损失了2dB左右。因此,对于地空通信采用分集接收可以有效地改善远距离通信和低空条件下的误码性能,在进行数据传输时尤为重要。而采用混合合并的方式对于机载设备是一种较好的选择,既满足了简化设备的要求,又能得到较好的性能。

表1 三种合并方式误码性能的比较

单通道 SNR(dB) / 合并方式	误码率($P_e=1\times10^{-3}$)		
	$\rho_f=0$, $\rho_p=0$	$\rho_f=0.6$, $\rho_p=0.3$	$\rho_f=0.3$, $\rho_p=0.6$
等增益	4.9	7.0	7.0
等增益+选择式	6.4	8.45	8.67
选择式	7.3	9.35	9.4

参考文献

[1]刘圣民,熊兆飞编.对流层散射通信技术.国防工业出版社,1982

[2]吴祈耀编.随机过程.国防工业出版社,1984

作者简介

徐松毅 男,中国电子科技集团公司第54研究所微波散射部高级工程师,主要研究方向为散射与微波通信系统总体。