基于帧间重叠谱减法的语音增强算法及实现

李宏伟，段艳丽，郭　英
(空军工程大学 电讯工程学院，陕西 西安 710077)

摘　要：采用短时谱分析、合成技术，对含噪语音进行帧间重叠谱减法消除噪音，这种算法符合语音特性连续变化的特点。实验证明，该方法有效去除了噪声干扰，得到了增强语音，保证了话音的可懂度和自然度不受损失。

关键词：时变谱；谱减法；语音增强
中图分类号：TN912.3　　**文献标识码**：A　　**文章编号**：1009-3516(2001)05-0048-03

谱减法[1]是处理宽带噪声最通用的技术，其基本思想是从含噪语音谱中减去噪声频谱估值。假定噪声为加性和不相关的，则含噪语音

$$x(m)=s(m)+e(m) \qquad -\infty<m<\infty \tag{1}$$

式中，$s(m)$和$e(m)$分别为语音和噪声成分。若用$X(n,\omega)$、$S(n,\omega)$、$E(n,\omega)$表示它们在n附近的L点傅立叶变换，则

$$X(n,\omega)=S(n,\omega)+E(n,\omega) \tag{2}$$

由于$s(m)$与$e(m)$独立，所以，$E[|X(n,\omega)|^2]=E[|S(n,\omega)|^2]+E[|E(n,\omega)|^2]$

对一个分析帧内的短时平稳过程，有

$$|S(n,\omega)|^2=|X(n,\omega)|^2-|E(n,\omega)|^2 \tag{3}$$

若$S(n,\omega)$的相位用$X(n,\omega)$近似，则

$$y(m)=IFT\{|S(n,\omega)|\exp(X(n,\omega))\} \tag{4}$$

整个算法过程如图1所示。噪声方差由对噪声$e(m)$的统计特性估计而得到。对含噪语音的每一帧进行同样处理，再将每帧恢复语音依次连接起来，就是消除噪音的增强语音。但由于恢复信号帧间不连续，使得恢复的语音含有一周期性的“嘟嘟”背景声，因此，提出用帧间重叠谱减法进行语音增强。

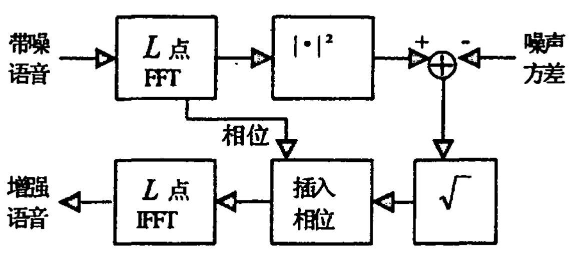


图1　基本谱减法原理框图

1　帧间重叠谱减法

1.1　分析、合成原理

设$x(m)$ $(-\infty<m<\infty)$是取样频率为F_S的含噪语音序列，$w(m)$ $(0\leqslant m\leqslant N-1)$为一窗序列。那么，位于$n$附近的语音段的傅立叶变换为

$$X(n,\omega)=\sum_{m=-\infty}^{\infty}w(n-m)x(m)e^{-j\omega m} \tag{5}$$

可见，$X(n,\omega)$是n和ω的二维函数，为时变谱。因此，对随机语音序列时—频分析，就是使窗$w(n-m)$沿着$x(m)$序列滑动，截取加窗语音段$w(n-m)x(m)$，对其进行傅立叶变换。窗的滑动速率即为$X(n,\omega)$时域抽样率F_s'，由时域抽样定理知，$F_s\geqslant 2B$[2]。B为窗的单边带宽，对汉明窗，$B=2F_S/N$。所以，$w(m)$采用汉明窗时，

收稿日期：2001-02-21
基金项目：空军科研部基金资助(KJ00104A)
作者简介：李宏伟(1966-)，男，陕西澄城人，讲师，硕士，主要从事信号与信息处理研究.

$$F_s \geqslant 4F_S/N \tag{6}$$

这样,只需每隔 $R=N/4$ 个样值,窗移动一次,计算 $w(n-m)x(m)$ 的傅立叶变换。即可得满足时域抽样定理、保持语音时变特性的频谱。显然,帧间重叠为 $3N/4$。所以 n 的取值为 $n=\mathrm{int}[m/4]\quad -\infty<m<\infty$。

离散傅立叶变换的点数 L 应满足频域抽样定理

$$L \geqslant N \tag{7}$$

以 $X(n,\omega_k)(k=0,1,\cdots L-1)$ 表示 $X(n,\omega)$ 在[0,2]间的 L 点等间隔采样,即 $w(n-m)x(m)$ 的 L 点 DFT·

$$X(n,\omega_k)=X(n,\omega)\mid_{\omega=\omega_k} \quad (8) \qquad \omega_k=2\pi k/L \quad (0\leqslant k\leqslant L-1) \tag{9}$$

由式(5)得

$$X(n,\omega_k)=\sum_{m=-\infty}^{\infty} w(n-m)x(m)e^{-j\omega_k m}=e^{-j\omega_k n}\sum_{m=-\infty}^{\infty} x(n-m)w(m)e^{j\omega_k m} \tag{10}$$

令 $h_k(m)=w(m)e^{j\omega_k m}$(即为中心频率位于 ω_k 的带通滤波器的冲激响应),则

$$X(n,\omega_k)=x(n)*h_k(n)e^{-j\omega_k n}=s(n)*h_k(n)e^{-j\omega_k n}+e(n)*h_k(n)e^{-j\omega_k n}=S(n,\omega_k)+E(n,\omega_k) \tag{11}$$

令 $$y_k(n)=s(n)*h_k(n),\text{则 } y_k(n)=S(n,\omega_k)e^{-j\omega_k n} \tag{12}$$

当 $\omega=\omega_k$ 时,分析谱减通道如图2所示。所有通道输出之和 $y(n)=\sum_{k=0}^{L-1}y_k(n)=\sum_{k=0}^{L-1}\sum_{m=-\infty}^{\infty}s(n-m)h_k(m)$

$$=\sum_{m=-\infty}^{\infty}s(n-m)\sum_{k=0}^{L-1}h_k(m)=\sum_{m=-\infty}^{\infty}s(n-m)\sum_{k=0}^{L-1}w(m)e^{j\omega_k m}=\sum_{m=-\infty}^{\infty}s(n-m)w(m)\sum_{k=0}^{L-1}e^{j\frac{2\pi}{L}km}=Lw(0)s(n) \tag{13}$$

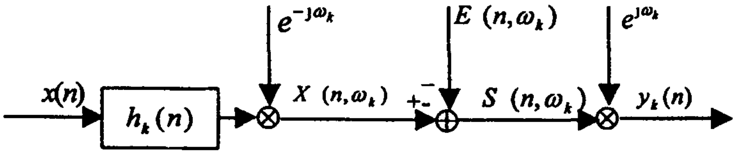


图2 第 k 个谱减分析通道

$y(n)$ 与语音信号 $s(n)$ 仅差了一个常系数。这样就证明了将谱减法用于滤波器组分析合成,可以从含噪语音中恢复出纯净语音。恢复语音

$$y(n)=\sum_{k=0}^{L-1}S(n,\omega_k)e^{j\omega_k n}=\sum_{k=0}^{L-1}[X(n,\omega_k)-E(n,\omega_k)]e^{j\omega_k n} \tag{14}$$

1.2 插值实现

频谱分析时并不需要对原始序列的每一个样点计算变换,只需要在输入信号每隔 R 个取样时才计算一次 $X(n,\omega_k)$。由 $X(n,\omega_k)$ 增强处理后得到的 $y(n)$ 序列是原始语音序列 $s(m)$ 每隔 R 样点的抽样序列。因此,以 $S(n,\omega_k)$ 合成 $s(m)$ 时,应用插值法填充 n 不等于 R 的整倍数时的 $S(n,\omega_k)$ 值[2]。为此,定义

$$V(n,\omega_k)=\begin{cases}S(n,\omega_k), & n=0,\pm R,\pm 2R,\cdots\\ 0, & n\neq 0,\pm R,\pm 2R,\cdots\end{cases} \tag{15}$$

当 ω_k 为某一定值时,$S(n,\omega_k)$、$V(n,\omega_k)$ 均为时间序列,简记为 S_n、V_n,其频谱记为 $S_n(j\Omega)$、$V_n(j\Omega)$。因为

$$S_n(j\Omega)=V_n(j\Omega) \tag{16}$$

而 S_n 的抽样率为 F_s,因此,可使 V_n 通过一截止频率为 $\frac{\pi}{R}$ 弧度的低通滤波器实现插值。该低通滤波器的冲击响应为 $g(m)$ 关于 $m=0$ 对称,为有限时宽,其宽度等于 $2RQ-1$,这里 Q 为任意正整数。内插后 $S(n,\omega_k)$ 为

$$Y(n,\omega_k)=\sum_{m=n-(RQ-1)}^{n+(RQ-1)}V(m,\omega_k)g(n-m) \tag{17}$$

$$y(n)=\sum_{k=0}^{L-1}Y(n,\omega_k)e^{j\frac{2\pi rk}{L}}=\sum_{k=0}^{L-1}\sum_{m=n-(RQ-1)}^{n+(RQ-1)}V(m,\omega_k)g(n-m)e^{j\frac{2\pi rk}{L}} \tag{18}$$

令 $v(m,r)=\sum_{k=0}^{L-1}V(m,\omega_k)e^{j\frac{2\pi rk}{L}}$ (19) 则 $y(n)=\sum_{m=n-(RQ-1)}^{n+(RQ-1)}v(m,n)g(n-m)$ (20)

将式(15)代入式(19),得

$$v(m,r)=\begin{cases}\sum_{k=0}^{L-1}V(m,\omega_k)^{j\frac{2\pi rk}{L}}, & m=0,\pm R,\pm 2R,\cdots\\ 0, & m\neq 0,\pm R,\pm 2R,\cdots\end{cases} \tag{21}$$

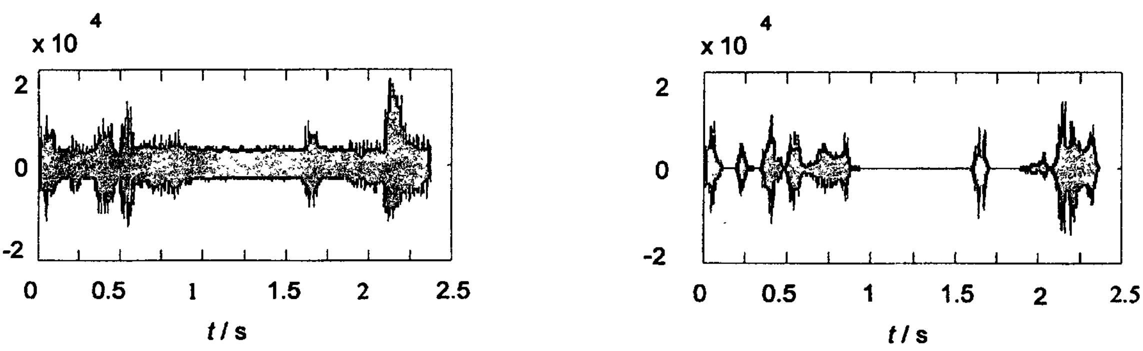
2 计算流程

由以上分析可知,去噪增强语音的过程是采用基本谱减法得到重叠各帧的语音频谱,利用相邻几个重叠帧的语音频谱,通过差值运算合成出语音信号。具体求得 $y(n)$ 的计算分以下几步进行:1)对每帧信号进行语音有无判定,在无语音期间估计出噪声谱;2)求得加窗段信号 $w(n-m)x(m)$ 的 L 点 FFT 得 $X(n,\omega_k)$;3)使窗 $w(n-m)$ 以步长 R 沿着 $x(m)$ 序列滑动,在 n 等于0, $\pm R$, $\pm 2R$,…求得 $X(n,\omega_k)$;4)假定 $s(m)$ 与 $e(m)$ 独立,对于帧内的短时平稳过程,由 $|S(n,\omega)|^2=|X(n,\omega)|^2-|E(n,\omega)|^2$ 进行谱减运算;5)由式(15)构造 $V(n,\omega_k)$;6)由式(21)计算 $v(m,r)$;7)由式(20)对 $v(m,r)$ 插值,便得到 $y(n)$。

3 应用事例

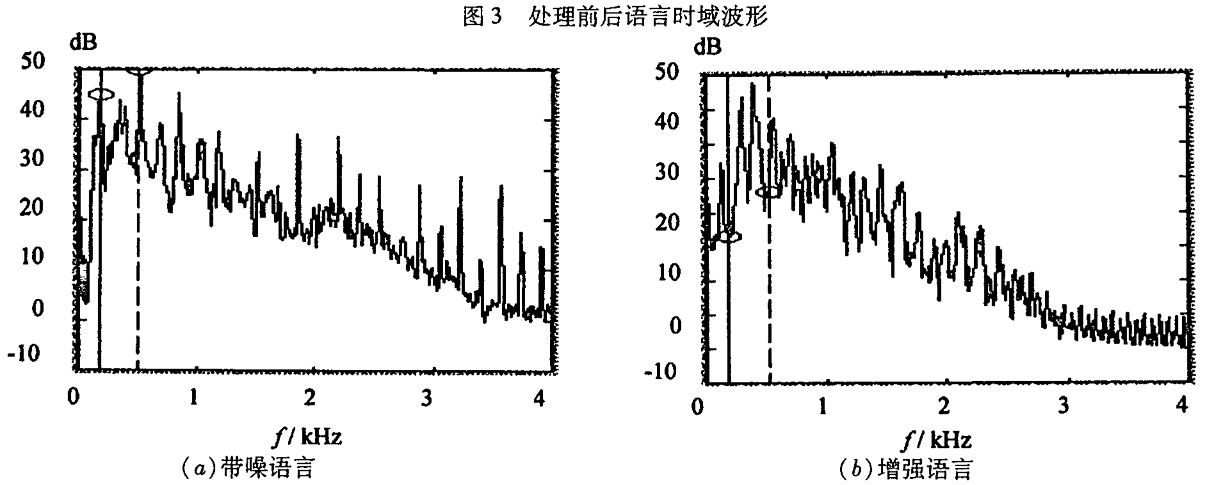
将上述算法在某一超短波通信系统的干扰抑制应用中,已取得了明显的消噪效果。该系统受到无法摆脱的周期性干扰,干扰谱为某一相对稳定的基波及其各次谐波之和,严重影响语音信号的接收,造成电台使用者产生明显的听觉疲劳。

图3(a)为受干扰信号,图3(b)是采用帧间重叠谱减法得到的增强语音,其对应的频谱幅度如图4(a)(b) 所示。从图中看出,噪声强度很大,噪声谱与语音谱重叠。这里,帧长 $N=240$,帧间重叠 $R=60$,FFT 的点数 $L=256$,$Q=2$。可以看出,该方法对噪声的各次谐波消除得比较彻底。非正式试听表明,该方法消除了周期性干扰在电台话音输出端引起的啸叫噪声,除清晰度略下降外,语音的可懂度和自然度未受损失。该方法对白噪声干扰消噪效果更好。若采用传统的梳状滤波器进行消噪处理,势必消除阻带内的语音谱,使话音质量受到损失[3-4]。采用普通谱减法进行消噪,恢复的语音含有一周期性的"嘟嘟"背景声,这是由于恢复信号帧间不连续造成的。



(a)带噪语言 (b)增强语言

图3 处理前后语言时域波形



(a)带噪语言 (b)增强语言

图4 处理前后语言频谱幅度

(下转第54页)