基于多带的 2.4 kbit/s 波形内插算法

汤一彬　　黄　蓉　　吴镇扬
(东南大学信息科学与工程学院　南京　210096)

摘　要：由于传统特征波形内插语音编码算法对特征波形相位信息的忽略，以及对特征波形的整体对齐，往往造成语音高频谐波分量丢失，从而导致语音的噪声感。为了提高合成语音的质量，该文引入语音多带清浊音标志，并以此为依据对波形内插编码模型中的慢渐变波形和快渐变波形的相位谱进行估计，在语音合成时则对特征波形采取部分对齐的方法，最后提出了一种基于多带的 2.4 kbit/s 特征波形内插算法。与传统算法相比，新算法明显提高了语音的清晰度。与标准 2.4 kbit/sMELP 算法相比，该算法合成语音质量亦略显优势。

关键词：语音编码；波形内插；多带
中图分类号：TN912.3　　**文献标识码**：A　　**文章编号**：1009-5896(2010)05-1077-06
DOI: 10.3724/SP.J.1146.2009.00613

Multiband Waveform Interpolation Speech Coding Algorithm at 2.4 kbit/s

Tang Yi-bin　　Huang Rong　　Wu Zhen-yang
(*School of Information Science and Engineering*, *Southeast University*, *Nanjing* 210096, *China*)

Abstract: In the traditional characteristic waveform interpolation speech coding, the high frequency harmonics of the synthetic speech are usually lost, which makes the speech feel noisy, due to the phase information neglect of the characteristic waveform and the whole characteristic waveform alignment. In order to improve the synthetic speech quality, the multiband surd/sonant flag is first introduced. The phase spectrum of the slowly/rapidly evolving waveform is estimated with the waveform interpolation speech coding model depending on the multiband surd /sonant flag. Then the partial characteristic waveform alignment is used in the speech synthesis section and a 2.4 kbit/s multiband waveform interpolation speech coding algorithm is proposed finally. Compared with the traditional characteristic waveform algorithm, the new algorithm can distinctly improve the speech definition. Compared with the standard 2.4 kbit/s MELP algorithm, the synthetic speech quality is also slightly better.

Key words: Speech coding; Waveform interpolation; Multiband

1　引言

在低速率语音编码领域中，特征波形内插(Characteristic Waveform Interpolation，CWI)算法被公认为一种极具潜力的编码方法[1]。其能够将语音信号经线性预测模型后获得的残差信号通过某些处理按一定的速率提取特征波(Characteristic Waveform，CW)，并将这些特征波序列表示为由时间轴和相位轴组成的 2 维基底上的特征波表面[2]。量化时，则可对该特征波表面进行多种形式的分解，如快慢渐变波分解形式[3]，矩阵奇异值分解形式[4]，小波分解形式[5]等。但总体思想都是将特征波表面分解成为准周期成分和类噪声成分，并分别根据各成分的性质进行不同的抽样速率提取和比特量化，以达到降低编码速率的目的。

对于目前大多数 CWI 算法，其合成语音都存在一定的噪声感，语音质量并不理想。在语谱图中表现为合成语音高频谐波分量的缺失，高频部分呈现为噪声性质。以基于特征波表面快慢渐变波分解的 CWI 算法为例，特征波被分解为慢渐变波(Slowly Evolving Waveform，SEW)和快渐变波(Rapidly Evolving Waveform，REW)两部分来分别表示表面中的准周期成分和类噪声成分。由于受到编码速率的约束，传统算法往往对 SEW 和 REW 的相位不进行比特编码。由于相位谱的缺失，算法在语音合成端采用 SEW 固定相位谱和 REW 随机相位谱的简单估计来弥补。这样不可避免地造成在合成端恢复出的相邻特征波的相关性减小，使特征波高频谐波噪声化。此外，在语音合成端内插产生瞬时特征波形时，传统算法需要对相邻特征波形进行全带对齐。全带对齐虽然可以提高特征波低频谐波的强度，使语音的可懂度增加，但也同时造成了高频谐波分量的噪声化，使语音质量的清晰度和自然度下降。

2009-04-24 收到，2009-09-25 改回
国家自然科学基金(60672094)资助课题
通信作者：汤一彬　norbie2008@126.com

因此，本文着眼于语音高频谐波分量的恢复，以期望在低速率的情况下进一步提高语音质量。

本文以基于特征波表面快慢渐变波分解的CWI算法为基础，首先对语音在多个频带上的清浊音程度进行标记，以此决定 SEW 和 REW 在该频带上的相位谱取值，并在内插生成瞬时特征波时仅对特征波的低频浊音谐波进行对齐，最后提出了一种基于多带 2.4 kbit/s 的特征波形内插算法。语音测试表明，新算法在增强语音低频谐波分量的同时，也能够较好地保留高频谐波分量，提高了语音的清晰度。与传统 CWI 算法[3]相比，该算法在语音质量上有很大的提高，亦略优于 2.4 kbit/s 的 MELP 编码标准[6,7]的语音质量。

2 基于快慢渐变波分解的 CWI 算法

2.1 特征波形的分解

传统 CWI 编码器是在语音信号经过线性预测模型后获得的残差信号中按一定速率提取特征波形，然后将其时域波形进行实数域的傅里叶级数(FS)表示：

$$s(n,\phi)\approx\sum_{k=1}^{\lfloor P(n)/2\rfloor}\left[A_k(n)\cos(k\phi)+B_k(n)\sin(k\phi)\right],\quad 0\le\phi<2\pi \tag{1}$$

$$\phi=\frac{2\pi m}{P(n)},\quad 0\le m<P(n) \tag{2}$$

式(1)，式(2)中，$P(n)$ 是时刻 n 时基音长度，$A_k(n)$ 和 $B_k(n)$ 是时刻 n 时所提特征波第 k 次谐波(对应频率为 $kf_s/P(n)$ Hz，f_s 为语音信号采样频率)的 FS 系数，ϕ 定义为 m 在 $P(n)$ 上归一到 2π 时的相位值，$s(n,\phi)$ 则为一个与时刻 n 和相位值 ϕ 有关的 2 维函数。

传统 CWI 算法将 $A_k(n)$ 和 $B_k(n)$ 通过一组低通滤波器后可简单的分离出其 SEW 和 REW 第 k 次谐波对应的 FS 系数，分别表示为 $A_k^{\mathrm{SEW}}(n)$，$B_k^{\mathrm{SEW}}(n)$ 和 $A_k^{\mathrm{REW}}(n)$，$B_k^{\mathrm{REW}}(n)$[2]，并将其各自转化为幅度谱和相位谱的形式，定义为

$$\left.\begin{aligned}F_k^{\mathrm{SEW}}(n)&=\sqrt{\left(A_k^{\mathrm{SEW}}(n)\right)^2+\left(B_k^{\mathrm{SEW}}(n)\right)^2}\\ \varphi_k^{\mathrm{SEW}}(n)&=\operatorname{arctg}\left(\frac{B_k^{\mathrm{SEW}}(n)}{A_k^{\mathrm{SEW}}(n)}\right)\end{aligned}\right\} \tag{3}$$

$$\left.\begin{aligned}F_k^{\mathrm{REW}}(n)&=\sqrt{\left(A_k^{\mathrm{REW}}(n)\right)^2+\left(B_k^{\mathrm{REW}}(n)\right)^2}\\ \varphi_k^{\mathrm{REW}}(n)&=\operatorname{arctg}\left(\frac{B_k^{\mathrm{REW}}(n)}{A_k^{\mathrm{REW}}(n)}\right)\end{aligned}\right\} \tag{4}$$

式(3)，式(4)中，$F_k^{\mathrm{SEW}}(n)$，$F_k^{\mathrm{REW}}(n)$ 分别定义为时刻 n 时 SEW 和 REW 第 k 次谐波的幅度，$\varphi_k^{\mathrm{SEW}}(n)$，$\varphi_k^{\mathrm{REW}}(n)$ 分别定义为时刻 n 时SEW和 REW 第 k 次谐波的相位。由于人耳对于相位谱不敏感，因此编码传输中往往只对 SEW 和 REW 的幅度谱进行量化，而对各自的相位谱不编码，或在码率允许的情况下仅以少量比特编码[8,9]。

由于语音信号低频谐波能量一般较大且在短时内较稳定，因此 k 较小时的谐波 FS 系数 $A_k(n)$ 和 $B_k(n)$ 经低通滤波器后其大部分数值将包含在 SEW 的 $A_k^{\mathrm{SEW}}(n)$ 和 $B_k^{\mathrm{SEW}}(n)$ 中，在式(3)中则表现为低频谐波信号成分主要集中在 $F_k^{\mathrm{SEW}}(n)$ 中，而在 $F_k^{\mathrm{REW}}(n)$ 中的信号成分较小。同理，对于语音信号的高频谐波，由于其常趋向于类噪声信号，即当 k 较大时在时间轴上相邻特征波的 $A_k(n)$ 和 $B_k(n)$ 变化较大，因此经滤波分离后，其高频谐波信号成分会更多的集中在 $F_k^{\mathrm{REW}}(n)$ 中，而在 $F_k^{\mathrm{SEW}}(n)$ 中的信号成分较小。

2.2 特征波形的重构和残差信号的恢复

在 CWI 解码端，首先对接收到的 SEW，REW 的幅度谱进行解码恢复，但是由于缺少对应的相位谱信息而导致特征波的不能恢复。因此，对于 SEW 和 REW 相位谱的估计成为了重构中不可缺少的环节。一般认为，由于 REW 类噪声，其相位谱可以随机噪声代替，而 SEW 的周期性较强，因此其相位谱用固定相位代替，从而得以顺利恢复出特征波形的 FS 系数 $\hat{A}_k(n)$ 和 $\hat{B}_k(n)$，表示为

$$\begin{aligned}\hat{A}_k(n)&=\hat{A}_k^{\mathrm{SEW}}(n)+\hat{A}_k^{\mathrm{REW}}(n)=F_k^{\mathrm{SEW}}(n)\\&\quad\cdot\cos\hat{\varphi}_k^{\mathrm{SEW}}(n)+F_k^{\mathrm{REW}}(n)\cos\hat{\varphi}_k^{\mathrm{REW}}(n)\end{aligned} \tag{5}$$

$$\begin{aligned}\hat{B}_k(n)&=\hat{B}_k^{\mathrm{SEW}}(n)+\hat{B}_k^{\mathrm{REW}}(n)=F_k^{\mathrm{SEW}}(n)\\&\quad\cdot\sin\hat{\varphi}_k^{\mathrm{SEW}}(n)+F_k^{\mathrm{REW}}(n)\sin\hat{\varphi}_k^{\mathrm{REW}}(n)\end{aligned} \tag{6}$$

式(5)，式(6)中，$\hat{A}_k^{\mathrm{SEW}}(n)$，$\hat{B}_k^{\mathrm{SEW}}(n)$ 和 $\hat{A}_k^{\mathrm{REW}}(n)$，$\hat{B}_k^{\mathrm{REW}}(n)$ 分别为恢复出的SEW和REW的FS系数，$\hat{\varphi}_k^{\mathrm{SEW}}(n)$，$\hat{\varphi}_k^{\mathrm{REW}}(n)$ 分别为 SEW 和 REW 的相位谱估计。此外，相邻特征波形之间还要进行对齐操作，以期望增强相邻特征波的相关性，从而更好地恢复出语音浊音段信号。若 $P(n)$ 和 $P(n+1)$ 相等，值都为 P 时，对齐操作定义为

$$\begin{aligned}T=\arg\max_{0\le T'<P}\sum_{k=1}^{\lfloor P/2\rfloor}\Big\{&\left[\hat{A}_k(n)\hat{A}_k(n+1)+\hat{B}_k(n)\hat{B}_k(n+1)\right]\\&\cdot\cos\left(\frac{2\pi kT'}{P}\right)+\left[\hat{B}_k(n)\hat{A}_k(n+1)-\hat{A}_k(n)\hat{B}_k(n+1)\right]\\&\cdot\sin\left(\frac{2\pi kT'}{P}\right)\Big\}\end{aligned} \tag{7}$$

$$\left.\begin{aligned}\widetilde{A}_k(n+1)&=\widehat{A}_k(n+1)\cos\left(\frac{2\pi kT}{P}\right)-\widehat{B}_k(n+1)\\&\quad\cdot\sin\left(\frac{2\pi kT}{P}\right)\\\widetilde{B}_k(n+1)&=\widehat{A}_k(n+1)\sin\left(\frac{2\pi kT}{P}\right)+\widehat{B}_k(n+1)\\&\quad\cdot\cos\left(\frac{2\pi kT}{P}\right)\end{aligned}\right\}\tag{8}$$

式(7)中，T 为对齐时需要偏移的样点数。式(8)中 $\widetilde{A}_k(n+1)$，$\widetilde{B}_k(n+1)$ 为时刻 $n+1$ 时对齐后的 FS 系数。当 $P(n)$ 和 $P(n+1)$ 不相等时，则首先需要进行 FS 系数的补齐操作。抽样时刻 n 时恢复出的特征波形则可表示为

$$\tilde{s}(n,\phi)=\sum_{k=1}^{\lfloor P(n)/2\rfloor}\left[\widetilde{A}_k(n)\cos(k\phi)+\widetilde{B}_k(n)\sin(k\phi)\right]\tag{9}$$

任意时刻 n_1 时残差信号的恢复则首先对抽样时刻特征波形进行内插，依靠产生的瞬时特征波形和其瞬时相位值来决定，表示为

$$\begin{aligned}\tilde{r}(n_1)&=\overline{s}(n_1,\phi_1)\\&=\sum_{k=1}^{\lfloor P(n_1)/2\rfloor}\left[\overline{A}_k(n_1)\cos(k\phi_1)+\overline{B}_k(n_1)\sin(k\phi_1)\right]\end{aligned}\tag{10}$$

式(10)中，$\tilde{r}(n_1)$ 为时刻 n_1 时恢复出的残差信号，$\overline{A}_k(n_1)$，$\overline{B}_k(n_1)$ 为时刻 n_1 时内插产生的瞬时特征波形 $\overline{s}(n_1,\phi)$ 的 FS 系数，ϕ_1 表示当前时刻 n_1 的瞬时相位值。更多 CWI 算法细节参见文献[2]。

3　基于多带 2.4 kbit/s 的 CWI 算法

传统 CWI 算法编码时 SEW 和 REW 幅度谱的量化误差和相位估计的不确定性，使得恢复出的特征波形相关性减小，因此在解码端恢复时，通过对齐方法增强相邻特征波的相关性。这种对齐方法从一定程度上可以理解为将特征波所有谐波的初始相位统一做出重新调整，使相邻波形更加匹配。由于特征波低频谐波分量能量一般较大，在谐波总能量中占主导地位，且低频信号的周期较长，因此整体对齐的方法对低频谐波分量的匹配非常有效。但由于高频谐波分量能量较小且信号周期较短，这种整体对齐对高频谐波分量的匹配却往往不理想，有时对齐后甚至更加不匹配，从而造成高频谐波分量的噪声化。此外，对于快慢波来说，由于高频谐波分量在 REW 中能量较大，而 REW 往往赋以随机相位谱，也导致了相邻特征波形高频分量相关性的减小，最终使得高频谐波分量噪声化。

3.1 多带清浊音标志

与 2.4 kbit/s 的 MELP 算法[6]相似，本文首先将语音划分为 4 个频带：0–1 kHz，1–2 kHz，2–3 kHz 和 3–4 kHz，分别计算语音在各带内的语音强度，并根据语音强度设置多带清浊音标志。语音强度判断在每个语音帧最后一个样点处进行。定义在某个语音帧最后一个样点附近的语音信号序列为 $\{\cdots,s_{-2},s_{-1},s_0,s_1,s_2,\cdots\}$，其中 s_0 为该帧的最后一个样点。对该信号序列分别进行带通频率为 0–1 kHz，1–2 kHz，2–3 kHz 和 3–4 kHz 的带通滤波器滤波得 4 组带通信号序列，各自对应信号序列为 $s^i=\{\cdots,s^i_{-2},s^i_{-1},s^i_0,s^i_1,s^i_2,\cdots\}$，$i=1,2,3,4$。对频带 1–2 kHz，2–3 kHz 和 3–4 kHz 对应的 s^i 还需进行 150 Hz 的低通滤波，得到带通信号 s^i 的包络信号序列 $g^i=\{\cdots,g^i_{-2},g^i_{-1},g^i_0,g^i_1,g^i_2,\cdots\}$，$i=2,3,4$。

对于频带 0~1 kHz 的语音强度定义为

$$Vbp_1=\frac{\sum_{k=-\lfloor\tau/2\rfloor-80}^{-\lfloor\tau/2\rfloor+79}s^1_k s^1_{k+\tau}}{\sqrt{\sum_{k=-\lfloor\tau/2\rfloor-80}^{-\lfloor\tau/2\rfloor+79}\left(s^1_k\right)^2\sum_{k=-\lfloor\tau/2\rfloor-80}^{-\lfloor\tau/2\rfloor+79}\left(s^1_{k+\tau}\right)^2}}\tag{11}$$

对于频带 1–2 kHz，2–3 kHz 和 3–4 kHz 的语音强度定义为

$$Vbp_i=\max\left(\frac{\sum_{k=-\lfloor\tau/2\rfloor-80}^{-\lfloor\tau/2\rfloor+79}s^i_k s^i_{k+\tau}}{\sqrt{\sum_{k=-\lfloor\tau/2\rfloor-80}^{-\lfloor\tau/2\rfloor+79}\left(s^i_k\right)^2\sum_{k=-\lfloor\tau/2\rfloor-80}^{-\lfloor\tau/2\rfloor+79}\left(s^i_{k+\tau}\right)^2}},\ 0.1+\frac{\sum_{k=-\lfloor\tau/2\rfloor-80}^{-\lfloor\tau/2\rfloor+79}g^i_k g^i_{k+\tau}}{\sqrt{\sum_{k=-\lfloor\tau/2\rfloor-80}^{-\lfloor\tau/2\rfloor+79}\left(g^i_k\right)^2\sum_{k=-\lfloor\tau/2\rfloor-80}^{-\lfloor\tau/2\rfloor+79}\left(g^i_{k+\tau}\right)^2}}\right)\quad i=2,3,4\tag{12}$$

式(11)，式(12)中，τ 为该帧的基音长度。

当带内语音强度大于某阈值时，则认为该带为浊音带，对应该带的清浊音标志置 1，反之则置 0。需要注意的是，当 0–1 kHz 频带对应的标志为 0 时，则全带都强制置 0；当 1–2 kHz 和 2–3 kHz 频带对应的标志都为 0 时，3–4 kHz 频带对应的标志强制置 0。多带清浊音标志的设置能够很好地指示各频带语音的性质，从而为恢复语音高频谐波提供辅助信息。更多多带清浊音标志设置细节参见文献[10]。

3.2 快慢波相位的恢复

由于恢复出的 SEW 和 REW 的 FS 系数 $\widehat{A}_k^{\mathrm{SEW}}(n)$，$\widehat{B}_k^{\mathrm{SEW}}(n)$ 和 $\widehat{A}_k^{\mathrm{REW}}(n)$，$\widehat{B}_k^{\mathrm{REW}}(n)$ 包含了第 k 次谐波的信息，其对应的谐波频率为 $kf_s/P(n)$ Hz。而多带清浊音标志指示了语音各频带的清浊音性质，因此本文规定当快慢波的第 k 次谐波频率在浊音带内时，其相位谱都用固定相位谱估计；反之，在清音带内时，其相位谱都用随机相位谱估计。

3.3 特征波重构的对齐

考虑到整体对齐对于语音高频谐波恢复的无效性，本文采取分条件的部分对齐方法。当多带清浊音标志全为 0 时，默认为全带清音，相关性较弱，不对齐；当频带 0–1 kHz 标志为 1 时，则搜索频带 1–2 kHz 的清浊音标志，如果搜索到其带内清浊音标志为 0 时则终止搜索。对齐时仅将搜索到的浊音频带内的谐波进行对齐，而对于 2 kHz 以上的高频谐波则始终不参加对齐过程。部分对齐定义为

$$
\begin{aligned}
T = \mathop{\arg\max}_{0\le T'<P} \sum_k \Big\{ & \Big[\widehat{A}_k(n)\widehat{A}_k(n+1) + \widehat{B}_k(n)\widehat{B}_k(n+1)\Big] \\
& \cdot \cos\left(\frac{2\pi k T'}{P}\right) + \Big[\widehat{B}_k(n)\widehat{A}_k(n+1) \\
& - \widehat{A}_k(n)\widehat{B}_k(n+1)\Big]\sin\left(\frac{2\pi k T'}{P}\right)\Big\}, \\
& k \in \left\{ x \mid 0 < \frac{xf_s}{P} \le 2000 \text{且} \frac{xf_s}{P} \text{在浊音带内} \right\}
\end{aligned} \tag{13}
$$

$$
\left.
\begin{cases}
\widetilde{A}_k(n+1) = \widehat{A}_k(n+1)\cos\left(\dfrac{2\pi k T}{P}\right) - \widehat{B}_k(n+1)\sin\left(\dfrac{2\pi k T}{P}\right) \\
\widetilde{B}_k(n+1) = \widehat{A}_k(n+1)\sin\left(\dfrac{2\pi k T}{P}\right) + \widehat{B}_k(n+1)\cos\left(\dfrac{2\pi k T}{P}\right)
\end{cases}, \quad k \in \left\{ x \mid 0 < \frac{xf_s}{P} \le 2000 \text{且} \frac{xf_s}{P} \text{在浊音带内} \right\} \right.
$$

$$
\left.
\begin{cases}
\widetilde{A}_k(n+1) = \widehat{A}_k(n+1) \\
\widetilde{B}_k(n+1) = \widehat{B}_k(n+1)
\end{cases}, \quad k\text{为其它} \right\} \tag{14}
$$

由式(13)，式(14)可见，由于本文采用部分对齐方法，并非所有谐波分量都参与对齐，因此新算法在合成端特征波对齐时的计算复杂度也可大为降低。

3.4 编码参数量化和编解码方案

本文基于多带2.4 kbit/s 的特征波形内插编码器相关参数的比特分配如表1所示。输入窄带语音分析帧长为25 ms，波形提取速率为320 Hz(即每帧提取8个特征波形)，线谱频率(LSF)参数采用22 bit的分裂矢量量化；基音周期采用7 bit标量量化；功率增益采用MELP算法[10]中的8 bit对数标量量化方法。SEW和REW的采样速率都降为40 Hz。量化时，SEW幅度谱的低12维分量进行8 bit矢量量化，其余分量采用5 bit的变维矢量量化；REW幅度谱采用7 bit变维矢量量化[2]。对于SEW和REW的相位谱，固定相位谱设为恒定0相位，随机相位谱则由$[0,2\pi]$上均匀分布的随机数生成。

多带 2.4 kbit/s CWI 算法的整体编解码方案如图 1 所示。

4 实验结果与分析

本文选取语音均来自TIMIT语音库，输入信号采样率调整为 8 kHz，并以传统 2.4 kbit/s 的 CWI 算法[3]和 2.4 kbit/s 的标准 MELP 算法[10]作为比较依据。需要说明的是，传统 2.4 kbit/s 的 CWI 算法的参数设置与原文献略有不同，帧长设为 25 ms，REW 的更新速率为 80 Hz，其它参数更新速率为 40 Hz。比特分配时，则仅对 SEW 幅度谱的低 12 维分量进行 9 bit 矢量量化，而高维分量仍采用文献[3]中与 REW 幅度谱互补的恢复方案，其它参数比特数分配则同表 1 中的对应参数。

图 2，图 3 分别为一段男声和女声语音的语谱图。图 2，图 3 中各子图(a)，(b)，(c)，(d)分别为原始语音以及传统 CWI 算法，MELP 算法，本文算法合成语音的语谱图。图中可见，传统 CWI 算法对男女声语音高频段的谐波结构造成了丢失，而本文算法在浊音段较好地保持了高频部分的谐波结构。此外，MELP 算法合成语音也能较好的保持高频谐波结构，并可见其高频部分的谐波特征强于本文算法。在总体听觉感受上，传统 2.4 kbit/s 的 CWI 算法合成语音有较为明显的噪声感，而后两种编码方法的合成语音则更能令人接受。

在对多带CWI与传统CWI算法和MELP算法进行主观R-A/B测试时，选取10名男女各半的听音人组成测试小组，测试语音材料则为24段男女语音各半的TIMIT语音。R是原始语音仅作参考，A和B为需要比较的编解码合成语音，听音人对A，B的音质做出偏爱选择，最后统计测听结果，得到主观评价表。

表2为本文基于多带2.4 kbit/s的CWI算法与传统2.4 kbit/s的CWI算法的主观R-A/B测试结果。其

表1 编码参数量化比特分配

参数	每帧比特(bit)	更新速率(Hz)
LSF	22(7+7+8)	40
基音周期	7	40
功率增益	8	40
多带清浊音标志	3	40
SEW	13(8+5)	40
REW	7	40
总比特	60	40

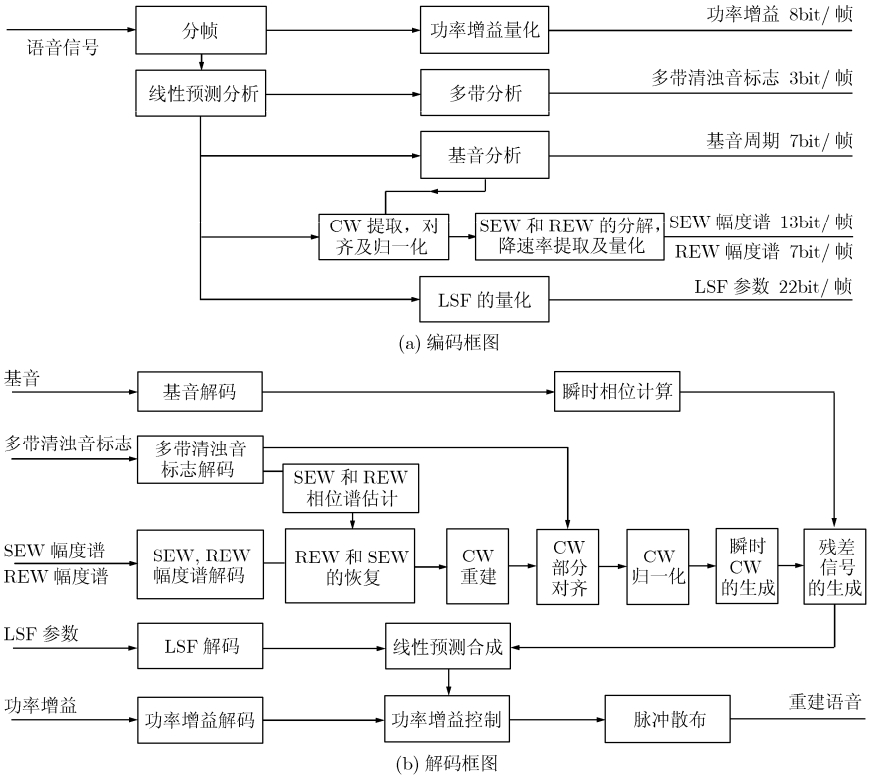

图1 基于多带2.4 kbit/s的CWI算法编解码框图

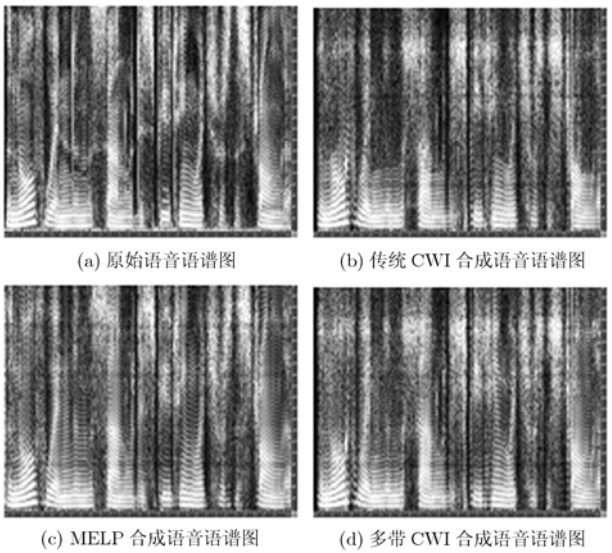

图2 男声语音语谱图

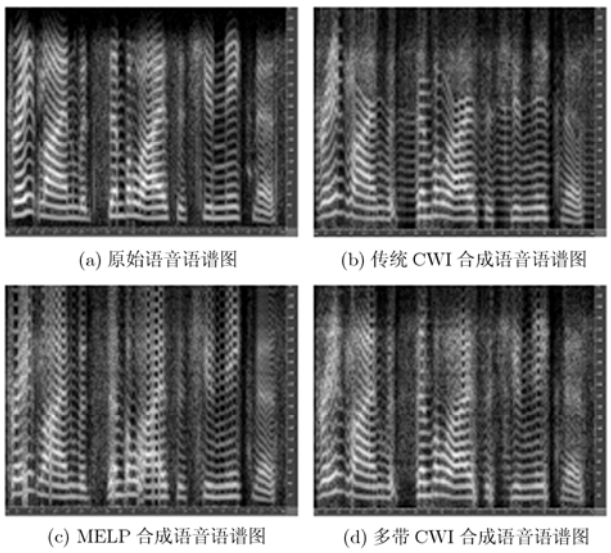

图3　女声语音语谱图

中，A代表多带CWI算法，B代表传统CWI算法。从表2可见，多带CWI算法的合成语音质量43.3%的偏爱大大优于传统CWI算法24.6%的偏爱程度。特别在男声语音的听觉效果上，由于多带CWI算法能够表现语音高频谐波特征，因此更以54.2%的优势超过传统算法22.5%的偏爱。

表3为本文基于多带2.4 kbit/s的CWI算法与2.4 kbit/s的MELP算法的主观R-A/B测试结果。其中，A代表多带CWI算法，B代表MELP算法。从表3可见，多带CWI算法的听觉质量略优于MELP算法，在女声语音时35.8%的偏爱度优于MELP算法的

表2 主观R-A/B测试结果(%)

	偏爱 A	无偏爱	偏爱 B
女声语音	32.5	40.8	26.7
男声语音	54.2	23.3	22.5
所有语音	43.3	32.1	24.6

表3 主观R-A/B测试结果(%)

	偏爱 A	无偏爱	偏爱 B
女声语音	35.8	40.0	24.2
男声语音	26.7	44.2	29.2
所有语音	31.3	42.1	26.7

24.2%，而在男声语音时略差，偏爱度为26.7%，而MELP算法为29.2%。这是由于MELP算法为了在低速率情况下保证语音可懂度，更侧重于加强语音的谐波结构。MELP算法中最多只对前10个谐波分量的归一化能量进行编码，而对其后所有高频谐波的能量都用前10个谐波分量的平均能量来代替。这样的高频能量的固定设置是相对较高的。由于男声谐波数目较多，在语音总能量一定的情况下，单个高频谐波所分配的能量变小，所以其高频谐波能量一般仅略强于真实语音的高频谐波能量。而多带CWI算法的高频谐波能量是通过SEW和REW来恢复的，是对真实语音高频谐波能量的编解码，其谐波能量并没有得到加强。因此这种增强了语音谐波结构的MELP算法合成出的男声语音更加浑厚有力，效果优于多带CWI算法。但对于女声语音，由于其谐波数目较少，因此单个高频谐波所分配的能量较多，MELP算法容易造成合成语音的第10个谐波以后的高频谐波能量大大强于真实语音谐波能量的情况。过强的谐波结构使得MELP算法合成出来的女声语音有时会有一定的机械感和沉闷感。而多带CWI算法则不存在这种情况，合成的语音更加自然。

5 结束语

本文提出了一种基于多带 2.4 kbit/s 的 CWI 算法。由于其引入多带清浊音标志，在 SEW 和 REW 的相位估计上采用新的估计方法，同时在合成端特征波采用部分对齐手段，使得语音的高频谐波分量得以较好的恢复。该算法与传统 CWI 算法相比，合成语音质量明显提高，语音更为清晰，噪声感较小，并且由于特征波只采用部分对齐，使得合成端的计算复杂度大为下降；与 2.4 kbit/s 的 MELP 算法相比，合成语音质量亦略显优势，减小了语音的机械感和沉闷感。但是该 CWI 算法的计算复杂度仍然较高，如何进一步减少计算复杂度，以及在噪声环境，丢帧等情况下的处理，也都是需要考虑的问题。

参 考 文 献

[1] Kleijn W B. A speech coder based on decomposition of characteristic waveforms[C]. IEEE ICASSP'95, Detroit, 1995: 508–511.

[2] 鲍长春. 数字语音编码原理[M]. 西安: 西安电子科技大学出版社, 2007, 第9章.

[3] Kleijn W B, Shoham Y, and Sen D, *et al.*. A low-complexity waveform interpolation coder[C]. IEEE ICASSP'96, Atlanta, 1996: 212–215.

[4] 王贵平, 鲍长春, 张鹏. 基于奇异值分解的低速率波形内插语音编码算法[J]. 电子学报, 2006, 34(1): 135–140.
Wang Gui-ping, Bao Chang-chun, and Zhang Peng. Low bit rates waveform interpolation speech coding based on singular value decomposition[J]. *Acta Electronica Sinica*, 2006, 34(1): 135–140.

[5] 王晶，匡镜明，谢湘. 基于小波变换的2.4 kbit/s 波形内插语音编码算法[J]. 通信学报, 2007, 28(5): 43–48.
Wang Jing, Kuang Jing-ming, and Xie Xiang. Waveform interpolation speech coding algorithm at 2.4 kbit/s based on wavelet transform[J]. *Journal on Communications*, 2007, 28(5): 43–48.

[6] Supplee L M, Cohn R P, and Collura J S, *et al.*. MELP: the new Federal Standard at 2400 bps[C]. IEEE ICASSP'97, Munich, 1997: 1591–1594.

[7] McCree A V and Barnwell T P. A mixed excitation LPC vocoder model for low bit rate speech coding[J]. *IEEE Transactions on Speech and Audio Processing*, 1995, 3(4): 242–250.

[8] Gottesman O and Gersho A. Enhanced waveform interpolative coding at 4 kbps[C]. IEEE Workshop on Speech Coding Proceedings, Haikko Manor Porvoo, 1999: 90–92.

[9] 陈悦, 鲍长春. 一种用于WI 语音编码的相位预测式矢量量化方法[J]. 电子与信息学报, 2007, 29(11): 2672–2675.
Chen Yue and Bao Chang-chun. A predictive phase vector quantization method in WI speech coding[J]. *Journal of Electronics & Information Technology*, 2007, 29(11): 2672–2675.

[10] Federal Information Processing Standards Publication. Specifications for the analog to digital conversion of voice by 2400 bit/second mixed excitation linear prediction[S], 1998.

汤一彬： 男，1982 年生，博士生，研究方向为语音信号处理.

黄 蓉： 女，1984 年生，硕士生，研究方向为语音信号处理.

吴镇扬： 男，1949 年生，教授，博士生导师，研究方向为听觉与视觉信号处理、多媒体信号处理.