

## 2. 4kb/s 多带激励语音编码算法研究

杨明 李秋云 邱锋海 莫福源

**摘要** 本文对多带激励语音编码算法进行了探讨。文中描述了多带激励的语音模型, 并重点介绍了其具体实现中的几项关键技术, 为提高低速率语音编码合成语音质量作了一些有益的研究, 在 2.4kb/s 的速率下获得了令人满意的通信语音质量。

**关键词** 语音编码 多带激励 基音周期

### 引言

在数字通信中, 语音信号直接数字化所需的数码率太高, 为了提高传输和存储的效率, 充分利用信道容量, 必须对数字语音信号进行压缩编码。通过降低编码速率, 可以使同样的信道容量能传输更多路的语音信号。在传输比特率限制十分严格的场合, 低速率语音编码具有特别重要的意义。由于现有的语音编码国际标准传输速率较高, 低速率编码方案的语音质量又大多不能令人满意, 因此低速率语音编码技术研究近来得到了广泛的重视。

本文论述的编码速率为 2.4kb/s 的语音编解码模拟系统采用的算法核心是基于改进的多带激励模型 (Multi-Band Excitation, 简称 MBE)。从算法角度, 该系统分为语音分析、模型参数量化编解码和语音合成三大部分。

基音提取在语音分析中是最重要的部分, 对 MBE 整体系统质量起着决定性的影响。本文采用了一种有效的基音提取和平滑算法, 准确度较高, 并注重消除了倍频干扰, 具有良好的基音连续性, 提高了合成语音的自然度。

MBE 系统要应用于低速率编码, 必须对参数量化有高效的方法。由于 MBE 中谱幅度参数较多, 标量量化效率太低, 本文采用线性预测 (Linear Prediction) 技术, 用全极点模型拟合 MBE 谱包络, 将线性预测系数转换成线谱频率 (Line Spectrum Frequency) 后再采用矢量量化的方法提高量化效率; 为保证合成语音的质量, 还采用了一种共振峰增强 (Formant Enhancement) 技术, 补偿全极点模型带来的误差。采用以上的方法, 在传输速率降到 2.4kb/s 情况下, 获得了满意的合成语音。

### 1. 多带激励语音模型

本文中语音编码的对象是带宽为 200~3400Hz 的电话带宽语音, 语音一般限带到 4KHz 并以 8KHz 采样, 其基本的编码技术为脉冲编码调制 PCM。

传统的声码器 (如 LPC (Linear Prediction Coding) 声码器等) 通常按清浊音的不同使用两元化激励信号: 完全由基音周期决定的周期信号或随机信号。从频域上看就是谐波谱和噪声谱两种。而在 MBE 模型中, 激励谱是谐波谱和噪声谱的混合。具体来说, 就是把语音频谱划分为多个频带, 对每个带进行二元清浊判断,

然后对不同的带采用相应的激励信号，最后将各带合成信号叠加，形成全带合成信号。正是由于频谱分带分析合成，故称多带激励。

### 1.1 MBE 模型的参数提取

MBE 模型需要估计的参数包括：基音频率、谱包络和各带的清浊音判决信息。与 LPC 模型不同，MBE 同时估计激励和谱包络参数。参数提取采用类似合成分析 (Analysis by Synthesis) 的方法，误差函数取为：

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ |S_w(\omega)| - |\hat{S}_w(\omega)| \right]^2 d\omega \quad (1)$$

或

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} |S_w(\omega) - \hat{S}_w(\omega)|^2 d\omega \quad (2)$$

(2) 式与 (1) 式的不同之处在于 (2) 式中包含相位信息。考虑到编码速率较低，实际采用 (1) 式，只考虑谱分量的模值。

提取参数时首先依次假设基音频率  $\omega_0$  (设基音周期为  $P$ ，则  $\omega_0 = 2\pi/P$ ) 为各种可能出现的值。对于每一个  $\omega_0$ ，将  $\omega = -\pi \sim \pi$  分为若干频带，每个频带的频率下限和上限依次是  $a_m$  和  $b_m$ ，同时认为谱包络在每个基音谐波范围内为定值，设第  $m$  个频带的谱包络为  $A_m$ ，则第  $m$  个频带的误差为：

$$\varepsilon = \frac{1}{2\pi} \int_{a_m}^{b_m} \left[ |S_w(\omega)| - |A_m| \cdot |E_w(\omega)| \right]^2 d\omega \quad (3)$$

由 (3) 式知，要求得一组最佳的  $A_m$  使得  $\varepsilon = \sum_{-M}^M \varepsilon_m$  最小，只需分别对每一频带求得相应的  $A_m$  使得  $\varepsilon_m$  最小。此最佳值由  $\varepsilon_m$  对  $|A_m|$  的偏导值为零求得如下：

$$|A_m| = \frac{\int_{a_m}^{b_m} |S_w(\omega)| \cdot |E_w(\omega)| d\omega}{\int_{a_m}^{b_m} |E_w(\omega)|^2 d\omega} \quad (4)$$

对浊音谐波段，(4) 式中激励谱由周期谱代入；对清音谐波段，由噪声谱代入。这样即可得到 MBE 模型的谱包络信息，即谱包络矢量  $\{A_m\}$ 。

基音频率估计时，假设语音为浊音，用周期谱代入求出各次谐波带内的最小误差  $\varepsilon_{m0}$ ，则整个频带内的最小误差可以由它们累加得到：

$$\varepsilon_0 = \sum_{m=-M}^M \varepsilon_{m0} \quad (5)$$

对于每个假设的基音频率都可以求得一个  $\varepsilon_0$ ，在真正的基音周期或它的整数倍周期上， $\varepsilon_0$  将取极小值，由此可估计基音周期，这就是 MBE 基音估计的频域计算方法。

实际计算中，考虑到计算量，先用时域方法求出基音周期的粗估值。时域基音周期估计误差函数计算公式如下：

$$\varepsilon_{uB} = \frac{\sum_{n=-\infty}^{\infty} s^2(n)w^2(n) - P \cdot \sum_{k=-\infty}^{\infty} \phi(kP)}{\left[1 - P \cdot \sum_{n=-\infty}^{\infty} w^4(n)\right] \cdot \left[\sum_{n=-\infty}^{\infty} s^2(n)w^2(n)\right]} \quad (6)$$

其中  $\phi(m)$  为  $s(n)w^2(n)$  的自相关函数:

$$\phi(m) = \sum_{n=-\infty}^{\infty} s(n)w^2(n)s(n-m)w^2(n-m) \quad (7)$$

基音周期的粗估值还需要经过基音平滑, 使得相邻语音帧的基音没有大的跳变, 并消除基音周期整数倍的影响。得到粗估值后, 再用频域方法在粗估值附近进行细搜索, 得到基音频率的最终估值。

清浊音判决取决于周期谱对语音谱的匹配程度。将归一化拟合误差  $\xi_m$  与一定的阈值相比较,

$$\xi_m = \frac{\varepsilon_m}{\frac{1}{2\pi} \int_{a_m}^{b_m} |S_w(\omega)|^2 d\omega} \quad (8)$$

$\xi_m$  小于阈值则可以判该谐波频带为浊音区, 反之为清音区。

## 1. 2 MBE 模型的语音合成

为了保证基音频率在帧间平滑过渡, MBE 模型对浊音带语音采用时域法合成; 同时, 由于在频域中比较容易实现带通滤波, 对清音带语音采用频域法合成。浊音带语音由一组正弦波叠加合成:

$$s_v(t) = \sum_m A_m(t) \cos(\theta_m(t)) \quad (9)$$

幅度函数  $A_m(t)$  在帧间进行线性插值。浊音带合成时, 清音带幅度取为零。相位信号由初始相位  $\phi_0$  和频率跟踪函数  $\omega_m(t)$  决定:

$$\theta_m(t) = \int_0^t \omega_m(\xi) d\xi + \phi_0 \quad (10)$$

其中  $\omega_m(t)$  是由相邻帧的第  $m$  个谐波的相位相互插值得出:

$$\omega_m(t) = m\omega_0(0) \frac{(N-t)}{N} + m\omega_0(N) \frac{t}{N} + \Delta\omega_m \quad (11)$$

其中  $N$  为帧长。选取适当的  $\phi_0$  和  $\Delta\omega_m$ , 以保证帧间相位连续。

当相邻帧基音周期变化较大时, 若仍强行插值, 会造成合成语音不自然。这时对前后两帧分别进行合成, 然后加窗叠接。

清音带语音的合成采用如下方法: 先将加窗的噪音信号进行傅里叶变换, 其中对应浊音段的频带幅度置为零, 对应清音段的平均幅度由  $|A_m|$  得到, 然后在进行傅里叶反变换即可得到清音带的合成语音。

最后将浊音和清音合成语音叠加起来, 就是 MBE 模型的合成语音信号。

## 2. 几项关键技术

以上介绍了 MBE 模型的核心算法，将其具体实现还会遇到很多细节上的问题。参数量化方法虽然对编码速率影响很大，由于篇幅所限，将另文进行讨论。这里只介绍作者对几项较重要的技术研究。

### 2.1 有限窗长对应的时域基音粗估误差函数

实际应用中，窗函数  $\omega(n)$  具有一定宽度，不妨设其为  $(2N+1)$ ，并围绕原点对称。在窗长范围内有  $L$  个假设基音周期，其中：

$$L = \left\lfloor \frac{2N+1}{P} \right\rfloor \quad (12)$$

符号  $\lfloor x \rfloor$  表示小于或等于  $x$  的最大整数。(6) 式的求和上下限应作相应变动，得到本文系统中采用的误差函数公式：

$$\varepsilon_{uB} = \frac{\sum_{n=-N}^N s^2(n)w^2(n) - P \cdot \sum_{k=-L}^L \phi(kP)}{\left[ 1 - P \cdot \sum_{n=-N}^N w^4(n) \right] \cdot \left[ \sum_{n=-N}^N s^2(n)w^2(n) \right]} \quad (13)$$

### 2.2 基音平滑以及消除倍频干扰

由于基音周期对应的误差函数值一定为由各假设值求出的  $\varepsilon_{uB}$  中的某个局部最小值，故提取基音的问题转化为从若干个局部最小值中选取最合理的一个(通常从前五个局部最小峰值中选)。为了消除倍频干扰，考察一下前十个局部最小值，若发现前几个局部最小值对应的周期都是后面某个局部最小值对应的周期的整数倍，应将后面的这一局部最小值位置适当提前。经过重新排序后，系统再对前五个局部最小峰值进行筛选。筛选的方法是在连续四帧选出的峰值中，每帧各取一个结点，连成一条路径，为路径的边选取合适的权重，这样就将基音平滑问题简化为求带权无环图的最短路径问题。采用著名的动态规划算法可以有效地对之求解。在结点权与边权的比例选择问题上，由于算法的原因，结点权越小，对应于准确基音周期的可能性越大；边权重越小，基音平滑性越好。在实际应用中，因为人耳对相邻帧出现的基音跳变比较敏感，判断基音周期时应更侧重于基音的平滑性。

### 2.3 浊音段合成语音的相位衔接

上文提到，浊音段合成时，应选取适当的  $\phi_0$  和  $\Delta \omega_m$ ，以保证帧间相位连续。具体来说， $\phi_0$  的选择应使本帧的  $\theta_m(0)$  (帧起始点相位) 的主值等于上帧  $\theta_m(N)$  (帧终止点相位) 的主值。对应 (10) 式的离散表达为：

$$\theta_m(n) = \sum_{l=0}^{n-1} \omega_m(l,0) + \phi_0 \quad (14)$$

频率跟踪函数取

$$\omega_m(n,0) = m\omega_0(-1)\frac{N-n}{N} + m\omega_0(0)\frac{n}{N}, 0 \leq n < N \quad (15)$$

将(15)式代入(14)式, 推导可得

$$\theta_m(N) = \phi_0 + \frac{N-1}{2}m\omega_0(0) + \frac{N+1}{2}m\omega_0(-1) \quad (16)$$

将  $\theta_m(N)$  作为下一帧的  $\phi_0$ , 即可保证帧间的相位衔接。

### 3. 结束语

MBE 作为正弦模型的一种, 在语音生成模型和模型参数提取方面都有其独到之处。应用于语音编码中, 由于辅音清晰度的改善, 合成语音在速率为 2.4kb/s 的模拟系统中得到了令人完全可以接受的通信质量。

#### 参考文献

- 1 D.W.Griffin and J.S.Lim, "Multi-band excitation vocoder", IEEE Trans. On ASSP, vol. 36, No. 8, Aug. 1988, pp. 1223-1235.
- 2 杨行峻 迟惠生, 《语音信号数字处理》, 北京, 电子工业出版社, 1995。
- 3 王田, "低速率 1.2kb/s~2.4kb/s 语音编码算法的研究", 清华大学博士论文, 1996。