

# 一种 2.4kbps 改进型 MBELP 编码<sup>1</sup>

钟 明 江太辉

(五邑大学信息科学研究所 广东江门 529020)

**摘 要** 该文给出了一种改进的 2.4kb/s 多带激励线性预测 (IMBELP) 语音编码算法。与传统的 MBELP 算法相比, 本算法在基音提取和清 / 浊音判决上采取了一些改进措施, 使得合成语音质量有一定的提高。本文详细介绍了改进后的 MBELP 算法, 并将其在基音提取和清 / 浊音判决的结果与传统的 MBELP 进行比较。

**关键词** 多带激励编码, 多带激励线性预测编码, 基音提取, 清 / 浊音判决

**中图分类号** TN912.3

## 1 引 言

近些年, 为了实现高质量、低码率的语音合成信号, 人们一直在研究 2.4kb/s 甚至更低码率的语音编码算法。美国 MIT 林肯实验室提出的多带激励 (MBE) 语音编码方案是其中一种比较成功的算法<sup>[1]</sup>。这种编码算法突破了传统算法 (如 LPC 编码) 二元激励的局限性, 它将语音谱按基音频率分成若干谐波频带, 对各个频带的信号分别进行清 / 浊音判决, 然后根据各带是清音还是浊音, 分别用清音谱或周期序列信号谱作为激励, 产生其合成信号, 最后将各带信号相加, 形成全带合成语音。因为这种激励模型比较符合实际语音的特性, 能够使合成语音谱同原语音谱在细致结构上拟合得较好, 所以基于这种算法的声码器较二元激励声码器在合成语音的自然度和鲁棒性上都有较大的提高。

但是对于传统的 MBE 编码, 因为其谐波频带数随基音的大小而变化, 对于各帧语音, 需要量化的谐波带谱幅度值数目不定, 变化范围大约为 8-60 左右, 这对于大多数的应用都是不可接受的。为了解决这个问题, 有人提出了多带激励线性预测 (MBELP) 编码的思想<sup>[2]</sup>。线性预测是一种经常用来描述语音信号谱包络的方式。如果将声道传输函数看成是  $P$  阶全极点函数, 则可以用  $P$  个系数来描述, 在解码端, 只须用周期序列信号谱与其相乘, 便可获得相应的谐波带谱幅度值。从而将每一帧需要量化的参数数目基本固定下来。这对于参数的量化和降低码率都是很有利的。

本文给出了一种改进型 MBELP 编码算法, 与普通的 MBELP 编码算法相比, 在基音提取和清浊音的判断上采取了一些改进措施。通过试听比较, 在 2.4kb/s 码率下, 与普通的 MBELP 相比, 合成语音有一定的提高。

下面详细介绍这种改进型多带激励线性预测 (IMBELP) 编码算法。

## 2 算法介绍

MBE 是一种频域逼近的编码方法, 在分析过程中要提取的参数包括语音的基音频率各谐波带的清浊音判断、以及各谐波带的谱幅度。图 1 是编解码的框图。

下面介绍各个参数的提取方法。

<sup>1</sup> 1999-03-08 收到, 1999-09-30 定稿

国家 863 高科技计划和五邑大学科研基金资助项目

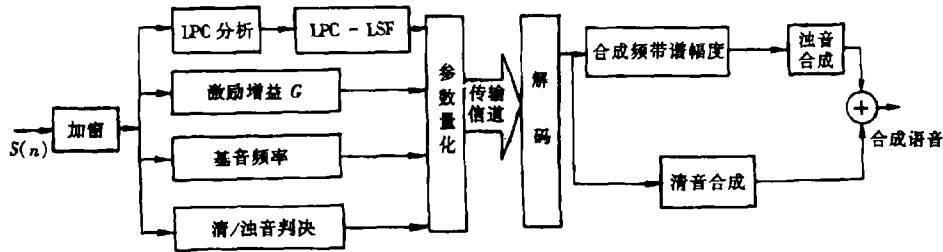


图 1 IMBELP 编解码框图

## 2.1 基音预测

基音预测是语音编码过程中非常普遍但又非常重要的一个步骤,在 MBE 编码中,它直接影响谐波带的划分、清浊音的判断、以及谱幅度的提取。传统的 MBE 编码对于基音的提取分两个步骤:第一个步骤是在时域上计算出整型的初始基音频率  $P$ ;第二个步骤是在频域上在初始基音频率  $P$  的附近进行更细致的搜索,确定更精确的基音频率值。实验表明,基音频率提取的正确与否,直接影响到合成语音的可懂度和自然度。因此,在本文的 IMBELP 编码算法中,对基音预测的方法采取了改进措施。

在传统的 MBE 编码中,基音频率的预测都是利用在正确的基音频率上合成语音与原始语音的拟合误差达到最小的原理,在时域和频域上对可能的基音频率值进行搜索。这是一种前向基音搜索算法,对于所有的语音帧,无论是浊音帧还是清音帧,都将提取出一个基音频率。而对于清音帧,这个提取出来的基音频率是无效的,如果利用它来提取其它参数,将会影响到参数的准确性。在本文的 IMBELP 编码中,采用了前向基音搜索算法和后向基音搜索算法相结合的方法来消除这种影响。该后向基音搜索算法如下:

$$P_0 = \begin{cases} \hat{P}_0, & \text{如果 } \text{sum}(\mathbf{V}) \geq \text{length}(\mathbf{V}) \times \text{th}, \\ P_{-1}, & \text{其它} \end{cases} \quad (1)$$

其中  $\mathbf{V}$  是当前帧包含所有频带清 / 浊音判决的向量 (第  $k$  频带为浊音带时  $v(k) = 1$ , 为清音带时  $v(k) = 0$ )。  $P_{-1}$  是前一帧的基音频率值,  $\hat{P}_0$  是当前帧搜索出的基音频率值,  $P_0$  是当前帧最终有效基音频率值,  $\text{th}$  是门限值。上式表明,如果当前帧中浊音带数目与总频带数的比例超过某一门限值,则该语音帧可认为是浊音帧,搜索出来的基音频率值有效。如果低于这一门限值,则该语音帧认为是清音帧,搜索出的基音频率值被认为是无效。考虑到基音频率的连续性,则直接采用上一帧语音的有效基音频率作为本帧语音的有效基音频率。我们可以从图 2 看出经过后向基音搜索算法后语音基音频率的平滑效果。

## 2.2 清 / 浊音判决

在本文的 IMBELP 编码算法中,频带清 / 浊音 (U/V) 的判断采用一种最多分成 9 个频带的判决方法。即将相邻的四个谐波分带合并成一个频带进行 U/V 判决,但合并后的频带数最多不超过 9 个。有关细节请参阅文献 [3]。

在传统的 MBE 编码中,仅仅是根据合成语音的频谱与原始语音谱的拟合程度来进行 U/V 判决。如果它们的拟合误差小于某个门限,则判为浊音,反之则为清音。这种 U/V 判决方法在准确提取基音的前提下能够取得较好的效果。但如果提取的基音不准,导致拟合误差增大,就会引起错误的 U/V 判决,从而影响到合成语音的质量。因此,在本文的 IMBELP 编码中,采用了双门限法来减少基音不准对 U/V 判决的影响。具体方法如下:

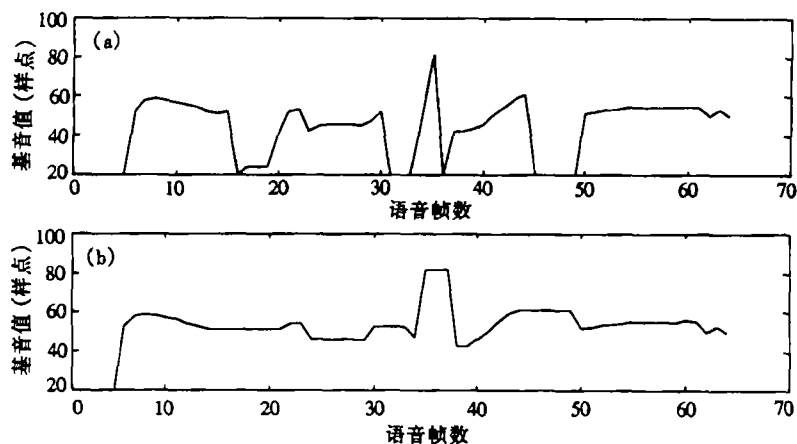


图2 二种算法基音搜索结果比较图  
(a) 仅用前向基音搜索算法预测出的基音值  
(b) 前后向基音搜索算法预测出的基音值

$$v(k) = \begin{cases} 1, & \text{如果 } \begin{cases} D_k \leq th_1 \text{ 或} \\ D_k \leq th_2 \text{ 且 } z_c \leq z_{cth} \end{cases} \\ 0, & \text{其它.} \end{cases} \quad (2)$$

其中  $D_k$  是频带的拟合误差,  $z_c$  是当前帧的过零率,  $z_{cth}$  是过零率门限。如当前帧过零率低于门限, 可以认为该语音帧中, 绝大多数频带是浊音带。在这种情况下, 应当提高清/浊音判决门限, 保证绝大多数频带判为浊音带。因此,  $th_2$  的值应当大于  $th_1$ , 但  $th_2$  也不能太大, 以防将无声段判为浊音帧。

通过增加过零率判断, 减少了在基音频率搜索不准时造成的频带 U/V 误判。图 3 显示了有无过零率判断时合成语音的质量差别。

### 2.3 频带谱幅度的提取

在 MBE 编码中, 语音频谱幅度的信息是用一组谱幅度抽样值  $\{A_m\}$  来表示的, 每个谐波带一个抽样值。因为在本文的 2.4kb/s IMBELP 编码中, 每帧编码的比特数较少, 所以只对频谱的幅度值处理, 把频谱的相位值丢弃。

正如前面所说,  $\{A_m\}$  是所有参数里最难量化的, 因为其数目的可变性导致难以用固定数目的比特来量化它, 尤其是对于低码率编码, 更是难以用较少的比特数来量化它。因此, 必须用其它方法来描述语音的谱幅度。

如果将语音产生的声道模型看成是全极点模型  $H(z) = 1/A(z)$ , 其中

$$A(z) = 1 - \sum_{k=1}^P a_k Z^{-k}, \quad (3)$$

$a_k$  是线性预测 (LP) 系数,  $P$  是线性预测阶数, 在本文的 IMBELP 编码中,  $P=10$ 。合成语音则可表示成

$$S(z) = H(z)E(z). \quad (4)$$

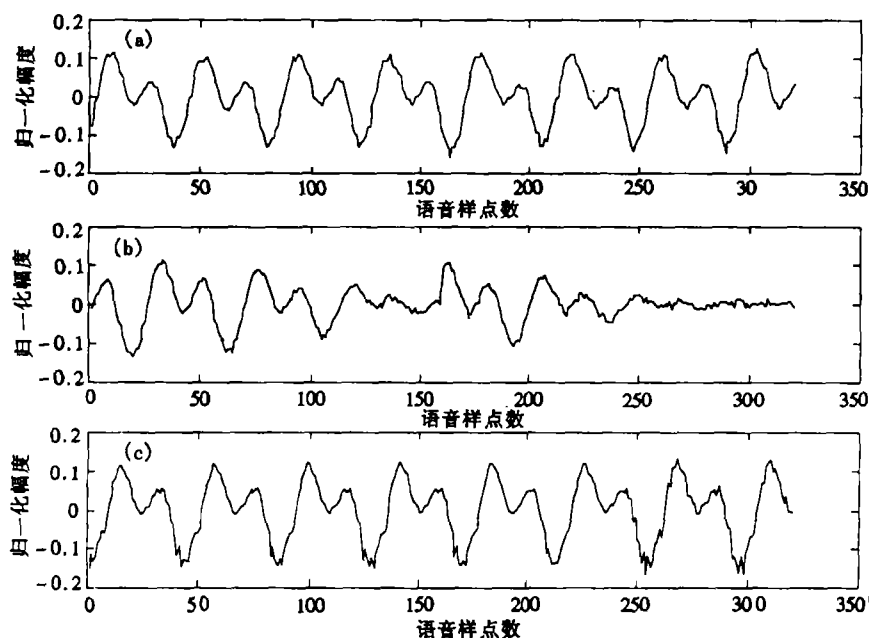


图 3 有无过零率合成语音比较图  
(a) 原始语音波形, (b) 无过零率处理的合成语音波形  
(c) 有过零率处理的合成语音波形

可以用  $P$  个 LP 系数  $\{a_k\}$  来代替需要传输的不定数目的  $\{A_m\}$ , 只须在解码端通过一些计算便可得到相应的频带谱幅度抽样。为了确定  $\{a_k\}$ , 需要计算出语音前  $P+1$  个自相关系数  $\{R_i\}$ 。这些自相关系数可以通过下式计算出<sup>[4]</sup>:

$$R_i = \sum_{m=0}^{M-1} A_m^2 \cos(i\omega_0), \quad (5)$$

其中  $\omega_0$  为基音角频率,  $M$  为频带谱幅度抽样的总数。在求出  $R_i$  后, 就可以利用德宾递推算法计算出  $\{a_k\}$ 。另外还须计算的参数是激励增益  $G$ , 具体计算方法参阅文献<sup>[3]</sup>。

在解码端, 通过  $\{a_k\}$  和  $G$ , 可以重建频带的谱幅度抽样值  $\{A'_m\}$ , 即

$$A'_m = \sqrt{G} / |A(e^{j\omega_0 m})|, \quad (6)$$

### 3 参数量化

2.4kb/s IMBELP 编码的语音采样频率为 8kHz, 每帧取 160 个采样点 (20ms)。基音频率和激励增益的量化采用标量量化, 所分配的比特数见表 1。对于基音频率, 直接量化其采样点数, 范围大约为 20-120 个点。因为 LP 系数不适于直接量化, 它的微小变化会造成极点位置的很大变化, 所以我们采用在数学上与 LP 系数  $\{a_k\}$  完全等价的另一种表示方式——线谱频率 LSF。LSF 参数具有误差相对独立的性质, 即某个频率点上的 LSF 的偏

表 1 2.4kbps 声码器比特分配表

参数	比特 / 帧	码率 (kbit/s)
基音频率	9	0.45
10 个 LSF 参数	24	1.2
激励增益	6	0.3
清 / 浊音判决	9	0.45
总数	48	2.4

差只对该频率附近的语音谱产生影响,而对其它 LSF 频率上的语音谱的影响不大。另外,如果保持 LSF 参数的有序有界性质,就可以保证全极点滤波器  $H(z)$  的稳定性。

在将 10 个 LP 系数转化为 10 个 LSF 后<sup>[3]</sup>,对这 10 个 LSF 参数采用分裂式矢量量化(SVQ)。将 10 个 LSF 参数分成 3 个子矢量,每个子矢量的维数分别为 3,3,4,每个子矢量分别用 8bit 量化。码本的训练采用 LBG 算法。

#### 4 仿真实验

我们使用美国 AD 公司的 EZ-Kit Lite 语音评价板作为仿真实验的平台。该实验板的信号数字处理器为 ADSP2181,所以我们的实验软件用 ADSP2181 汇编语音写成。在评价板上模拟了上述 2.4kb/s 改进型多带激励线性预测编解码器及语音合成。我们分别在较安静和强噪声(如播放一些音乐)环境下录制了“五邑大学”这一句话的男声录音。对这句话的录音进行取样(取样率为 8kHz),按改进型多带激励方法提取基音频率  $P$  和 LSF 等参数,然后按 IMBELP 语音合成的步骤进行语音合成。合成语音采用清音在频域合成,浊音在时域合成的方式。通过试听评分,其合成语音与 2.4kb/s MBELP 声码器的合成语音比较,音质有所提高,尤其在强噪声环境下合成语音的自然度有明显提高。图 4 是上述男声的时域原始波形和 2.4kb/s IMBELP 声码器的合成语音的时域波形。

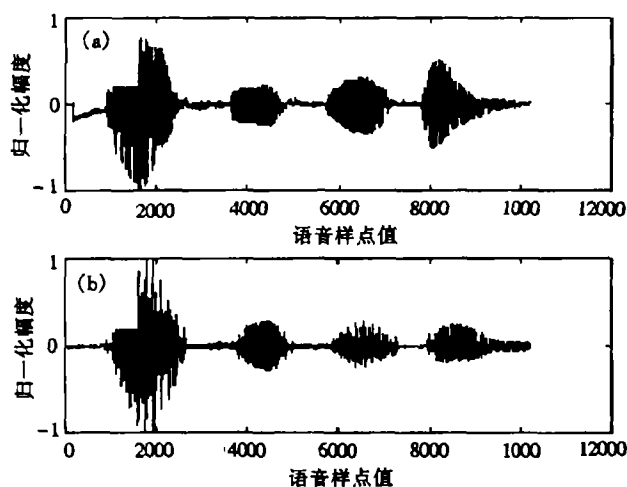


图 4 IMBELP 声码器合成语音波形

(a) 原始语音波形, (b) 2.4kb/s IMBELP 声码器合成波形

## 参 考 文 献

- [1] Griff D W, Lim J S. Multi-band excitation vocoder. IEEE Trans. on ASSP, 1988, ASSP-36(8): 1223-1235.
- [2] Yeldener S, Kondo A M, Evans B G. High quality multiband LPC coding of speech at 2.4kbit/s, Electron. Lett., 1991, 27(14): 1287-1289.
- [3] 杨行峻等编著. 语音信号数字处理, 北京: 电子工业出版社, 1995, 265-286.
- [4] Wang Tian, Tang Kun. A high quality MBE-LPC-FE speech coder at 2.4kbps and 1.2kbps. Proc. ICASSP'96, Atlanta., 1996, 208-211.

## AN IMPROVED MBELP CODER AT 2.4kbps

Zhong Ming     Jiang Taihui

*(Institute of Information Science, Wuyi University, Jiangmen, Guangdong 529020)*

**Abstract** This paper gives an improved MBELP speech coding algorithm at 2.4kb/s. Compared with conventional MBELP algorithm, this algorithm has some improvements on pitch estimation and unvoice/voice decision. The synthetical speech of vocoder based on this algorithm is better than that based on conventional one. After a brief introduction, the improved MBELP algorithm is discussed in detail. And the results of pitch estimation and U/V decision based on the improved MBELP are compared with those based on conventional MBELP.

**Key words** MBE coder, MBELP coder, Pitch estimation, U/V decision

钟 明: 男, 1974 年生, 硕士研究生, 从事通信与语音处理研究工作.

江太辉: 男, 1945 年生, 教授, 从事数字信号处理的教学与研究工作.