

高质量 4 ~ 8kb/s 变速率有限状态 ACELP 语音编码算法研究

唐 昆, 崔慧娟, 刘志勇, 冯重熙

(清华大学电子工程系, 微波与数字通信国家重点实验室, 北京 100084)

摘 要: 4 ~ 8kb/s 变速率有限状态代数码激励线性预测语音编码(VR-FS-ACELP)是一种具有延时较短、合成语音质量高、算法复杂度较低的语音编码算法。在线性预测(LP)参数量化上,使用预测式分裂式矢量量化,获得很高的量化效率。在自适应码本搜索上,采用了有限状态控制分数延时搜索的算法,有效地降低了运算量。对于随机码本,采用了具有多模结构的代数码本,提高语音合成质量。对于激励码序列的增益,采用了预测式矢量量化,有效地提高了量化精度。经非正式听音测试,其中 4kb/s 的合成语音质量超过了北美 8kb/s VSELP,接近长途质量,而 6kb/s 和 8kb/s 合成语音质量达到了长途质量,与 G. 729kb/s CS-ACELP 相当。

关键词: 变速率(VR); 有限状态(FS); 多模代数码本; 矢量量化(VQ)

中图分类号: TN912.32 **文献标识码:** A **文章编号:** 0372-2112 (2000) 01-0021-05

4 ~ 8kb/s Variable Rate-Finite State-Algebraic Code Excited Linear Prediction Speech Coding Algorithm

TANG Kun, CUI Hui-juan, LIU Zhi-yong, FENG Chong-xi

(Department of Electronic Engineering of Tsinghua University State Key Lab. on Microwave and Digital Communication, Beijing 100084, China)

Abstract: 4 ~ 8kb/s Variable Rate-Finite State-Algebraic Code Excited Linear Prediction (VR-FS-ACELP) is a high quality speech coding algorithm which has shorter time delay and low calculation burden. For line spectrum pairs, this algorithm uses a highly efficient split vector quantizer based on intra-frame and inter-frame correlations. For adaptive codebook searching, it uses a finite state fractional delay adaptive codebook which has low calculation complexity and high performance. This algorithm also uses a state controlled multimode algebraic codebook and a predictive vector quantizer for gains to further improve the quality of synthesized speech. The non-formal tests show that the speech quality of 4kb/s algorithm is better than that of IS-54 8kb/s VSELP and near to toll quality and the speech quality of 6kb/s and 8kb/s algorithms reaches toll quality.

Key words: variable rate(VR); finite state (FS); multimode algebraic codebook; vector quantization (VQ)

1 引言

变速率语音压缩编码除了可以作为通常的固定速率语音压缩编码使用之外,可以使数字话路倍增系统、语音包交换系统、码分多址移动通信等系统根据业务量动态调整编码速率,使系统的话音质量和容量得到合理的配置,最大限度地发挥系统的效能,因此近些年来变速率语音压缩编码得到了普遍的重视。国际电信联盟(ITU)在制订、公布了 G. 728 16kb/s LD-CELP 和 G. 729 8kb/s CS-ACELP 语音编码标准以后,也分别将这两个算法进一步扩充到其它速率,准备制订变速率语音编码标准。

从使用角度看,不仅要求算法合成语音质量高,还要求编码延时小、程序和数据所需存储器容量少、编码算法复杂度低等。码激励线性预测(CELP)是一种高效率的语音编码技术。

这类算法使用自适应码序列和随机码序列激励一个时变线性滤波器得到合成语音。由于采用了感觉加权、分析合成、矢量量化和后滤波等技术,CELP 能够在中低速率上给出高质量的合成语音。但在 8kb/s 及以下的速率上,由于编码速率低,激励码本尺寸小,合成语音质量较差。为了提高这一速率的合成质量,往往需要加大处理帧长,这会使编码延时加大。另外还需要增加算法复杂度和存储器容量,才能得到令人比较满意的合成语音质量。本文作者在 CELP 模型基础上提出了一种具有延时较短的 4、6、8kb/s 变速率有限状态代数码本激励线性预测(VR-FS-ACELP)算法。在线性预测(LP)参数量化上,利用了语音帧间的相关性,对线谱对(LSP)参数使用预测式分裂式矢量量化,在保证量化质量同时有效地压缩所需存储器容量,仅用 19bit 就能使 LP 参数平均谱失真小于 1dB。在自适应

收稿日期:1998-05-06;修订日期:1998-07-24

码本搜索上,采用了有限状态控制分数延时搜索的算法,在保证合成语音质量的同时,有效地降低了运算量.对于随机码本,采用了具有多模结构的代数码本,这种码本无需存储,并能提高语音合成质量.对于激励码序列的增益,采用了预测式矢量量化,有效地提高了量化精度.算法还采用了自适应后滤波技术,进一步改善了合成语音质量,经非正式听音测试,其中 4kb/s 的合成语音质量超过了北美 8kb/s VSELP,接近长途质量, MOS 分约为 3.9,而 6kb/s 的 8kb/s 合成语音质量达到了长途质量,与 G.729 8kb/s CS-ACELP 相当, MOS 分约为 4.0.编码速率的选择由应用系统控制,也可以由输入语音信号的能量自适应控制,这时需要在算法中加入相应的能量检测判决和控制模块.

2 VR-FS-ACELP 算法的基本工作原理

图 1 画出了 VR-FS-ACELP 算法编码器的工作原理框图.

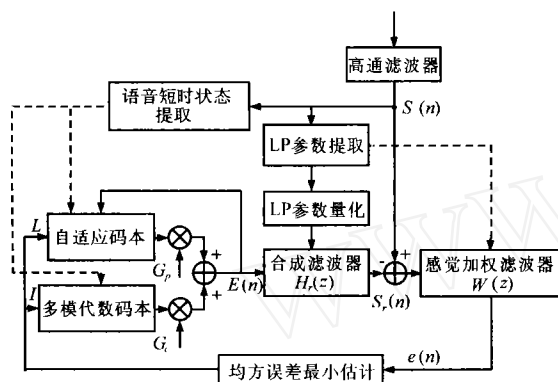


图 1 VR-FS-ACELP 算法编码器工作原理图

8kHz 采样的输入语音信号首先经过高通滤波器去除直流量得到语音信号 $S(n)$, 然后按每一帧提取 10 阶 LP 参数并转换成 LSP 参数进行量化, 对于 4kb/s 每帧为 20ms, 对于 6kb/s 和 8kb/s 每帧分别为 15ms 和 10ms. 每一帧分两个子帧进行处理, 用未量化和量化后的 LP 参数分别构成感觉加权滤波器 $W(z)$ 及合成滤波器 $H(z)$, 分别为

$$W(z) = A(z/\gamma_1)/A(z/\gamma_2) \quad (1)$$

$$\text{和} \quad H(z) = 1/\hat{A}(z) \quad (2)$$

式中: $A(z)$ 和 $\hat{A}(z)$ 是未量化和量化后的 LP 逆滤波器的传递函数; γ_1 和 γ_2 为常数, $\gamma_1 = 0.9$, $\gamma_2 = 0.6$.

从自适应码本和多模代数码本中取出的激励序列分别乘以各自的增益并相加后激励 $H(z)$, 产生本地重建语音信号 $S_r(n)$. 通过对 $S(n)$ 和 $S_r(n)$ 之间的感觉加权均方误差的计算, 确定使误差最小的最佳激励矢量 $P_{L_{OPT}}(n)$ 和 $V_{I_{OPT}}(n)$. 将它们在码本中的序号 L 和 I 、增益值 G_L 和 G_I 以及 LP 参数量化值传送给解码器, 就能够在解码器中合成重建语音.

自适应码本搜索方法以及多模代数码本结构的选择取决于语音帧的短时特性, 由语音短时特性参数控制. 该参数从 $S(n)$ 中每帧提取一次, 并用 1bit 编码, 也需要传送给解码器, 以便正确选择多模代数码本结构来合成语音.

自适应码本使用移位寄存器来存储码矢量, 利用码矢量

的叠接性来节省存储空间. 每一子帧更新一次, 将这一子帧的最佳激励序列 $E_{opt}(n)$ 存入自适应码本, 同时将最早的激励序列移出码本. 自适应码本的搜索和多模代数码本的搜索是分别进行的, 这样可以大幅度地减小运算量, 而性能几乎不受影响. 表 1 给出了该算法的 bit 分配, 其中 U/V 表示清音或浊音, UV 表示清音到浊音的过渡音, VU 表示浊音到清音的过渡音.

表 1(a) 4kb/s VR-FS-ACELP 算法的比特分配

参数	U/V		UV/VU		
	第一子帧	第二子帧	全帧	第一子帧	第二子帧
LSP 系数 $\{f_i\}$			19		
语音清浊状态			1		
最佳分数延时 L_{OPT}	8	4	12	8	8
最佳代数码字 I_{OPT}	18	18	36	18	14
增益矢量 $\{G_p, G_c\}$	6	6	12	6	6
总计			80		

表 1(b) 6kb/s VR-FS-ACELP 算法的比特分配

参 数	第一子帧	第二子帧	全帧
LSP 系数 $\{f_i\}$			19
最佳分数延时 L_{OPT}	8	5	13
最佳代数码字 I_{OPT}	23	23	46
增益矢量 $\{G_p, G_c\}$	6	6	12
总 计			90

表 1(c) 8kb/s VR-FS-ACELP 算法的比特分配

参 数	第一子帧	第二子帧	全帧
LSP 系数 $\{f_i\}$			19
最佳分数延时 L_{OPT}	8	5	13
最佳代数码字 I_{OPT}	17	17	34
增益矢量 $\{G_p, G_c\}$	7	7	14
总 计			80

3 语音编码参数的提取和量化

3.1 LP 参数的提取和量化

LP 参数每帧提取一次, 其阶数为 10. 语音信号 $S(n)$ 经过加窗后计算自相关函数, 并对此相关函数加指数窗实现 80Hz 频带展宽, 以利于系统稳定, 然后用 Levinson-Durbin 算法求出 LP 系数 $a_i, i = 1, \dots, 10$. 由于每帧 bit 数很少, 为了获得高质量的合成语音, 必须对 LP 参数采用很高效率的量化方法, 以节省出 bit 来保证激励码本的尺寸足够大. LSP 参数具有比 A 参数更高的量化效率, 因此 VR-FS-ACELP 算法不直接对 A 参数量化, 而是将 A 参数 $a_i, i = 1, \dots, 10$, 转换成 LSP 参数 $f_i, i = 1, \dots, 10$ 再进行量化.

B.S. Atal 等人^[1]提出了一种 24bits 的分裂式矢量量化算法, 它将 10 维的 LSP 矢量分裂成两个 5 维的子矢量, 并对这两个子矢量分别进行 12bits 的矢量量化. 这种量化方法利用了 LSP 系数的帧内相关性, 取得了比非均匀标量量化法高得多的量化性能和量化效率, 平均频谱失真已接近 1dB, 但它需要两个 12bits 的量化表, 所用的比特数、计算复杂度和量化表的存储量都比较大. 为了进一步提高量化效率和量化性能, 提出了一种 LSP 系数的预测式分裂式矢量量化算法 (Predicted

Split Vector Quantization, PSVQ), 将 LSP 系数量化所需的比特数降到了 19bits, 同时保持了较高的量化性能。

设 $f_i^n (i=1, 2, \dots, 10)$ 为第 n 帧的第 i 个 LSP 频率, 为了减小计算和量化对象的动态范围, 可以利用 LSP 频率的有序有界性, 将第 i 个 LSP 频率的平均值 $\bar{f}_i, (i=1, 2, \dots, 10)$ 从 f_i^n 中减去, 以得到第 n 帧的第 i 个无偏 LSP 频率 \tilde{f}_i^n 。

$$\tilde{f}_i^n = f_i^n - \bar{f}_i, \quad i=1, 2, \dots, 10 \quad (3)$$

令 $\hat{f}_i^n (i=1, 2, \dots, 10)$ 为 \tilde{f}_i^n 的量化值, 则 \tilde{f}_i^n 的预测值 \tilde{f}_i^n 可以由下式得到

$$\tilde{f}_i^n = \sum_{k=1}^{M_p} \alpha_k^i \tilde{f}_i^{n-k}, \quad i=1, 2, \dots, 10 \quad (4)$$

式中: $M_p=2$ 为预测器的阶数; $\alpha_k^i, i=1, 2, \dots, 10, k=1, 2, \dots, M_p$ 是预测器的系数, 其值由实验确定, \tilde{f}_i^n 的预测误差 r_i^n 为

$$r_i^n = \tilde{f}_i^n - \hat{f}_i^n, \quad i=1, 2, \dots, 10 \quad (5)$$

经过这样的预测处理后, r_i^n 的动态范围比原来 LSP 频率 f_i^n 的动态范围减小了很多, 因而非常适合于作为量化的对象以提高量化的精度。大量的统计实验也证明了这一点。

在 PSVQ 中, 将 LSP 频率的预测误差矢量 $\{r_i\}$ 分裂成两个子矢量, 前 5 个分量构成一个子矢量, 后 5 个分量构成另一个子矢量, 对这两个子矢量分别进行矢量量化, 并利用人耳对不同频率敏感程度不同的特性, 采用加权欧氏距离作为量化失真的度量。设 k_1 和 k_2 分别为每两个矢量量化后的序号, $\hat{r}_i(k_1), i=1, 2, \dots, 5$ 和 $\hat{r}_i(k_2), i=6, 7, \dots, 10$ 分别为两个矢量量化表中对应于序号 k_1 和 k_2 的预测误差子矢量, $\hat{f}_i(k_1)$ 和 $\hat{f}_i(k_2)$ 分别为对应于 k_1 和 k_2 的量化后无偏 LSP 频率的子矢量, 则有

$$\hat{f}_i(k_1) = \tilde{f}_i + \hat{r}_i(k_1), \quad i=1, 2, \dots, 5 \quad (6a)$$

$$\hat{f}_i(k_2) = \tilde{f}_i + \hat{r}_i(k_2), \quad i=6, 7, \dots, 10 \quad (6b)$$

再将它们分别加上各自的平均分量 $\bar{f}_i, i=1, 2, \dots, 10$ 就得到量化后的 LSP 参数。两个预测误差矢量码本是在加权欧氏距离最小准则下用“模拟退火”算法训练出来的, 其性能比常用的 LBG 算法要好, 平均每个码本能够节约 1bit^[2]。

3.2 有限状态自适应码本搜索算法

自适应码本的延时搜索范围是 19~143, 覆盖了绝大多数基音周期值。为了提高合成语音质量, 特别是女声和童声的语音质量, 采用了分数延时自适应码矢量搜索。对于 19~85 个样点的基音周期范围, 使用 1/3 个样点的搜索精度, 对于 86~143 个样点的基音周期范围, 仍采用整数样点的搜索精度。这样, 自适应码本的码字每子帧可用 8bits 表示。分数延时激励矢量可以用整数延时激励矢量的插值来得到。

自适应码本的搜索通常采用合成分析的闭环搜索算法, 即使利用码矢量的叠接性, 采用快速算法, 运算量也相当大, 特别是对于具有分数延时的搜索。因此, VR-FS-ACELP 先用开环的方法在每子帧利用加权语音算出开环最佳基音周期, 然后再用闭环的方法在延时为开环最佳基音周期的一个小范围内搜索出最佳自适应码矢量。实验表明, 这种开环-闭环分两

步自适应码本搜索方法, 可以使运算量降低很多, 而合成语音质量基本不受影响。

由于语音信号具有短时平稳性, 对于浊音段语音帧, 两个子帧的基音周期相差很小, 最佳自适应激励码序列延时值相差不大, 可以采用差分搜索和编码。第一子帧使用 8bits 编码, 第二子帧只在第一子帧延时值周围搜索最佳激励矢量, 并用 4 或 5bits 编码, 节省下来的 bit 用来加强代数码本。计算机模拟表明, 采用这种方法能够进一步改善浊音段合成语音的质量。对于清音段, 由于自适应码本起的作用不大, 代数码本对合成语音质量更大些, 因此也可以采用差分搜索和编码。在 4kb/s 编码中, 由于每个子帧有 10ms 长, 对于过渡段语音, 无论是清音到浊音, 还是浊音到清音, 两个相邻子帧的基音周期变化常会超出 4bits 差分编码的范围, 采用这种搜索方法会影响合成语音质量。为此, 论文作者提出一种有限状态分数延时自适应码本搜索算法, 这种算法需要对语音信号进行简单的清浊音判决, 并根据判决的结果决定采用什么样的方法来对自适应码本进行搜索。

设 L_i^* 为第 i 子帧的最佳开环整数延时, 子帧语音的长时相关性可以用该子帧的加权语音 $S_{ue}(n)$ 与它延时 L_i^* 后的值加权 $S_{ue}(n-L_i^*)$ 的归一化互相关系数 g_i 来表示。

$$g_i = \frac{(\sum_{n=0}^{N-1} S_{ue}(n) S_{ue}(n-L_i^*))^2}{\sum_{n=0}^{N-1} S_{ue}^2(n) \sum_{n=0}^{N-1} S_{ue}^2(n-L_i^*)} \quad (7)$$

式中 N 为子帧长度。 g_i 在每帧的平均值 g_{ave} 可以用来表示语音的清浊性, g_{ave} 越大, 说明语音的浊音性越强, 反之则清音性越强。相邻子帧最佳开环整数延时 L_i^* 的变化值 ΔL_i^* 可以表示语音的平稳程度, ΔL_i^* 越小, 说明语音越平稳, 反之则越不平稳, 其中 $\Delta L_i^* = L_i^* - L_{i-1}^*$ 。由此可得到语音的状态转移图, 如图 2 所示。

图中共有 4 个状态, 状态 0 表示语音的清音状态 U , 状态 1 表示语音从清音到浊音的过渡状态 UV , 状态 2 表示语音平稳的浊音状态 V , 状态 3 表示语音从浊音到清音的过渡状态 VU 。图中的状态输入 $\alpha_1 \alpha_2$ 被定义为

$$\alpha_1 = \begin{cases} 1, & g_{ave} \geq g_{th} \\ 0, & g_{ave} < g_{th} \end{cases} \quad (8a)$$

$$\alpha_2 = \begin{cases} 1, & \Delta L_i^* \leq \Delta L_{th}^*, \forall i \\ 0, & \text{otherwise} \end{cases} \quad (8b)$$

其中 g_{th} 和 ΔL_{th}^* 分别为 g_{ave} 和 ΔL_i^* 的阈值, 需由实验统计得到。

由语音的状态转移图和式 (7)、(8) 可以方便地得当前语音所处的状态, 并根据不同的语音状态对分数延时自适应码本采用不同的搜索方法以提高基音检测的准确性。状态 0 和状态 2 对应的是语音平稳的清音段和浊音段, 第二子帧的

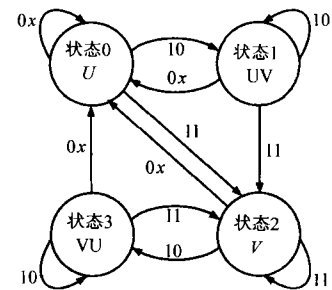


图 2 语音的有限状态转移图

码本搜索都在第一子帧最佳分数延时的基础上采用差分搜索的方法搜索出本子帧的最佳分数延时,并用差分量化的方法对它进行量化以减少码本的量化比特.状态1和状态3对应的是语音不平稳的过渡段,这种语音仍有较强的准周期性,而且相邻子帧的基音周期变化很大,因而不能采用前后子帧的差分搜索和差分量比以免造成太大的基音检测误差,必须在19~143的范围内对自适应码本进行重新搜索以保证最后合成语音的质量.在4kb/s编码中采用了这种搜索算法使各种类型语音段的合成语音质量都令人比较满意.对于6或8kb/s,由于子帧长度相对较短,两个相邻子帧的基音周期变化没有那么大,并且差分编码使用5bits表示,因此不需要采用有限状态表示.

3.3 多模代数码本

在传统的 CELP 算法中,随机激励码本中的激励序列采用的是完全随机的序列,这种激励序列对于清音段语音是合适的,但对于过渡段和浊音段的语音来说并不十分合适,因为这两种语音信号均具有较强的准周期性,仅用模型中的一个长时基音滤波器(自适应码本)来表示它们的准周期性是不够的.为了使合成语音在过渡段和浊音段与原始语音也能很好匹配,要求随机激励码本也能产生具有一定的准周期性的激励序列^[3].代数码本就具有这种能力,但由于4kb/s编码速率较低,采用单模式的固定代数码本已不能满足合成高质量语音的要求,因此在 VR-FS-ACELP 算法中,我们采用具有特殊结构的多模代数码本.多模代数码本能够根据语音的不同状态来调整其结构和 bit 分配,使不同状态的语音都能获得高的语音合成质量.

在这个多模代数码本中,每个随机码矢量 $V_l(n)$, $n=0, 1, \dots, N-1$ 最多只能有 N_c 个非零脉冲,这 N_c 个非零脉冲幅度和位置的取值都要受到一定的限制以满足一定的代数结构和比特分配要求.若设 S_i 和 P_i 分别为第 i ($i=0, 1, \dots, N_c-1$) 个非零脉冲的幅度和位置,则 S_i 的取值可以是一个定值(+1或-1),也可以是一个变值(± 1),而 P_i 的取值则具有一定的代数结构,它的第 k 个取值 P_i^k 可以用下式来表示:

$$P_i^k = O_f + i \cdot N_d + k \cdot N_c \cdot N_d, i=0, 1, \dots, N_c-1 \quad (9)$$

其中, N_d 为相邻非零脉冲的最小间隔, $O_f=0, 1, \dots, N_d-1$ 为非零脉冲相对于位置0的共同偏移值.当非零脉冲的位置超出子帧的范围即 $P_i^k > N-1$ 时则表示该非零脉冲实际上并不存在.因此,随机码矢量 $V_l(n)$ 中非零脉冲的个数实际上可以取 $\leq N_c$ 的任何整数,这样可以更好地保证随机码矢量一定的准周期性和一定的随机性,从而保证在提高浊音段和过渡段语音合成质量的同时不会影响清音段语音的合成质量.

由式(8),表2给出了4kb/s的多模代数码本在浊音和清音模式下随机码矢量中非零脉冲幅度和偶位置的取值情况,其中子帧长度 $N=80$,非零脉冲的个数 $N_c=5$, $N_d=2$, S_i 为+1、-1、+1、-1和 ± 1 .非零脉冲的位置也可以都取奇,另外 S_i 也可以都取符号相反的数值,这样清、浊音段的代数码本可用18bits表示.对于过渡段语音, $N_d=4$,代数码本则用14bits表示.

6kb/s算法的代数码本 $N=60$, $N_c=5$, $N_d=1$,每子帧用

23bits表示,其中5bits表示5个非零脉冲的幅度,18bits表示非零脉冲的位置.8kb/s算法的代数码本 $N=40$, $N_c=5$, $N_d=1$,每子帧非零脉冲的位置用17bits表示,非零脉冲的幅度分别为+1, -1, +1, -1和 ± 1 .

表2 4kb/s算法非零脉冲幅度和位置的取值

脉冲序号	脉冲幅度	脉冲位置
0	+1	0, 10, 20, 30, 40, 50, 60, 70, (80)
1	-1	2, 12, 22, 32, 42, 52, 62, 72, (82)
2	+1	4, 14, 24, 34, 44, 54, 64, 74, (84)
3	-1	6, 16, 26, 36, 46, 56, 66, 76, (86)
4	± 1	8, 18, 28, 38, 48, 58, 68, 78, (88)

随机码本的搜索也是在感觉加权均方误差最小的准则下,用合成分析的方法在闭环下进行的,就是在随机码本中搜索出能使 C_l^2/G_l 最大的随机码矢量 $V_l(n)$, $n=0, 1, \dots, N-1$, 其中

$$C_l = \sum_{n=0}^{N-1} T(n) V_l(n) \quad (10)$$

$$G_l = \sum_{n=0}^{N-1} V_l^2(n) \quad (11)$$

式中: N 为子帧长度; $T(n)$, $n=0, 1, \dots, N-1$ 为随机码本搜索的目标矢量; $V_l(n)$, $n=0, 1, \dots, N-1$ 为随机码本在序号为 l 时的输出码矢量 $V_l(n)$ 通过加权合成滤波器 $H_w(z) = H_r(z) \cdot W(z)$ 的零状态响应,即 $V_l(n)$ 与 $H_w(z)$ 零状态冲激响应 $h(n)$, $n=0, 1, \dots, N-1$ 的卷积,也就是

$$V_l'(n) = \sum_{m=0}^n V_l(m) h(n-m), n=0, 1, \dots, N-1 \quad (12)$$

对于多模代数码本, $V_l(n)$ 中最多只能有 N_c 个非零脉冲,即

$$V_l(n) = \sum_{i=0}^{N_c-1} S_i \delta(n - P_i), n=0, 1, \dots, N-1 \quad (13)$$

式中: $\delta(n)$ 为单位脉冲(Unit Pulse). S_i 和 P_i 分别为第 i 个非零脉冲的幅度和位置.将式(12)和(13)代入式(10)和(11),便可得到 C_l 和 G_l 的简化表达式

$$C_l = \sum_{i=0}^{N_c-1} S_i R(P_i) \quad (14)$$

$$G_l = \sum_{i=0}^{N_c-1} \varphi(P_i, P_i) + 2 \sum_{i=0}^{N_c-2} \sum_{j=i+1}^{N_c-1} S_i S_j \varphi(P_i, P_j) \quad (15)$$

其中: $R(i)$, $i=0, 1, \dots, N-1$ 为 $T(n)$ 与 $h(n)$ 的互相关; $\varphi(i, j)$, $i, j=0, 1, \dots, N-1$ 为 $h(n)$ 的协方差矩阵.它们的表达式为

$$R(i) = \sum_{n=i}^{N-1} T(n) h(n-i), i=0, 1, \dots, N-1 \quad (16)$$

$$\varphi(i, j) = \sum_{n=\max(i, j)}^{N-1} h(n-i) h(n-j), i, j=0, 1, \dots, N-1 \quad (17)$$

由于矩阵 $\varphi(i, j)$ 是一个对称矩阵,非零脉冲的位置 P_i 要受到一定的限制,因而在码本搜索之前只需计算和存储真正需要的那些 $R(i)$ 和 $\varphi(i, j)$ 以减小算法的运算复杂度.

由于 $V_l(n)$ 中最多只有 N_c 个非零脉冲,因而对这种多模

代数码本可以进行非常快速有效的搜索。

3.4 增益的矢量量化

由于语音信号是一种短时平稳的准周期信号,CELP 模型中自适应码本和随机码本的增益在相邻的子帧之间存在着一定的相关性,因此,我们可以采用线性预测的方法来对码本的增益进行预测处理以进一步去除中间的冗余度,从而达到提高增益量化精度和量化效率的目的。由于自适应码本增益的大小与语音信号准周期性的强弱密切相关,它在各子帧之间变化较大,相邻子帧间的相关性较弱,不太适合采用预测的方法来处理,而随机码本的增益则由于语音信号的短时平稳性而具有较强的帧间相关性,因而可以采用线性预测的方法来减小量化对象的动态范围,提高增益的量化精度和量化效率,以进一步提高低速率语音编码的语音合成质量。具体的预测和量化方法可以参阅文献[4]。

4 算法的实时模拟及性能

4~8kb/s VR-FS-ACELP 算法充分利用了语音的基本特征,采用了有限状态分数延时自适应码本、多模代数码本和预测式矢量量化等技术,获得了很高的语音编码质量。经计算机实时模拟,通过人耳主观听觉实验等性能测试表明,4.0kb/s 算法的编码质量要明显优于北美的 8.0kb/s VSELP 算法,非正式 MOS 分可达 3.9 左右,不同测试语句编码加权信噪比也都高于北美的 8.0kb/s VSELP 算法,如表 3 所示。而 6kb/s 和 8kb/s 算法不同测试语句编码加权信噪比虽然略低于 ITU G. 729 8kb/s CS-ACELP,但主观听音几乎听不出差别。测试采用 A、B 选择方法,语音采用中国科学院声学所提供的测试语句。

表 3 语音信号编码质量的测试结果

测试语句	d1.dat	d2.dat	d3.dat	d4.dat	d5.dat	d6.dat	平均	非正式 MOS
8.0kb/s VSELP	14.24	13.14	11.81	11.66	11.63	13.59	12.68	3.7
4.0kb/s VR-FS-ACELP	14.98	14.68	13.27	12.10	12.44	14.09	13.59	3.9
6.0kb/s VR-FS-ACELP	17.10	16.98	15.42	14.34	14.84	16.88	15.93	3.9~4.0
8.0kb/s VR-FS-ACELP	17.67	17.38	15.91	15.08	16.18	17.38	16.60	4.0
8.0kb/s CS-ACELP	18.20	18.09	16.15	15.49	16.16	17.68	16.96	4.0

参考文献

- [1] K.K. Paliwal and B.S. Atal, Efficient vector quantization of LPC parameters at 24 bits/frame. IEEE Trans. on Speech and Audio Processing, Jan. 1993, 1(1):3~14
- [2] Zeger K. Global optimal vector quantizer design by stochastic relaxation. IEEE Trans. on Signal Processing, Feb 1992, 40(2):310~322
- [3] Adoul J. P., Mabilieu P. Fast CELP coding based on algebraic codes. Proc ICASSP, 1987, 1957~1960
- [4] 刘志勇等. 一种高质量 8kb/s ACELP 语音编码算法及其实时实现. 电子学报, 1997, 25(7):72~74



唐 昆 1970 年毕业于清华大学并留校任教,现为清华大学电子工程系教授,从事数字通信、语音编码、自适应信号处理等领域的研究工作。



崔慧娟 1970 年毕业于清华大学并留校任教,现为清华大学电子工程系副教授,从事信源编码、多媒体通信系统等领域的研究工作。