

# 高质量的 4 kb/s 散布脉冲 CELP 语音编码算法

鲍长春

(北京工业大学电子信息与控制工程学院,北京 100022)

**摘 要:** 本文提出了一种散布脉冲 CELP(DP-CELP) 语音编码算法,激励矢量由特殊结构的代数码书与固定形式的散布脉冲的卷积获得,这种激励源有效地改善了重建语音质量,但未增加代数码书搜索的复杂度.非正式的主观听力测试表明,这种 4 kb/s DP-CELP 语音编码算法的合成语音质量非常接近 G.723.1 中 6.3 kb/s 语音编码器.

**关键词:** 语音编码; 矢量量化; 线性预测; 散布脉冲码书

**中图分类号:** TN912.32 **文献标识码:** A **文章编号:** 0372-2112 (2003) 02-0309-05

## A High Quality Dispersed-Pulse CELP Speech Coding Algorithm at 4 kb/s

BAO Chang-chun

(Electronic Information and Control Engineering College, Beijing Polytechnic University, Beijing 100022, China)

**Abstract:** A dispersed-pulse CELP (DP-CELP) speech coding algorithm is proposed. The excitation is generated by means of convoluting the algebraic codevector with a fixed dispersion pulse. The constructed speech quality is efficiently improved with this excitation, but the search complexity for algebraic codebook not increased. The informal subjective test shows that the speech quality for this DP-CELP algorithm at 4 kb/s closes to that of a G.723.1 codec at 6.3 kb/s.

**Key words:** speech coding; vector quantization; linear prediction; dispersed-pulse codebook

### 1 引言

八十年代出现的码激励线性预测 (CELP) 语音编码技术<sup>[1]</sup>是近十几年来最成功的,它用线性预测技术提取声道参数,应用听觉加权技术和波形编码准则去优化激励信号,即在闭环基础上,从一个具有许多典型激励矢量的码书中寻找主观失真最小的激励矢量,从而以低于 16 kb/s 码率获得高质量的合成语音,现有的许多语音编码标准都基于这种技术,如美国联邦标准的 4.8 kb/s CELP<sup>[2]</sup>,ITU-T 的 G.728,16 kb/s LD-CELP<sup>[3]</sup>;G.729,8 kb/s CS-ACELP<sup>[4]</sup>和 G.723.1,5.3/6.3 kb/s 双速率 CELP<sup>[5]</sup>,北美 TDMA 移动电话标准 IS54 8 kb/s VSELP<sup>[6]</sup>,北美 CDMA 移动电话标准 IS96 8.5 kb/s QCELP<sup>[7]</sup>,日本的 TDMA 全速率数字移动电话标准 JDC 6.7 kb/s VSELP 和半速率数字移动电话标准 JDC 3.6 kb/s PSFCELP<sup>[8]</sup>.目前许多 4kb/s 语音编码标准候选算法也基于 CELP 技术,如较具潜力的 eX-CELP 候选算法<sup>[9]</sup>.

从目前的 4~8kb/s CELP 类语音编码研究中看到,激励源的选择是一个关键的问题,人们普遍认同代数码书是最为成功的,因为它不需要存贮,同时又能以不太多的比特数表达激励信息.但当编码速率降至 4kb/s 时,为了提高量化效率,激励矢量的长度必须大于 8~10ms,这时还用很少的符号脉冲来描述激励信号就会导致语音质量的急剧下降,这也就是说传统的 CELP 算法在 4kb/s 速率质量难以提高的根本原因

在于对模型激励信号的描述还不够精确,而不在于目前的声道模型本身.为了克服这个问题,本文提出了一种散布脉冲 CELP(DP-CELP) 语音编码算法.激励矢量由特殊结构的代数码书与固定形式的散布脉冲的卷积获得,在不增加代数码书搜索复杂度的前提下,较好地解决了传统代数码书中符号脉冲不足带来的缺陷,有效地改善了重建语音质量.非正式的主观听力测试表明,这种 4kb/s DP-CELP 语音编码算法的合成语音质量非常接近 G.723.1 中 6.3kb/s 语音编码器.

### 2 DP-CELP 编译码器结构

图 1 为 DP-CELP 编码器原理框图.编码器以 CELP 编码模型为基础,类似于 G.729,包括三个模块:(1)短时线性预测;(2)自适应码书搜索;(3)散布脉冲码书搜索.合成语音借助于激励信号通过一个基于短时预测器系数的合成滤波器获得,合成滤波器的传递函数为

$$H(z) = 1/A(z) = 1/\left(1 - \sum_{i=1}^p a_i z^{-i}\right) \quad (1)$$

式中,  $i = 1, \dots, p$  为线性预测(LP)系数,  $p = 10$  为预测阶数,  $a_i$  由线性预测分析技术求得.激励信号由自适应码书和散布脉冲码书的加权和形成,两个权值分别为两个码书的增益.自适应码书表征了信号的周期性,它的作用相当于一个基音预测器.散布脉冲码书由特殊结构的代数码书与固定形式的散

布脉冲的卷积得到. 两个码书的下标和码书增益是由合成分析 (ABS) 过程确定的, 即通过使输入语音与合成语音之间的感性加权均方差最小来确定. 感觉加权滤波器的传递函数为

$$W(z) = A(z) / A(z') \quad (2)$$

其中  $A(z)$  的表达式同式 (1),  $0 < \alpha < 1$ , 本文取  $\alpha = 0.8$ , 加权合成滤波器使用了量化的 LP 系数. 在 DP-CHLP 编码器中需要量化和传送的参数是: 最佳固定码书的下标及增益, 最佳自适应码书的下标及增益和线性预测系数.

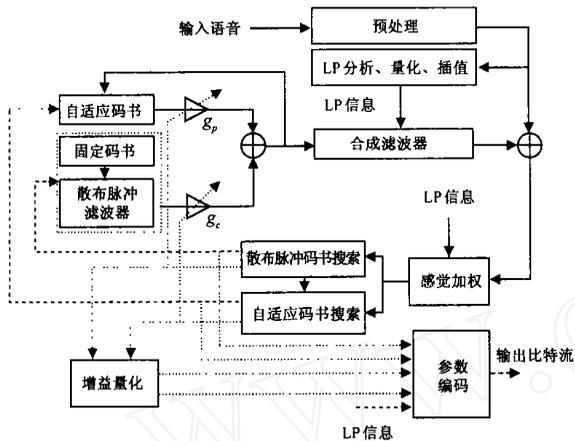


图 1 DP-CHLP 编码器原理框图

激励矢量是分层进行获取的, 首先对自适应码书搜索, 其次对散布脉冲码书搜索. 对于来自散布脉冲码书或自适应码书的第  $k$  个码矢量  $c^{(k)}$ , 线性合成滤波器  $1/A(z')$  的输出  $s_k$  等于零输入响应  $s_0$  与零状态响应  $g^{(k)} Hc^{(k)}$  之和, 即

$$s_k = s_0 + g^{(k)} Hc^{(k)} \quad (3)$$

其中  $g^{(k)}$  为第  $k$  个码字的增益,  $H$  为加权合成滤波器  $1/A(z')$  的脉冲响应矩阵, 它是一个  $N \times N$  的下三角 Toeplitz 矩阵, 在它的第一列里包含了滤波器的脉冲响应序列  $\{h_i\}_{i=1,2,\dots,N-1}$ .

令加权输入语音矢量为  $x$ , 则对应激励矢量  $c^{(k)}$  的误差为:

$$e_k = x - s_k = x_0 - g^{(k)} Hc^{(k)} \quad (4)$$

其中目标矢量  $x_0 = x - s_0$ . 通过使得平方误差  $E^{(k)} = e_k^T e_k$  最小, 即令  $\partial E^{(k)} / \partial g^{(k)} = 0$  可得  $g^{(k)}$  和  $E^{(k)}$  的表达式分别为:

$$g^{(k)} = \frac{x_0^T Hc^{(k)}}{Hc^{(k)T} Hc^{(k)}} \quad (5)$$

$$E^{(k)} = x_0^T x_0 - \frac{(x_0^T Hc^{(k)})^2}{Hc^{(k)T} Hc^{(k)}} \quad (6)$$

通过选择使  $E^{(k)}$  最小的序号  $k$ , 就可以得到最佳匹配码字, 或者等价的说, 选择使式 (6) 右边第二项最大的序号  $k$ , 对应这个序号的码字为所求的激励矢量.

图 2 给出了 DP-CHLP 译码器原理框图. 对语音参数解码后, 将来自固定码书的激励矢量与固定的散布脉冲矢量卷积, 再乘以对应的增益得到类噪声激励矢量, 将来自自适应码书的激励矢量乘以对应的增益得到类脉冲激励矢量, 将类噪声激励矢量和类脉冲激励矢量相加通过合成滤波器即可得到重

建的合成语音, 之后加入后滤波器, 以改善重建语音质量.

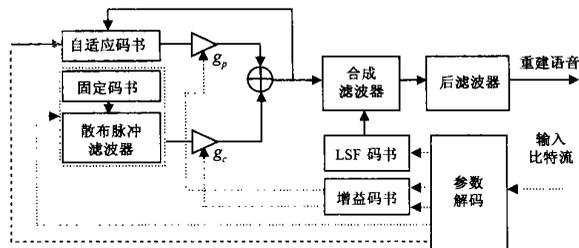


图 2 DP-CHLP 译码器原理框图

本文使用了 20ms 语音帧, 每帧计算和传输一次 LP 参数, 每帧划分为两个 10ms (80 个样点) 的子帧, 每子帧传送一次自适应码书和固定码书参数, 编码器的比特分配如表 1 所示. 运用预测矢量量化技术, 将 LP 参数转换到 LSF 域进行 20bit 矢量量化. 为了减少算法的复杂度, 基音延迟仅用整数延迟, 在两个子帧均用 7bit 编码, 自适应码书增益和散布脉冲码书增益均采用非均匀标量量化, 其中自适应码书增益用 4bit 量化, 散布脉冲码书增益用 5 bit 量化. 在每个子帧, 散布脉冲码书用 14bit 量化, 其中 13bit 表示脉冲的位置, 1bit 表示码书模式.

表 1 4kb/s DP-CHLP 语音编码器比特分配

参数	第一子帧	第二子帧	全帧
LSF 参数			20
自适应码书增益	4	4	8
散布脉冲码书增益	5	5	10
基音延迟	7	7	14
散布脉冲码书标号	14	14	28
总计			80

在接下来的各节里, 我们将描述 DP-CHLP 算法的主要功能模块, 即线性预测分析及量化、自适应码书搜索算法、散布脉冲码书的结构及搜索算法和自适应后滤波.

### 3 LP 分析、量化及内插

输入语音为 16 比特精度. 在 LP 分析之前首先要进行信号比例缩放和高通滤波, 信号比例缩放是为了减少实际定点处理中的溢出问题, 而高通滤波则用来滤除不必要的低频成分, 其传递函数为

$$H(z) = (1 - z^{-1}) / (1 - 0.992z^{-1}) \quad (7)$$

线性预测分析每 20ms 进行一次, 窗口的形状同 G 729 一样, 是一个 240 点的非对称窗, 但未引入额外的延时, 即 LP 分析窗由过去语音帧的 80 个样点和当前语音帧的 160 个样点组成. 窗的第一部分是半个哈密窗 (Hamming), 第二部分是余弦函数周期的四分之一. 首先在 240 点加窗的语音上计算自相关系数, 接着进行 60Hz 带宽展宽<sup>[4]</sup>, 最后应用 Levinson-Durbin 递推算法求解 LP 系数.

将计算得到的 LP 系数转换到 LSF (Line Spectrum Frequency) 域, 然后对 LSF 参数进行预测分裂矢量量化<sup>[11]</sup>. 设  $\bar{f}_i^{(n)}, i = 1, 2, \dots, p$  为第  $n$  帧的第  $i$  个 LSF 频率, 将第  $i$  个 LSF 频率的平均值  $\bar{f}_i, i = 1, 2, \dots, p$  从  $\bar{f}_i^{(n)}$  中减去, 得到第  $n$  帧的第  $i$  个无偏 LSF 频率  $\bar{f}_i^{(n)}, i = 1, 2, \dots, p$

$$\bar{r}_i^{(n)} = \bar{r}_i^{(n-1)} - \bar{r}_i, i = 1, 2, \dots, p \quad (8)$$

令  $\hat{r}_i^{(n)}, i = 1, 2, \dots, p$  为  $\bar{r}_i^{(n)}$  的量化值, 则  $\bar{r}_i^{(n)}$  的预测值  $\tilde{r}_i^{(n)}, i = 1, 2, \dots, p$  由下式得到<sup>[11]</sup>

$$\tilde{r}_i^{(n)} = \hat{r}_i^{(n-1)} + \hat{r}_i^{(n-1)}, i = 1, 2, \dots, p \quad (9)$$

其中,  $\hat{r}_i$  为帧内预测系数,  $\hat{r}_i$  为帧间内插系数. 式中上标括号中的值表示帧序号, 下标为 LSF 参数分量. 通过使如下平方预测误差最小

$$E_i = \sum_{n=1}^{N_f} [r_i^{(n)}]^2 = \sum_{n=1}^{N_f} [\bar{r}_i^{(n)} - \tilde{r}_i^{(n)}]^2 \quad (10)$$

即令  $\partial E_i / \partial \hat{r}_i = 0$  和  $\partial E_i / \partial \hat{r}_i = 0$ , 可确定系数  $\hat{r}_i$  和  $\hat{r}_i$ . 式中,  $N_f$  为训练数据的总帧数. 这时  $\bar{r}_i^{(n)}, i = 1, 2, \dots, p$  的预测误差  $r_i^{(n)}, i = 1, 2, \dots, p$  为

$$r_i^{(n)} = \bar{r}_i^{(n)} - \tilde{r}_i^{(n)}, i = 1, 2, \dots, p \quad (11)$$

经过这样的预处理后, 预测误差的动态范围比原来 LSF 频率  $\bar{r}_i^{(n)}$  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效率. 在实际量化中, 我们将 LSF 参数的预测误差矢量  $\{r_i^{(n)}\}$  分裂成两个子矢量, 令前四个分量构成一个子矢量, 后六个分量构成一个子矢量, 对这两个子矢量分别用 10 比特进行矢量量化<sup>[12]</sup>, 并利用人耳对不同频率敏感程度不同的特性, 采用 Paliwal 和 Atal 建议的加权欧氏距离作为量化失真测度<sup>[13]</sup>, 在 20 比特的平均谱失真为 0.9175dB, 达到了透明量化效果.

量化的 LP 参数用在第二子帧上, 而第一子帧的 LP 参数要由当前帧和前一帧的 LP 参数的线性内插获得, 线性内插在 LSF 域进行, 即第  $n$  子帧的第  $i$  个 LSF 参数可由下式确定:

$$LSF_n(i) = (1 - a_n) LSF_p(i) + a_n LSF_c(i) \quad (12)$$

式中  $LSF_c$  表示当前的一组 LSF 参数, 而  $LSF_p$  表示前一帧的一组 LSF 参数, 本文使用的加权系数为:  $a_1 = 0.5, a_2 = 1$ .

#### 4 自适应码书搜索

自适应码书由 147 位移寄存器 (SR) 构成<sup>[14]</sup>, 自适应码书的搜索过程就是搜索使式 (6) 右边第二项最大的码矢量, 该式的计算复杂度主要来源于卷积项  $Hc^{(k)}$ , 这里, 该卷积项用  $P_u(n)$  表示, 下标  $u$  表示整数基音延迟 20 ~ 147, 对应于自适应码书的 128 个码字序号. 当搜索完自适应码书和散布脉冲码书后, 两个码书中优化激励矢量的加权和移入 SR 更新自适应码书.

令  $N$  为子帧长, 当  $u > N - 1$  时, 第  $u$  个码矢量  $c^{(u)}$  的第  $n$  个元素为  $c^{(u)}(n) = SR(u - 1 - n), 0 \leq n < N - 1$ ; 当  $20 \leq u \leq N$  时, 第  $u$  个码矢量  $c^{(u)}$  的第  $n$  个元素为  $c^{(u)}(n) = SR(u - 1 - n \% u), 0 \leq n < N - 1$ , 其中, 符号  $\%$  表示求余. 根据这两种情况, 激励矢量  $c^{(u)}$  与感知加权合成滤波器的脉冲响应  $h(n)$  卷积  $P_u(n)$  具有如下递推公式<sup>[14]</sup>:

$$\begin{cases} P_u(0) = SR(u - 1), n = 0 \\ P_u(n) = SR(u - 1)h(n) + P_{u-1}(n - 1), 0 < n < u \\ P_u(n) = SR(u - 1)h(n) + P_{u-1}(n - 1) + P_u(n - u) \\ \quad - P_{u-1}(n - u), u \leq n < N - 1 \end{cases} \quad (13)$$

由于感觉加权因子的衰减作用, 当  $n > 40$  时,  $h(n)$  很小, 实验表明, 将  $h(n)$  截断到 40, 减小了自适应码书搜索的复杂度, 语音质量无感性降低. 确定了自适应码矢量后, 可根据式 (5) 获得自适应码书增益  $g_p$ , 并且  $g_p$  被限制在 0 ~ 1.4 之间 (码书搜索过程中仅考虑互相关项  $xHc^{(k)}$  大于零的基音延迟), 用 4 比特进行非均匀标量量化, 其中, 1 比特用于符号量化, 3 比特用于幅度量化, 幅度码书用 LBG 算法获得.

#### 5 散布脉冲码书结构及搜索

散布脉冲码书结构如图 3 所示, 它由固定码书中的 4 个非零脉冲与散布脉冲的卷积之和得到, 4 个非零脉冲的幅度及位置由两种模式构成, 如表 2 所示, 其中子帧长度为 80 个样点, 相邻脉冲的最小间隔为 2. 对于每种模式的码书, 前三个脉冲分别用 3bit 编码, 第四个脉冲用 4bit 编码, 码书模式用 1bit 编码, 这样固定码书可用  $3 + 3 + 3 + 4 + 1 = 14$  比特表示. 散布脉冲由截止频率为 3400Hz 的 FIR 滤波器的脉冲响应构

表 2 14bit 固定码书结构

脉冲	模式 1		模式 2	
	幅度	位置	幅度	位置
$m_0$	+1	0, 10, 20, 30, 40, 50, 60, 70	-1	1, 11, 21, 31, 41, 51, 61, 71
$m_1$	-1	2, 12, 22, 32, 42, 52, 62, 72	+1	3, 13, 23, 33, 43, 53, 63, 73
$m_2$	+1	4, 14, 24, 34, 44, 54, 64, 74	-1	5, 15, 25, 35, 45, 55, 65, 75
$m_3$	-1	6, 16, 26, 36, 46, 56, 66, 76	+1	7, 17, 27, 37, 47, 57, 67, 77
		8, 18, 28, 38, 48, 58, 68, 78		9, 19, 29, 39, 49, 59, 69, 79

成, FIR 滤波器采用窗函数法设计, 对应的滤波器系数为  $d_0, d_1, \dots, d_{N-1}$ , 其中  $d_i = 0, i = 21, 22, \dots, N - 1$ , 则图 3 中的散布脉冲矩阵  $D$  为一个  $N \times N$  的下三角 Toeplitz 矩阵.

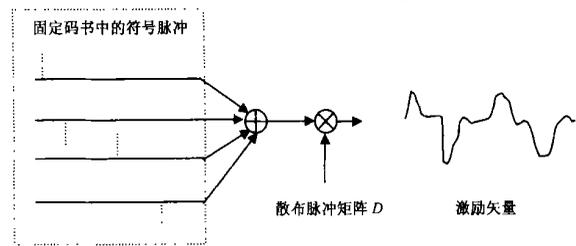


图 3 散布脉冲码书的基本结构

根据式 (6), 散布脉冲码书的搜索过程就是选择使下式最大的散布脉冲序列,

$$E_{\max}^{(k)} = \frac{C_k^2}{G_k} = \frac{(x_0^T H c^{(k)})^2}{H c^{(k)} c^{(k)T} x_0} \quad (14)$$

其中  $x_0$  是从减去零输入响应的加权语音  $x_0$  中再减掉自适应码书贡献  $P_u(n)$  的目标矢量, 即

$$x_0 = x_0 - g_p P_u(n) \quad (15)$$

经散布脉冲滤波器修正的脉冲响应矩阵  $H = HD$  是一个与  $H$  具有相同结构的下三角 Toeplitz 矩阵.

式 (14) 中的互相关项  $C_k = x_0^T H c^{(k)}$ , 为了减少计算量, 码书搜索前, 后向滤波的目标矢量  $d = x_0^T H$  可用下式先求出,

$$d(n) = \sum_{i=n}^{N-1} x_0(i) h(i - n), n = 0, 1, \dots, N - 1 \quad (16)$$

式(14)中的能量项  $G_k = H c^{(k)} = (x^{(k)})^T c^{(k)}$ , 其中  $H = H^T H$  为修正的脉冲响应  $h(n)$  的互相关矩阵, 其第  $(i, j)$  个元素可表示为

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

因为激励矢量  $c^{(k)}$  仅有 4 个非零脉冲, 所以, 互相关项  $C_k$  和能量项  $G_k$  可以分别表示为

$$C_k = s_0 d(m_0) + s_1 d(m_1) + s_2 d(m_2) + s_3 d(m_3) \quad (18)$$

$$G_k = \sum_{i=0}^3 \phi(m_i, m_i) + 2 \sum_{i=0}^2 \sum_{j=i+1}^3 s_i s_j \phi(m_i, m_j) \quad (19)$$

其中  $s_0, s_1, s_2$  和  $s_3$  为脉冲幅度, 如表 2 所示. 从上面的分析可以看出, 散布脉冲码书的搜索过程与固定码书是一样的, 未增加计算量, 仅在搜索前多计算一次矩阵乘法, 即  $H = HD$ .

因激励矢量的长度为 10ms, 可能几倍于基音周期长度, 这时散布脉冲激励中因缺少基音周期性可能导致语音质量下降, 为了改善音质, 特别是女性讲话者, 当基音延迟小于子帧长度时, 散布脉冲码矢量附加了一个固定增益的基音贡献, 即

$$c^{(k)}(n) = c^{(k)}(n) + 0.8c^{(k)}(n-u) \quad (20)$$

其中,  $u$  是从自适应码书中得到的基音延迟, 这种变化等价于在码书搜索前将  $h(n)$  进行如下修正,

$$h(n) = h(n) + 0.8h(n-u) \quad (21)$$

实验表明这种方法有效改善了合成语音质量.

确定了散布脉冲矢量后, 可根据式(5)获得散布脉冲码书增益  $g_c$ , 并用 5 比特进行非均匀标量量化, 其中, 1 比特用于符号量化, 4 比特用于幅度量化, 幅度码书用 LBG 算法获得.

## 6 自适应后滤波

本文使用的自适应后滤波器类似于 G 729, 由长时后滤波器和短时后滤波器级联组成<sup>[13]</sup>, 包括谱的倾斜(tilt)补偿和自动增益控制. 长时后滤波器针对基音谐波, 用来衰减基音谐波之间的谱的谷值; 短时后滤波器则针对共振峰, 用来衰减共振峰之间的语谱的谷值. 倾斜补偿滤波器用于平滑自适应后滤波器的频谱倾斜度, 后滤波器系数每 10ms 更新一次.

长时后滤波器传递函数定义为:

$$H_p(z) = (1 + p_g z^{-T}) / (1 + p_g g_l) \quad (22)$$

其中, 因子  $p$  控制长时滤波的数量, 其值为  $p = 0.5$ .  $T$  是基音延迟,  $g_l$  是增益系数. 通过使如下归一化相关函数最大来确定  $T$  和  $g_l$ ,

$$R(k) = \left[ \sum_{n=0}^{N-1} \gamma(n)\gamma(n-k) \right]^2 / \sum_{n=0}^{N-1} \gamma^2(n)\gamma^2(n-k) \quad (23)$$

式(23)中,  $\gamma(n)$  是  $\mathcal{A}(z/n)$  滤波得到残差信号,

$$\gamma(n) = \mathcal{S}(n) - \sum_{i=1}^{10} \hat{a}_i \mathcal{S}(n-i) \quad (24)$$

基音延迟  $T$  在  $u-1, u, u+1$  中选取, 其中  $u$  为译码器得到的延迟. 如果  $R(k) < 0.5$ , 则  $g_l = 0$ , 否则

$$g_l = \frac{\sum_{n=0}^{N-1} \gamma(n)\gamma(n-k)}{\sum_{n=0}^{N-1} \gamma^2(n-k)}, \quad 0 \leq g_l \leq 1.0. \quad (25)$$

短时后滤波器定义为:

$$H_f(z) = \frac{A(z/n)}{A(z/d)} = \frac{1 - \sum_{i=1}^{10} \hat{a}_i z^{-i}}{1 - \sum_{i=1}^{10} \hat{a}_i z^{-i}} \quad (26)$$

其中  $n = 0.5, d = 0.8$ .

倾斜补偿滤波器  $H_t(z)$  的传递函数为

$$H_t(z) = (1 - 0.4k_1 z^{-1}) \quad (27)$$

其中,  $k_1$  是由合成语音  $\mathcal{S}(n)$  估计的第一个自相关系数. 经  $H_t(z)$  后的语音表示为  $\mathcal{S}_f(n)$ .

自适应增益控制用来补偿重建语音信号  $\mathcal{S}(n)$  和后滤波信号  $\mathcal{S}_f(n)$  之间的增益差. 当前子帧的增益比例因子  $G$  为:

$$G = \frac{\sum_{n=0}^{N-1} |\mathcal{S}(n)|}{\sum_{n=0}^{N-1} |\mathcal{S}_f(n)|} \quad (28)$$

增益比例调整后的后滤波信号  $\mathcal{S}_f(n)$  为:

$$\mathcal{S}_f(n) = g^{(n)} \mathcal{S}_f(n), \quad n=0, \dots, N-1 \quad (29)$$

其中  $g^{(n)}$  逐点更新为:

$$g^{(n)} = 0.85g^{(n-1)} + 0.15G, \quad n=0, \dots, N-1 \quad (30)$$

## 7 算法性能

为评价 4kb/s DP-CELP 语音编码算法的性能, 本文进行了非正式的 A/B 听力测试. 11 名听众通过试听比较了 4kb/s DP-CELP 编码器(B)和 6.3kb/s G 723.1 MP-MLQ 编码器(A)的质量, 放音时 A 和 B 无固定先后次序. 测试汉语语音由 8 个男性讲话和 8 个女性讲话组成, 共 16 句. 每名参加测试者要听完 A 和 B 产生的 16 对句子, 每听完一对句子后作出选择: 偏爱 A 或偏爱 B 或无偏爱(A、B 质量听不出差别), 最后对 11 名听众的试听结果进行平均, 得到最后主观评价结果. 测试结果如表 3 所示.

表 3 A/B 测试结果

编码器	偏爱 6.3 kb/s	偏爱 4 kb/s	无偏爱
测试语音	G 723.1 MP-MLQ	DP-CELP	
男声语音	30.6818 %	30.6818 %	38.6363 %
女声语音	22.7270 %	23.8636 %	53.4091 %
所有语音	26.7045 %	27.2727 %	46.0227 %

从表 3 可以看出, 4 kb/s DP-CELP 的语音质量与 6.3 kb/s G 723.1 MP-MLQ 相当, 这说明将散布脉冲码书和预测分裂式矢量量化技术应用于 CELP 模型中, 在 4 kb/s 可产生高质量的合成语音.

## 参考文献:

- [1] M R Schroeder, B S Atal. Code-excited linear predictive (CELP): high quality speech at very low bit rates [A]. Proc. IEEE ICASSP [C], 1985. 937 - 940.
- [2] Joseph P Campbell, Jr, Thomas E Tremain, Vanoy C Welch. The DoD 4.8 kbps standard (proposed federal standard 1016) [A]. Advances in Speech Coding [C]. Boston: Kluwer Academic Publishers, 1991. 121 - 133.
- [3] J H Chen, et al. A low-delay CELP coder for the CCITT 16 kb/s speech coding standard [J]. IEEE Journal on Selected Areas in Communications, 1992, 10(5): 830 - 849.

- [ 4 ] ITU-T Recommendation G.729. Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP) [S].
- [ 5 ] ITU-T Recommendation G.723.1. Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s[S].
- [ 6 ] I Gerson, M A Jasiuk. Vector sum excited linear prediction (VSELP) speech coding at 8 kb/s [A]. Proc. IEEE ICASSP[C]. 1990. 461 - 464.
- [ 7 ] W Gardner, P Jacobs, C Lee. QCELP: A Variable Rate Speech Coder for CDMA Digital Cellular. Speech and Audio Coding for Wireless Networks (edited by B. S. Atal, V. Cuperman, and A. Gersho.) [M]. Boston, MA: Kluwer, 1993. 85 - 92.
- [ 8 ] K Mano, T Moriya, S Miki, H Ohmuro, et al. Design of pitch synchronous innovation CELP coder for mobile communications[J]. IEEE Journal on Selected Areas in Communications, 1995, 13(1): 31 - 40.
- [ 9 ] Jes Thyssen, Yang Gao, et al. A candidate for the ITU-T 4 kbit/s speech coding standard[A]. Proc. ICASSP[C]. 2001. 681 - 684.
- [ 10 ] K K Paliwal, B S Atal. Efficient vector quantization of LPC parameters at 24 bits/frame[J]. IEEE Transactions on Speech and Audio Processing, 1993, 1(1): 3 - 14.
- [ 11 ] Bao Changchun, Dai Yisong. One-step interpolation predictive vector quantization of LSP parameters[J]. The Journal of China University of Posts and Telecommunications, 1996, 3(1): 21 - 26.
- [ 12 ] 卓力, 鲍长春. 一种高效、透明的线谱频率参数矢量量化器[J]. 信号处理, 1999, 15(增刊): 154 - 158.
- [ 13 ] Juin-Hwey Chen, et al. Adaptive postfiltering for quality enhancement of coded speech[J]. IEEE Transactions on Speech and Audio Processing, 1995, 3(1): 59 - 71.
- [ 14 ] 鲍长春. 低比特率数字语音编码基础[M]. 北京: 北京工业大学出版社, 2001. 179 ~ 217.

#### 作者简介:

鲍长春 男, 1965 年出生, 博士, 现为北京工业大学电子信息与控制工程学院副院长, 教授, 博士生导师, 北京市高等院校青年学科带头人, 中国电子学会高级会员, 中国电子学会信号处理学会委员, 中国电子学会青年工作委员会主任, 通信学报、信号处理学报和北京工业大学学报编委, 长期从事语音压缩编码的研究工作, 在该领域主持科研课题 10 余项, 发表论文 30 余篇, 著有《低比特率数字语音编码基础》一书, 研究兴趣包括数字信号处理、语音信号处理和语音压缩等。

www.cnki.net