

# 低复杂性代数码激励线性预测语音编码<sup>\*</sup>

成 德 源

(深圳大学电子工程系, 深圳 518060)

**摘 要** 基于码激励线性预测 (CELP) 语音编码技术, 提出一种低复杂性 8 kb/s 代数码激励线性预测 (ACELP) 语音编码. 代数码本中的码矢量具有多值脉冲幅度, 码本搜索采用次最佳部分码本快速搜索法, 基频延迟仅采用整数延迟, 增益系数与 LSP 频率采用标量量化. 解码器中采用自适应基频前置滤波器与频谱后置滤波器, 以进一步提高再生语音的质量.

**关键词** 代数码激励线性预测, 码激励线性预测, 语音编码

**分类号** TN 912.31

近年来, 码激励线性预测 (CELP) 语音编码技术已广泛应用于比特率从 4~16 kb/s 的电话频带语音压缩. 4.8 kb/s CELP 已被定为美国联邦政府标准. 相对于其它的语音压缩方法, 比特率从 4~16 kb/s 范围的 CELP 语音编码技术可提供最好的再生语音的质量<sup>[1]</sup>. CELP 语音编码技术有三个显著的特征: ① 解码器接收到的数据是一个用于再生语音的综合滤波器的参数和一个激励信号. ② 综合滤波器是一个以线性预测 (LP) 分析为基础的时变滤波器. ③ 激励信号的编码采用综合分析法, 并且矢量量化技术被采用于激励信号的编码中. 编码器所确定的激励信号通常由码本中的码矢量和与码矢量对应的增益系数来表示.

虽然 CELP 码有不少优点, 但是它的编码复杂性较高, 码本存储占用的空间也较大. 为降低 CELP 码中码本存储占用空间, Salami 等提出码率为 8 kb/s 代数 CELP 语音编码 (ACELP)<sup>[2]</sup>. ACELP 编码技术结合了 CELP 编码的基本技术, 并采用一个具有高代数结构的码本, 以节省码本存储占用空间. 然而 ACELP 编码复杂性很高, 约需要 24~29 MIPS, 而且 ACELP 码采用 40 个语音抽样点作为一帧, 这给目前常用的语音实时处理系统实现实时编码处理带来一定的难度.

考虑到以上几点, 本文提出一种低复杂性 ACELP 语音编码. 语音的抽样频率为 8 kHz, 每连续 64 个抽样点组成一帧, 基频延迟仅采用整数延迟, 增益系数与 LSP 频率采用标量量化, 代数码本中的码矢量具有多值脉冲幅度, 码本搜索采用次最佳部分码本快速搜索法.

## 1 ACELP 编码器组成

ACELP 编码器组成如图 1 所示. 输入语音的抽样频率  $f_s = 8$  kHz. 每连续 64 抽样点组

\* 广东省高教厅重点扶持学科资助项目

收稿日期: 1997-04-09 成德源, 男, 52 岁, 副教授

成一个语音子帧. 每子帧语音为 8 ms. 每两个语音子帧组成一个语音帧. 在每一帧语音中, 第一个子帧称为奇子帧, 第二个子帧称为偶子帧. 编码器每接收到一帧语音后, 对当前帧语音中的奇子帧和前一帧语音中的偶子帧进行编码处理. LP分析采用互相覆盖的长度为三个子帧的窗口, 即由当前帧语音的两个子帧和前一帧语音中的偶子帧组成当前的 LP分析帧. LP分析帧的语音为 24 ms, 192个抽样点. 通过 Hamming窗口后, 计算 10阶当前 LP分析帧的 LP分析参数, 并转换成 LSP频率  $f_i, i=1, 2, \dots, 10$ . 在编码过程中, 此当前 LSP频率用于当前帧语音中的奇子帧的编码. 然后再将当前 LP分析帧的 LSP频率与前一个 LP分析帧的 LSP频率的平均内插用于前一帧语音中的偶子帧的编码. 当前帧语音, 前一帧语音, 当前 LP分析帧语音, 及子帧语音之间的关系见图 2.

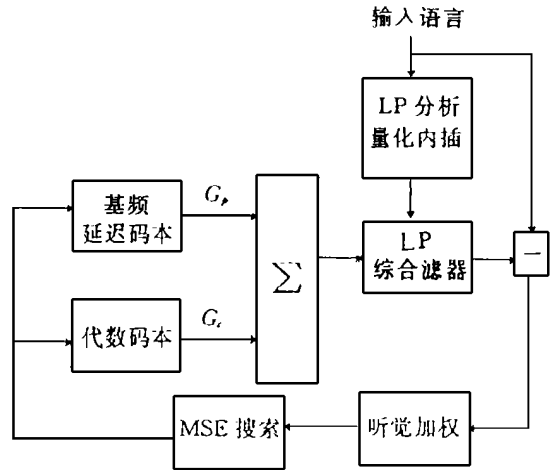


图 1 ACELP编码器原理

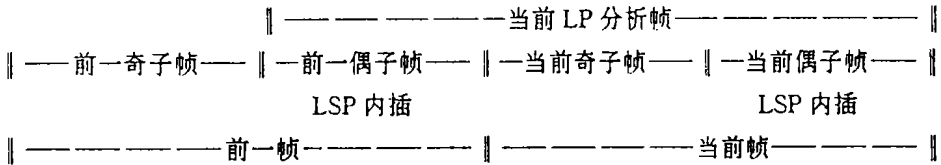


图 2 帧与子帧间互相关系

在图 1 中, LP综合滤波器是一个全极点滤波器, 可表示为

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

其中  $a_i, i=1, 2, \dots, p$ , 称为 LP分析参数,  $p$  为滤波器的阶, 取  $p=10$ .

在编码器中, 最佳基频延迟码矢量与代数码本中的最佳激励码矢量均由综合分析法来确定. 综合分析法中的听觉加权滤波器可表示为

$$W(z) = A(zN_1) / A(zN_2) \quad (2)$$

其中  $A(z)$  即为式 (1) 中的 LP倒置滤波器,  $N_1, N_2$  为听觉加权系数, 取  $N_1=0.9, N_2=0.6$ .

为了降低编码复杂性, LSP频率采用标量量化. 由于相邻 LSP频率之间的差值的标准偏离远比 LSP频率本身的标准偏离小<sup>[3]</sup>, 所以第一个 LSP频率及 9个相邻 LSP频率之间的差值作标量量化, 共采用 34个 bits, 分配见表 1.

## 2 自适应基频延迟码本

(自适应基频延迟码本中的码矢量长度为 64点, 与语音子帧同长. 考虑到语音信号中存

在基频长期周期性, 语音信号通过 LP 倒置滤波器后的残留信号可作为自适应基频延迟码本. 残留信号中延迟为  $T$  个抽样点的 64 点即构成基频延迟为  $T$  的基频延迟码矢量.

尽管分数基频延迟能产生更好的再生语音质量, 但由于其码本搜索复杂性相当高, 故仅采用整数基频延迟, 延迟范围为 20~147 个抽样点, 即 2.5~18.375 ms. 由于整数基频延迟所引起的再生语音质量的下降, 将设法由采用多值脉冲幅度代数码本来弥补.

在自适应基频延迟码本搜索中, 奇子帧与偶子帧采用不同的方法. 对于每个奇子帧语音, 采用延迟范围为 20~147 个抽样点的闭环全部码本搜索. 根据综合分析法, 以均方误差作为失真量度, 在残留信号中延迟 20~147 个抽样点的范围内, 获得一个最好的基频延迟  $T_b$ . 延迟为  $T_b$  的基频延迟码矢量所产生的再生语音与原始语音相比具有最小听觉加权均方误差  $d_{\min}$ . 为避免  $T_b$  实际上是基频的倍频, 将延迟为  $T_b/2$  的基频延迟码矢量所对应的听觉加权均方误差  $d_{1/2}$  与  $d_{\min}$  相比较. 如果  $0.85d_{1/2} < d_{\min}$ , 则最佳基频延迟确定为  $T_{opt} = T_b/2$ . 否则, 最佳基频延迟  $T_{opt} = T_b$ .

考虑到相邻奇偶子帧语音之间的相关性, 对于每个偶子帧语音, 在前一奇子帧的最佳基频延迟  $T_{opt}$  的基础上, 仅在  $T_{opt}-16$  至  $T_{opt}+15$  的延迟范围内, 用综合分析法搜索残留信号, 获得一个最佳的基频延迟. 于是, 需用 7 个 bits 来表示奇子帧的最佳基频延迟, 5 个 bits 来表示偶子帧的最佳基频延迟. 平均用 6 个 bits 来表示子帧的最佳基频延迟.

在搜索最佳基频延迟码矢量的同时, 得到与之相应的增益系数  $g_p$ . 根据  $g_p$  的统计特性, 采用 LBG 算法设计一个 4 bits 的标量量化码本, 并采用二进制树搜索法进行快速搜索.

### 3 代数码本

代数码本的码矢量的长度为 64 个样点, 与每一子帧语音长度相同. 码本具有如下代数结构: 每一码矢量中仅有 8 个非零脉冲, 其幅度可为 +1 或 +0.5, 以及 -1 或 -0.5, 其中 4 个脉冲的幅度还可以为零. 这些非零脉冲仅可出现在码矢量中的一些位置, 如表 2 所示. 整个码本中的码矢量可以同时改变其符号. 虽然这码本包含有巨大数量的码矢量, 但若表示其代数结构, 仅需要 33 个 bits.

表 2 代数码本中脉冲幅度和可出现的位置

脉冲序号	幅度	可出现的位置								
1	+ 0.5 或 + 1.0	0	8	16	24	32	40	48	56	
2	- 0.5 或 - 1.0	1	9	17	25	33	41	49	57	
3	+ 0.5 或 + 1.0	2	10	18	26	34	42	50	58	
4	- 0.5 或 - 1.0	3	11	19	27	35	43	51	59	
5	+ 0.5 或 + 1.0 或 0	4	12	20	28	36	44	51	X	
6	- 0.5 或 - 1.0 或 0	5	13	21	29	37	45	53	X	
7	+ 0.5 或 + 1.0 或 0	6	14	22	30	38	46	54	X	
8	- 0.5 或 - 1.0 或 0	7	15	23	31	39	47	55	X	

X 表示脉冲幅度为零

对每一个子帧语音而言, 当在确定最佳代数码矢量时, 该子帧的最佳基频延迟码矢量已经确定. 于是将代数码本的每一个码矢量与最佳基频延迟码矢量相加后, 一一通过 LP 综合滤波器, 用综合分析法确定最佳代数码矢量及其相应的增益系数  $g_c$ . 实验证明此多值脉冲幅度代数码本在很大程度上改善了再生语音的质量. 与 Salami<sup>[2]</sup>提出的单值脉冲幅度代数码本相比, 再生语音的信噪比提高约 1 dB.

如果代数码本搜索采用全部码本搜索法, 则可用一个包含有 8 个循环的多层循环来实现. 每个循环相应于搜索对应脉冲的最佳位置和幅度. 对于如此巨大的码本, 考虑到目前可采用的 DSP 处理器和实时处理系统硬件计算速度, 全部码本搜索法使实现实时处理变得非常困难. 于是, 必须考虑采用次最佳部分码本搜索法而又不降低再生语音的质量.

次最佳部分码本搜索法由 3 个多层循环组成. 第一个多层循环包含有 4 个循环, 以确定前 4 个对应脉冲的最佳位置和幅度. 码本的符号也在第一个多层循环中确定. 第二和第三个多层循环各包含有 2 个循环, 分别确定 2 个对应脉冲的最佳位置和幅度. 实验表明, 在次最佳部分码本搜索法中, 即使仅仅是一小部分的码本被搜索, 其再生语音质量的降低几乎不能被听觉出来.

对应于最佳码矢量的增益系数的量化采用 4 bits 标量量化. 不是线性量化增益系数本身, 而是对增益系数的对数作线性标量量化.

上述 ACELP 语音编码的比特率为 8 kb/s, bits 分配如表 3 所示. 由于采用次最佳部分码本快速搜索法, 编码复杂性约为 12 MIPS. 于是, 采用单片 DSP 处理器和语音实时处理系统, 例如采用由美国 D2 技术公司研制的单片 TMS320C50 语音实时处理系统<sup>[5]</sup>, 完全可以实现实时 ACELP 语音编码.

表 3 bits 分配表

参数	bits / 子帧
LSP 频率	34 / 2
基频延迟	(7 + 5) / 2
基频增益	4
代数码本	33
码矢量增益	4
总计	64

## 4 前置滤波器与后置滤波器

为进一步改善再生语音质量, 在解码器中采用前置滤波器与后置滤波器, 如图 3 所示. 在 LP 综合滤波器之前, 采用一个自适应基频前置滤波器, 以提高激励信号的周期性. 前置滤波器可表示为

$$H_p(z) = (1 + z^{-T}) / (1 + g) \quad (3)$$

其中,  $T$  为基频延迟,  $g$  为平衡前置滤波器的输入信号与输出信号之间的能量的增益系数.

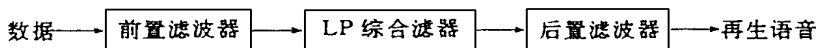


图 3 解码器原理

在前置滤波器之后是 LP 综合滤波器, 其参数可从接收到的 LSP 频率中求得. 在 LP 综合滤波器之后, 采用一个由听觉加权滤波器与一个倾斜补偿滤波器级联而成的频谱后置滤波器. 听觉加权滤波器与编码器中的一样, 但听觉加权系数为  $V_1 = 0.6$  和  $V_2 = 0.9$ , 与编码时的听觉加权系数正好相反. 倾斜补偿滤波器用于平滑整个后置滤波器的频谱倾斜度, 可表示为

$$H_t(z) = 1 + z^{-1} \quad (4)$$

其中, 如果  $r_{\omega} \leq 0$ , 则  $\_ = 0$ , 如果  $r_{\omega} > 0$ , 则  $\_ = 0.5 r_{\omega}$ ,  $r_{\omega}$  为第一个反射系数值, 可从 LSP 频率中求得.

为评估再生语音的质量, 1995年7月由加州大学圣巴巴拉分校进行了一次 MOS 测试, 结果其英语的 MOS 成绩为 3.9.

## 参 考 文 献

- 1 Gersho A. Advances in speech and audio compression IEEE Proc. Data Compression, 1994, 6 (82): 900~ 918
- 2 Salami R, Laflamme C, Adoul J P, et al. A toll quality 8 kb/s speech codec for the personal communication system (PCS). IEEE Trans. Veh. Technol, 1994 (43): 808~ 816
- 3 Ramachandran R, Sondhi M, Seshadri N, et al. A two codebook format for robust quantization of line spectral frequencies. IEEE Trans. Speech Audio Processing, 1995, 3 (3): 157~ 168
- 4 Laflamme C, Adoul J P, Su H Y, et al. On reducing computational complexity of codebook search in CELP coder through the use of algebraic codes. In IEEE Int. Conf. Proc. ICASSP90, 1990. 177~ 180
- 5 Voice Developing System V DS50. D2 Technologies Inc. 1995

## A Low Complexity ACELP Speech Codec

*Cheng Deyuan\**

**Abstract** An 8 kb/s ACELP speech codec is proposed. Its encoding algorithm is based on Code-Excited Linear Prediction (CELP) with a multi-level pulse amplitude Algebraic Codebook (ACELP). Only integer pitch delay is searched. There are eight non-zero pulses with fixed amplitudes, each of +1, or +0.5, or -1, or -0.5, or zero, in each algebraic codevector of length of 64 samples. An efficient non-exhaustive codebook search method is developed. The codevector's gain and LSP frequencies are scalarly quantized. A pitch prefilter and a spectral postfilter are adopted in speech decoder to enhance the reconstructed speech quality.

**Keywords** ACELP, CELP, speech coding

\* Department of Electronic Engineering, Shenzhen University, Shenzhen 518060