

码激励线性预测语音编码器中的非均匀和部分搜索域代数码书

白燕宁 鲍长春

(北京工业大学语音与音频信号处理研究室 北京 100022)

摘要 该文基于代数码激励线性预测(ACELP)语音编码算法提出了非均匀和部分搜索域代数码书。非均匀代数码书由代数码书的脉冲非均匀统计特性确定,部分搜索域代数码书则由代数码书矢量的周期性确定,该方法有效地弥补了低比特率情况下代数码书中脉冲数不足的缺点。在使用上述两项技术时,为保持基音的连续性,该编码器对语音段和非语音段采用了不同的基音估计方法。主观和客观的听力测试表明,当该技术应用于4kb/s 散布脉冲码激励线性预测(DP-CELP)语音编码器时,重建语音的质量得到明显改善,尤其是对女性讲话者。

关键词 语音编码, 码激励线性预测, 散布脉冲码书, 部分搜索域代数码书

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2006)11-2041-04

Non-uniform and Part-searching-area Algebraic Codebook for Code Excited Linear Prediction Speech Coder

Bai Yan-ning Bao Chang-chun

(Speech and Audio Signal Processing Lab, Beijing University of Technology, Beijing 100022, China)

Abstract This paper presents a non-uniform and part-searching-area algebraic codebook based on Algebraic Code Excited Linear Prediction(ACELP) speech coding algorithm. The non-uniform algebraic codebook is determined by the non-uniform statistical properties of the algebraic codebook, and the part-searching-area is determined by the periodicity of the algebraic codebook excitation vector, which makes up the insufficient numbers of signed pulses in algebraic codebook at low bit rate. In order to preserve the continuity of pitch, different pitch detection methods are employed for speech/silence frame when these two techniques are used. Subjective and objective test results indicate that the reconstructed speech quality of 4kb/s DP-CELP speech coder is improved based on these techniques, especially for the female speakers.

Key words Speech coding, Code excited linear prediction, Dispersed-pulse codebook, Part-searching-area algebraic codebook

1 引言

码激励线性预测(Code Excited Linear Prediction, CELP)是最近20年来语音编码中最为成功的一项技术,它是许多语音编码标准的基础,这其中包括16kbit/s的ITU-T/G.728(LD-CELP),8kbit/s的ITU-T/G.729(CS-ACELP)^[1]和6.3/5.3kbit/s双速率的ITU-T/G.723.1(MP/MLQ-ACELP)^[2]。

CELP语音编码模型的一个主要部分是它的激励码书。在各种码书中,代数码书被公认为是最为成功的,因为它不需要存储,同时又能以不太多的比特数来表达激励信息。虽然代数码激励线性预测(ACELP)在中速率上可以取得高质量的合成语音,但是当速率低于4kbit/s时,要想取得通信质量仍然是一项挑战。国际电信联盟ITU-T在过去几年中一直在进行具有通信质量的4kbit/s速率语音编码器的标准化工作,但是直到现在,仍然没有一个候选方案能够满足所有的性能指标要求。

合成语音质量下降的主要原因就在于低比特速率下符号脉冲数目的不足,从而导致对激励信号的描述不够精确。为了解决这一问题,各国学者已经提出了一系列的改进方法^[3-6],其中一些技术已经被成功地应用于一些低速率语音编码器中。

本文基于4kb/s DP-CELP语音编码算法^[6]提出了改善语音质量的非均匀和部分搜索域代数码书。

本文的第2节将详细介绍非均匀和部分搜索域代数码书的结构以及基音估计方法,第3节给出改进的4kb/s DP-CELP语音编码器,第4节给出编码器的主客观性能,并对实验结果进行分析,最后对全文进行总结。

2 非均匀和部分搜索域代数码书

在原先的DP-CELP语音编码器^[6]中,代数码书有两种模式,每种模式都包含4个非零脉冲。模式1和模式2分别对应由偶数位置和奇数位置上的脉冲构成的码书,两种模式的脉冲符号是相反的。具体的位置和符号由表1给出。

2005-03-21收到,2005-10-31改回
北京市教委科技发展项目(KM200310005024),国家自然科学基金(60372063)和北京市自然科学基金(4042009)资助课题

表 1 14bit 代数码书结构
Tab.1 14bit algebraic codebook structure

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

2.1 非均匀代数码书

改进的代数码书具有非均匀的结构^[7]。我们的统计实验显示代数码书的激励脉冲在子帧的开始部分出现的概率要明显大一些。图 1 给出了实验结果。在我们的实验中，子帧长度为 80 个样点，每子帧包含 4 个符号脉冲。脉冲的位置被分成 16 个区域，每 5 个样点组成一个区域。我们总共测试了 10,000 帧的数据，因为每帧包含两个子帧，所以总共统计了 80,000 个脉冲的位置。我们看到同其他区域相比，区域 1 和 2 有更多的脉冲个数。因此为了提高合成语音的质量，有理由使得代数码书的脉冲在子帧的开始部分更加密集一些。

改进后的脉冲的位置和符号由表 2 给出。与原先的码书相比，候选脉冲密度在子帧中是不同的，子帧的开始部分具有更大的密度。实验表明，该方法对男性和女性讲话者的合成语音质量均有改善。

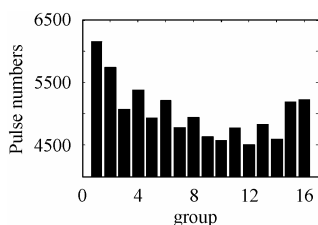


图 1 代数码书的非均匀统计特性

Fig1 Non-uniform statistical properties of algebraic codebook

表 2 14bit 非均匀代数码书(子码书 1)

Tab.2 14bit non-uniform algebraic codebook (sub-codebook 1)

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

2.2 部分搜索域代数码书

虽然非均匀的代数码书能够改善重建语音的质量，但是

非零脉冲的个数并没有增加，每子帧仍然是 4 个。为了进一步提高合成语音质量，我们采用部分搜索域代数码书来增加每子帧的非零脉冲个数。

在中速率的语音编码器中，代数码书激励通常表现出均匀的脉冲分布特性。但是在低比特率的情况下，代数码书激励则表现出很强的周期性^[5]，尤其在平稳的语音段和过渡段更为明显一些。理论上，我们希望代数码书在结构上能够反映这一特点。

在 DP-CELP 语音编码器中，当基音延迟小于子帧长时，代数码书 $C(n)$ 要由式(1)进行修正，

$$C'(n) = \begin{cases} C(n), & n = 0, \dots, T-1 \\ C(n) + \beta C(n-T), & n = T, \dots, 79 \end{cases} \quad (1)$$

这里 β 是基音增益，且 $0.2 < \beta < 0.8$ 。这种变化等价于在代数码书搜索过程中，通过修正声管模型的单位脉冲响应 $h(n)$ 来实现。

由于当基音周期足够小时，一些非零脉冲可以由式(1)产生，因此很自然就会产生另一个想法：为什么不通过减小搜索区域的方法来进一步增加非零脉冲的密度呢？在原先的 DP-CELP 编码器中，搜索域与子帧的大小是相同的，两个候选位置的平均间隔为 2，但是如果我们把搜索域的长度定为子帧长的一半时，平均间隔就减少到了 1。图 2 给出了部分搜索域代数码书的思想。候选脉冲被限制在搜索域而不是子帧内。位于搜索域外面的脉冲由式(1)产生。看起来缩小搜索域的大小在编码过程中是一种冒险行为，但是实验结果显示，该方法确实改善了合成语音的质量。

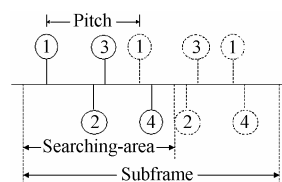


图 2 后向基音增强示意图

Fig.2 Sketch map of backward pitch enhancement

在我们的方案中，当基音周期小于 38 个样点时，搜索域的大小为 55，式(1)中的 β 等于 0.96。此时的代数码书结构由表 3 给出。在此情况下，两个候选位置的平均间隔是 1.5。进一步，当基音周期小于 28 个样点时，搜索域的大小为 40， β 值等于 0.97。代数码书结构由表 4 给出。在此情况下，两个候选位置的平均间隔为 1。

因此，改进的代数码书总共包含 3 个子码书。对每一子帧而言，一旦根据基音周期将最优的子码书确定，那么搜索最佳码矢的过程将只在该子码书中进行。由于子码书的选择是根据基音周期确定的，所以并不需要增加额外的比特。

表 3 14bit 非均匀代数码书结构(子码书 2)
Tab.3 14bit non-uniform algebraic codebook (sub-codebook 2)

脉冲	符号和位置	
	模式 1	模式 2
m_0	(+) 0,5,10,15,20,26,36,46	(-) 0,5,10,15,20,25,35,45
m_1	(-) 1,6,11,16,21,28,38,48	(+) 1,6,11,16,21,27,37,47
m_2	(+) 2,7,12,17,22,30,40,50	(-) 2,7,12,17,22,29,39,49
m_3	(-) 3,8,13,18,23,32,42,52	(+) 3,8,13,18,23,31,41,51
	4,9,14,19,24,34,44,54	4,9,14,19,24,33,43,53

表 4 14bit 非均匀代数码书结构(子码书 3)
Tab.4 14bit non-uniform algebraic codebook (sub-codebook 3)

脉冲	符号和位置	
	模式 1	模式 2
m_0	(+) 0,5,10,15,20,25,30,35	(-) 0,5,10,15,20,25,30,35
m_1	(-) 1,6,11,16,21,26,31,36	(+) 1,6,11,16,21,26,31,36
m_2	(+) 2,7,12,17,22,27,32,37	(-) 2,7,12,17,22,27,32,37
m_3	(-) 3,8,13,18,23,28,33,38	(+) 3,8,13,18,23,28,33,38
	4,9,14,19,24,29,34,39	4,9,14,19,24,29,34,39

2.3 自适应码书搜索

我们发现当使用上述两种技术时，保持语音段基音周期的连续性是非常重要的。在原来的 DP-CELP 语音编码器中，自适应码书参数(即基音周期)是通过闭环方法确定的。我们发现使用该方法很难在语音段保持基音的连续性，对女性讲话者而言更是如此。通过实验分析，我们发现原因是这样的：

在 4kbit/s 的速率上，为了降低编码合成滤波器系数和码书增益的比特速率，分析帧长通常是比较长的。在自适应码书实现基音滤波器的方法中，当延迟小于子帧长时，自适应码书的输出矢量需要用到当前子帧的激励信号，而当前子帧的激励信号目前还是未知的。原先的 DP-CELP 编码器中采用元素复制的方法解决这一问题。但是当基音延时足够小时，每个元素不得被复制多次，这时就很难保证求得的基音周期是准确的。如果采用类似于 G.729 中的方法，即：使用将来帧的数据，那么这一问题可以得到解决，但是会增加额外的编码延时。

为了不增加编码延时而又能够保持连续性，在改进的编码器中，对语音和非语音段采用了不同的基音估计方法。首先，使用开环的基音估计方法，接着，根据一些参数，比如尖峰度(peakiness)^[8]，将每帧分类为语音帧或非语音帧，然后使用不同的闭环基音估计方法：如果该帧属于语音帧，在开环基音延时附近的一个小范围内(8 个样点)进行搜索，从而确定最终的基音延时。如果该帧属于非语音帧，则在整个搜索区域内(20 至 143 个样点)重新进行闭环搜索以确定最终的基音延时。

在分类器参数的选择上，采用的是基于最小风险的贝叶斯决策。这是因为：使用上面提到的基音检测方法，将非语

音帧划分成语音帧的后果要比将语音帧划分成非语音帧的后果严重一些。所以我们设计的分类器会尽量避免第一类错误判决的发生。需要说明的是，这里并不需要额外的比特来表示该帧是语音帧还是非语音帧。实验显示这种方法在保持基音搜索准确性的同时，也保持了语音段基音周期的连续性。

3 改进的 4kbit/s CELP 语音编码算法

图 3 给出了改进的编码器的系统框图。输入语音的采样速率为 8000Hz，帧长是 20ms。每帧划分成两个 10ms 的子帧。线谱频率参数的内插，自适应码书的激励和代数码书的索引和增益都是每子帧计算一次。编码器参数的比特流分配如表 5 所示。除上文提到的自适应码书搜索和代数码书搜索外，其它几大模块说明如下。

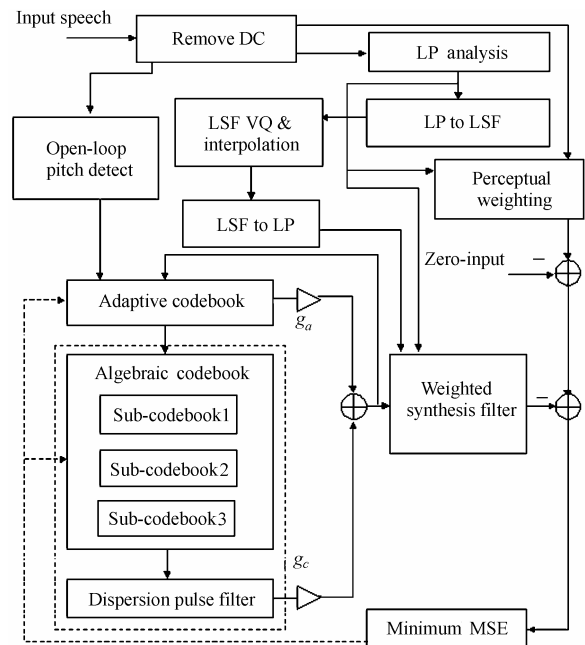


图 3 语音编码器的框图

Fig.3 Block diagram of the speech coder

表 5 4kbit/s 语音编码器比特流分配

Tab.5 Bit allocation of 4kbit/s speech coder

参数	比特数/帧
LSFs	20bit
自适应码书	8 bit/子帧 共 16 bit
固定码书模式	1 bit/子帧 共 2 bit
固定码书	13 bit/子帧 共 26 bit
自适应码书增益	4 bit/子帧 共 8 bit
固定码书增益	4 bit/子帧 共 8 bit
总计	80 bit

3.1 线性预测分析与线谱频率参数的量化

线性预测分析每帧进行一次。窗口的形状同G.729一样，但是没有引入前向延时。计算得到的线性预测系数被转化到

线谱频率参数(LSF)域进行量化。在本编码器中,我们使用一种一步插值预测式矢量量化器来量化线谱频率参数^[6,9]。

3.2 感觉加权

感觉加权基于未量化的 LP 滤波器系数,它由下式给出:

$$W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)} \quad (2)$$

同 ITU-T 的 G.729 标准一样,感觉加权滤波器的频率响应由两个参数 γ_1 和 γ_2 控制,它们依赖于信号谱的形状自适应变化。

3.3 散布脉冲滤波器

同原来的 DP-CELP 语音编码器一样,固定码书使用了散布脉冲滤波器的技术。散布脉冲矢量由截止频率为 3.4kHz 的 FIR 滤波器的冲激响应构成。该矢量用来散开信号脉冲的能量以使 ACELP 编码器的性能得到改善。

4 实验结果

为了评测 4kbit/s CELP 语音编码算法的性能,我们进行了大量的主观 A/B 测试。测试语句由 16 个中文语句组成,包括 8 句女性讲话者的语音(F1-F8)和 8 句男性讲话者的语音(M1-M8)。主观测试结果显示大多数语句的感性质量都有改善。

此外,我们还根据 ITU-T 的 P.862 标准^[10]进行了客观质量测试。该标准产生于 2001 年,称作 PESQ(Perceptual Evaluation of Speech Quality)法,是一种客观测试方法,它适用于窄带电话网络中的端到端的语音质量评测。PESQ 的测试结果在表 6 中给出。可以看到,改进的 4 kb/s DP-CELP 的

表 6 PESQ 实验结果

Tab.6 PESQ experiment result

测试文件	MOS (旧)	MOS (新)	测试文件	MOS (旧)	MOS (新)
F1	3.048	3.168	M1	2.774	2.812
F2	2.943	3.155	M2	2.867	3.013
F3	2.708	2.863	M3	3.141	3.161
F4	3.304	3.299	M4	3.073	3.112
F5	3.024	3.142	M5	2.926	2.969
F6	3.127	3.148	M6	3.128	3.279
F7	2.624	2.895	M7	2.965	3.084
F8	2.909	2.977	M8	3.183	3.217
Avg.	2.961	3.081	Avg.	3.007	3.081

性能优于原先的编码器,而且对女性讲话者的改善更为明显一些。

5 结束语

本文提出了 ACELP 语音编码器中的非均匀和部分搜索域代数码书搜索算法,并将其应用到一种 4 kb/s DP-CELP 语音编码算法中。主观和客观测试结果均表明:与原算法相比,重建语音的质量得到了改善。

参考文献

- [1] ITU-T Recommendation G.729. Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP), 1996.
- [2] ITU-T Recommendation G.723.1. Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s, 1996.
- [3] Yasunaga K, *et al.*. Dispersed-pulse codebook and its application to a 4kb/s speech coder. *IEEE Proc, ICASSP, 2000, Istanbul, Turkey, III*: 1503-1506.
- [4] Gao Y, *et al.*. eX-CELP: A speech coding paradigm. *IEEE Proc, ICASSP, 2001, Salt Lake City, Utah, II*: 689-692.
- [5] Rao A V, Ahmadi S, *et al.*. Pitch adaptive windows for improved excitation coding in low-rate CELP coders. *IEEE Trans. on Speech Audio Processing*, 2003, 11(6): 648-659.
- [6] 鲍长春. 高质量的 4kb/s 散布脉冲 CELP 语音编码算法. *电子学报*, 2003, 31(2): 309-313.
- [7] 李悦,唐昆等. 高质量 3.35kb/s MPD-USACELP 语音编码算法研究. *清华大学学报(自然科学版)*, 2004, 44(10): 1410-1413.
- [8] Chu W C. *Speech coding algorithms—Foundation and evolution of standardized coders*. New Jersey: Wiley-Interscience, 2003: 471-474.
- [9] Bao Changchun. Harmonic excited LPC (HE-LPC) speech coding at 2.3kb/s. *IEEE Proc. ICASSP, 2003, Hongkong, I*: 784-787.
- [10] ITU-T Recommendation P.862. *Perceptual evaluation of speech quality (PESQ)*, 2001.

白燕宁: 男, 1978 年生, 硕士生, 研究兴趣为语音信号处理、低比特率数字语音编码、DSP 开发。

鲍长春: 男, 1965 年生, 教授, 博士生导师, 主要从事低比特率语音编码、语音增强与听觉建模、语音合成、基于 IP 分组网络的信源编码、语音传输、大容量语音存储等方面的研究工作。