

## ◎数据库、信息处理◎

## 具有后向基音检测的 8 Kbit/s LD-aCELP 语音编码

武淑红<sup>1,2</sup>, 张刚<sup>1</sup>, 赵哲峰<sup>1</sup>WU Shu-hong<sup>1,2</sup>, ZHANG Gang<sup>1</sup>, ZHAO Zhe-feng<sup>1</sup>

1. 太原理工大学 信息工程学院, 太原 030024

2. 太原理工大学 计算机与软件学院, 太原 030024

1. College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China

2. College of Computer and Software, Taiyuan University of Technology, Taiyuan 030024, China

E-mail: shh\_wu@163.com

WU Shu-hong, ZHANG Gang, ZHAO Zhe-feng. 8 Kbit/s LD-aCELP speech coding with backward pitch detection. *Computer Engineering and Applications*, 2009, 45(17): 119-121.

**Abstract:** This paper makes improvement in G.728 algorithm aiming at reducing coding rate and presents an 8 Kbit/s speech coding algorithm whose delay is 2.5 ms. Adaptive codebook structure is introduced in the proposed algorithm, which is composed of the recent excitation. The algorithm uses double codebook structure which contains adaptive codebook and normalized fixed codebook. The exact value of gain is calculated and quantized. Fixed quantization is used in adaptive codebook and adaptive quantization is used in fixed codebook. For codebook searching, the algorithm detects backward pitch, and then searches the adaptive codebook subtly round the pitch period  $T$ . After searching 64 adaptive code words, 8 gains, 256 fixed code words and 8 gains, the best excitation is obtained. The algorithm uses 20 bit per framework. Testing with the average segment SNR and Perceptual Evaluation of Speech Quality (PESQ), the improved algorithm is close to G.728 in speech coding quality.

**Key words:** Adaptive codebook; Pitch detection; Low Delay Code Excited Linear Prediction (LD-CELP); Perceptual Evaluation of Speech Quality (PESQ)

**摘要:** 以降低码率为目的对 G.728 算法进行改进, 提出了一个延迟为 2.5 ms 的 8 Kbit/s 的语音编码算法。算法引入了由最近的历史激励构成的自适应码书和归一化的固定码书的双码书结构。计算增益真值并量化, 增益量化时对自适应码书用固定量化, 固定码书用自适应量化。码书搜索时先进行后向基音检测, 在基音周期  $T$  附近对自适应码书进行精细搜索。搜索 64 个自适应码矢、256 个固定码矢和各自 8 个增益值获得最佳激励, 每帧耗费 20 bit。用平均分段信噪比和感知语音质量评价 (PESQ) 测试, 改进算法编码质量接近于 G.728。

**关键词:** 自适应码书; 基音检测; 低延迟码激励线性预测; 感知语音质量评价

**DOI:** 10.3778/j.issn.1002-8331.2009.17.036 **文章编号:** 1002-8331(2009)17-0119-03 **文献标识码:** A **中图分类号:** TN911.7

## 1 引言

低码率、低复杂度的 LD-CELP 语音编码算法在通信领域具有十分重要的意义。目前 G.728 算法是 ITU 颁布的唯一一种 LD-CELP 国际标准<sup>[1]</sup>, 虽然它的延迟很小 (0.625 ms), 但编码速率过高 (16 Kbit/s)。ITU 的另一个编码标准 G.729 速率达到 8 Kbit/s<sup>[2]</sup>, 但延迟为 10 ms。以上两种编码算法均不满足 ITU 对 8 Kbit/s 语音编码算法的要求<sup>[3]</sup>。

本文提出了一种码率为 8 Kbit/s 的低延迟 LD-CELP 语音编码算法, 以 G.728 结构为基础将激励矢量长度由 5 样点提高到 20 样点, 参数更新的帧长与 G.728 相同, 因此算法保留了 G.728 大多数特征。一个重要区别是引入了自适应 (adaptive) 码

书结构和后向基音搜索, 因此将算法记作 LD-aCELP。下面先给出 LD-aCELP 算法基本原理, 然后介绍基于后向基音检测的码书搜索算法, 最后给出实验结果。实验结果表明引入自适应码书以及后向基音检测后, 编码速率由 G.728 的 16 Kbit/s 降低至 8 Kbit/s, 码率资源降低 50%。与标准 728 相比, SNR 下降 4.50%, PESQ 仅降低 1.19%, 编码质量基本相当。

## 2 算法原理

### 2.1 LD-aCELP 原理

算法使用自适应码书  $u_n$  和固定码书  $y_n$  两个码书搜索最佳激励矢量。

**基金项目:** 国家自然科学基金 (the National Natural Science Foundation of China under Grant No.60372058, No.60772101)。

**作者简介:** 武淑红 (1969-), 女, 博士研究生, 副教授, 主要研究领域为语音编码、计算机应用; 张刚 (1953-), 男, 博士, 教授, 博士生导师, 主要研究领域为语音编码与嵌入式系统; 赵哲峰 (1975-), 男, 博士研究生, 主要研究领域为语音编码与嵌入式系统。

**收稿日期:** 2009-02-17 **修回日期:** 2009-03-18

$$u_s = [u_{s,0}, u_{s,1}, \dots, u_{s,19}] \quad s=0, 1, \dots, N \quad (1)$$

$$y_j = [y_{j,0}, y_{j,1}, \dots, y_{j,19}] \quad j=0, 1, \dots, M \quad (2)$$

将 20 个连续的语音样点组成的输入语音矢量  $s(n)$  通过感觉加权滤波器, 得到加权的语音矢量  $v(n)$ , 减去零输入响应  $r(n)$  后得到用于自适应码书搜索的目标矢量  $x(n)$ :

$$x(n) = v(n) - r(n) \quad (3)$$

用  $\hat{G}_i, \hat{g}_k$  分别表示自适应码书增益和固定码书增益的量化值, 加权的综合滤波器的冲激响应矩阵为  $H(Z)$ 。对历史的激励信号序列先进行后向基音检测, 搜索基音周期  $T$ , 然后在  $T$  附近较小的范围内对自适应码书进行更精细的搜索, 具体搜索方法后面介绍。

对于搜索范围内的每一个自适应码矢  $u_s$ , 将  $\hat{G}_i u_s$  作为当前合成滤波器的自适应激励矢量。目标矢量信号  $x(n)$  减去已滤波的自适应激励的贡献  $\hat{G}_i H u_s$  得到用于固定码书搜索的新的目标矢量  $x'(n)^{[4-5]}$ 。

$$x'(n) = \frac{[x(n) - \hat{G}_i H u_s]}{\sigma(n)} \quad (4)$$

其中  $\sigma(n)$  是用 G.728 相同方法得到的后向增益预测值。用新的目标矢量  $x'(n)$  搜索固定码书  $y_j$  及其增益  $\hat{g}_k$  直到求出最佳自适应激励  $\hat{G}_i u_s$  和最佳固定码书激励  $\sigma(n) \hat{g}_k y_j$ , 得到综合滤波的最佳激励<sup>[6]</sup>:

$$e(n) = \hat{G}_i u_s + \sigma(n) \hat{g}_k y_j \quad (5)$$

每个输入矢量构成一个自适应搜索周期(称为帧), 每帧更新一次 LPC 系数。因为采用后向自适应预测技术<sup>[7]</sup>, 向解码器传送的信息只是激励矢量和增益的索引, 因此只有 20 个样点的缓冲延迟, 对于 8 KHz 的采样率就是 2.5 ms 的延迟。

## 2.2 自适应码书

自适应码书采用 147 样点的存储移位寄存器(SSR, Storage Shift Register)来存储码字, 相邻码字有叠接 1 的特征, 码字长度  $L=20$ 。开始时, SSR 中所有存储单元全为零。自适应码书每帧刷新一次, 在完成码书搜索的情况下, 当前帧搜索到最佳激励  $e(n)$ , 一方面作为合成线性预测滤波器的输入, 一方面也是下一个进入 SSR 中的元素。每次将  $e(n)$  推入 SSR 中, SSR 原有的内容依次向上推<sup>[8]</sup>。

## 3 搜索算法

### 3.1 增益的量化

训练数据采用中科院声学所的汉语语音数据文件, 电话信道采集的含噪数据。共 1 600 句话, 其中男、女声各 800 句, 发声 417 分钟, 约 200 万个矢量。对训练数据进行计算得到自适应码书及固定码书各自的增益精确值  $G$  与  $g$ 。用固定量化方法量化自适应码书增益, 采用叠代法<sup>[9]</sup>计算最佳量化特性。固定码书增益采取自适应量化方法<sup>[10]</sup>。分别将它们量化成 3 bit(符号位占 1 bit), 量化结果如表 1 所示。

### 3.2 码书搜索

自适应码书搜索范围是在后向基音值  $T$  附近搜索 64 个矢量(每帧 6 bit), 固定码书尺寸为 256(每帧 8 bit), 两个码书增益均量化为 3 bit, 总共 20 bit, 码率为 8 Kbit/s。基音周期反映了语音激励源的重要特征, 自适应码书较好地保留了信号的周期性, 准确地进行实时基音周期检测对于高质量语音编码具有重

表 1 LD-aCELP 算法增益的量化值

自适应码书			固定码书		
增益输入范围	索引	量化输出	增益输入范围	索引	量化输出
$(-\infty, -2.101\ 28]$	0	-2.738 20	$(-\infty, -14.854\ 0]$	0	-18.335 6
$(-2.101\ 28, -1.152\ 73]$	1	-1.464 36	$(-14.854\ 0, -9.183\ 85]$	1	-11.372 5
$(-1.152\ 73, -0.625\ 28]$	2	-0.841 10	$(-9.183\ 85, -5.368\ 89]$	2	-6.995 18
$(-0.625\ 28, 0.017\ 33]$	3	-0.409 46	$(-5.368\ 89, 0.002\ 20]$	3	-3.742 59
$(0.017\ 33, 0.648\ 20]$	4	0.444 12	$(0.002\ 20, 5.368\ 89]$	4	3.742 59
$(0.648\ 20, 1.095\ 68]$	5	0.852 28	$(5.368\ 89, 9.183\ 85]$	5	6.995 18
$(1.095\ 68, 1.996\ 59]$	6	1.339 08	$(9.183\ 85, 14.854\ 0]$	6	11.372 5
$(1.996\ 59, +\infty)$	7	2.654 09	$(14.854\ 0, +\infty)$	7	18.335 6

要意义。算法对激励信号进行的后向基音检测<sup>[11]</sup>, 准确性不能得到保证, 但用来确定自适应码书的搜索范围, 可以在较小的范围内对自适应码书进行更精细的搜索, 减少自适应码书的比特数, 将节省的比特用于扩大固定码书尺寸, 提高编码质量。

基音检测模块采用对历史的激励信号进行基音检测, 每 20 个采样点搜索一次基音周期。在搜索自适应码书时:

- (1) 若搜索到新的基音周期  $T$ , 则以  $T$  为基准;
- (2) 若没有搜索到新的基音周期, 则以上一次搜索到的基音周期为基准。

由 2.2 可知, 自适应码书采用 147 样点存储移位寄存器(SSR)来存储码字, 相邻码字有叠接 1 的特征。在确定了用于自适应码书搜索的基音周期  $T$  后, 自适应码书的搜索方案是, 从 SSR 的右端向左移动一个周期  $T$ , 然后在该位置两边各搜索 32 个矢量, 因为 SSR 中相邻码字有叠接 1 的特征, 所以相当于在自适应码书中搜索 64 个码字。

在 SSR 左移一个周期位置后, 按照以下三种情况确定所搜索的 64 个码字:

- (1) 当  $T \leq 51$  时, 搜索 SSR 中最左边的 64 个码字;
- (2) 当  $T \geq 115$  时, 搜索 SSR 中最右边的 64 个码字;
- (3) 其他情况, 在  $[146 - T - 32, 146 - T + 31]$  之间搜索。

LD-aCELP 按照上述范围对 64 个自适应码矢  $u_s$ 、8 个增益  $\hat{G}_i$ 、256 个固定码矢  $y_j$ 、8 个增益  $\hat{g}_k$  的组合进行搜索。对每帧语音信号, 首先对每个  $u_s$  和  $\hat{G}_i$  的组合计算用于固定码书搜索的目标矢量, 然后对每个  $y_j$  和  $\hat{g}_k$  的组合进行搜索, 确定索引  $i, s, k$  和  $j$  使下式均方误差  $D$  最小。

$$D = \|x(n) - \hat{G}_i H u_s - \sigma(n) \hat{g}_k H y_j\|^2 \quad (6)$$

## 4 实验结果与讨论

### 4.1 测试与评价标准

语音测试评价标准采用平均分段信噪比 SNR 以及 ITU P. 862 的 PESQ<sup>[12]</sup>进行测试评价。PESQ 算法得到的评测结果与主观评价指标 MOS 分的相关度可以达到平均 0.935, 是现有的基于心理声学模型的客观评价算法中效果最好的指标<sup>[13]</sup>。

### 4.2 实验结果与讨论

实验室独立采集了 30 句语音为实验数据, 男、女声各 15 句, 共有 15 489 个矢量, 实验数据不包括在训练数据中。图 1 为一句原始语音与 G.728、LD-aCELP 算法合成语音波形图, 同 G.728 编码效果比较, 视听效果无明显差别。

表 2 是 LD-aCELP 算法的平均分段 SNR 和 PESQ 值与 G.728 的比较。LD-aCELP 算法平均分段 SNR 达到 18.013 8 dB, PESQ 值为 3.575; G.728 的分段 SNR 为 18.861 7 dB, PESQ 值

表 2 LD-aCELP 算法与 G.728 测试比较

实验语句	平均分段 SNR/dB		PESQ		实验语句	平均分段 SNR/dB		PESQ	
	LD-aCELP	G.728	LD-aCELP	G.728		LD-aCELP	G.728	LD-aCELP	G.728
PCM90SL.hex	17.477 5	19.443 8	3.611	3.661	PCF10SL.hex	20.033 5	21.178 8	3.559	3.544
PCM80SL.hex	17.896 2	19.756 7	3.653	3.864	PCF0.hex	20.277 5	20.504 9	3.727	3.558
PCM70SL.hex	20.200 5	20.839 1	3.774	3.658	JOE10.hex	15.617 8	16.019 0	3.427	3.524
PCM60SL.hex	15.221 4	16.293 1	3.856	3.971	JOE9.hex	17.614 2	17.452 5	3.407	3.466
PCM50SL.hex	18.636 0	20.531 2	3.638	3.740	JOE8.hex	19.675 1	18.889 0	3.662	3.566
PCM40SL.hex	17.288 0	19.247 3	3.603	3.777	JOE7.hex	19.877 2	19.793 0	3.572	3.449
PCM30SL.hex	19.528 3	21.028 6	3.745	3.899	JOE6.hex	16.138 7	16.496 7	3.586	3.605
PCM20SL.hex	17.228 6	19.367 6	3.831	3.905	JOE4.hex	17.736 1	17.969 3	3.585	3.710
PCM10SL.hex	17.401 4	19.559 5	3.811	4.019	JOE3.hex	17.574 3	17.685 2	3.562	3.587
PCM100SL.hex	17.867 7	19.616 9	3.757	3.996	JOE2.hex	16.346 8	17.151 8	3.386	3.409
PCF90SL.hex	17.529 1	18.273 2	3.285	3.311	JOE1.hex	19.502 4	19.732 8	3.448	3.479
PCF70SL.hex	20.714 4	22.015 6	3.236	3.216	LGJ00S.hex	16.364 4	17.348 8	3.735	3.638
PCF50SL.hex	19.817 2	20.296 7	3.378	3.379	LGJ01S.hex	16.111 8	16.606 1	3.739	3.672
PCF40SL.hex	17.442 3	19.269 1	3.169	3.399	ZHG2.hex	16.551 7	16.018 2	3.509	3.430
PCF30SL.hex	19.227 2	20.448 6	3.629	3.610	平均值	18.013 8	18.861 7	3.575	3.618
PCF20SL.hex	19.187 3	20.628 4	3.363	3.490					

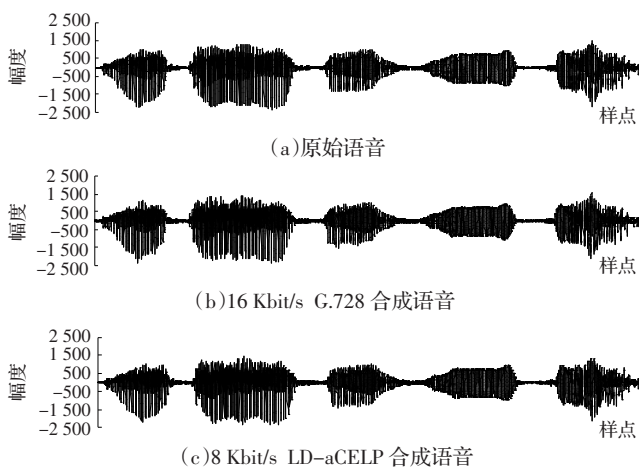


图 1 LD-aCELP 算法与 G.728 合成语音波形

为 3.618。LD-aCELP 算法的 SNR 值比 G.728 略有降低,但对 PESQ 值,LD-aCELP 算法平均 PESQ 值仅比 G.728 低 0.043,而且有 33.3%的语句比 G.728 都是有增加的。

实验结果表明在满足 2.5 ms 低延迟的要求下,编码速率由 G.728 的 16 Kbit/s 降低至 8 Kbit/s,码率降低 50%,平均分段 SNR 降低了 4.50%,PESQ 仅降低了 1.19%,LD-aCELP 算法的编码质量接近标准 G.728。

## 参考文献:

[1] CCITT.Recommendation G.728,Coding of Speech at 16kbit/s Using

Low-Delay Code Excited Linear Prediction[S].Geneva,1992.

- [2] ITU-Recommendation G72 Coding of Speech at 8Kbit CS-ACELP[S].1996.
- [3] 皇甫丽英.G.728 算法改进及低码率 LD-CELP 算法研究[D].太原:太原理工大学,2002.
- [4] 范睿,鲍长春,李锐.基于 ACELP 的嵌入式语音编码算法[J].通信学报,2007,28(10):48-54.
- [5] 鲍长春.数字语音编码原理[M].西安:西安电子科技大学出版社,2007:175-176.
- [6] 王艳,黄建国,李钊.一种用于语音编码的快速自适应码书搜索算法[J].计算机工程与应用,2007,43(15):69-71.
- [7] 张刚.甚低码率低延迟语音编码算法研究[D].太原:太原理工大学,2007.
- [8] 郑建华.基于 CELP 低速率语音编码算法的研究及实现[D].南宁:广西师范大学,2008.
- [9] 张起贵,张刚.低延迟码激励语音编码算法的最佳增益滤波器[J].电子与信息学报,2006,28(8):1533-1536.
- [10] 张刚,谢克明,张雪英,等.低延迟码激励语音编码算法的增益优化与混合窗设计[J].电子与信息学报,2005,27(6):892-895.
- [11] 李晋声,倪维桢,赵振纲.16 kbit/s LD-CELP 算法的再研究[J].北京邮电大学学报,1995,18(4):43-50.
- [12] ITU T Rec.P 862 Perceptual Evaluation of Speech Quality(PESQ)-an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs[S].Geneva,Switzerland,2001.
- [13] 张军,张德运,傅鹏.一种改进的心理声学语音质量客观评价算法[J].微电子学与计算机,2007,24(3):203-206.