

文章编号: 1673- 0062(2010) 02- 0047- 05

ACELP 语音编码算法改进与仿真

刘 华, 欧阳宏志, 左望霞

(南华大学 电气工程学院, 湖南 衡阳 421001)

摘要: 代数码激励线性预测算法 (ACELP) 是目前诸多低速率语音编码标准的算法核心, 包括 3G 语音标准 VSELP、AMR、AMR-NB、AMR-WB。该算法基于码激励线性预测模型, 通过对码本的有效搜索, 确定基音延迟, 算法时间复杂度为 $O(n^3)$ 。本文在 ACELP 算法基础上, 对 ACELP 中自适应码本搜索过程进行改进, 提出 E-ACELP 算法。通过 AMR 标准中 8 种速率情况的仿真, E-ACELP 算法码本搜索时间减少、时间复杂度下降。基于 E-ACELP 算法, 语音编码标准性能和效率得到提高。

关键词: 代数码激励线性预测; 自适应码本搜索; 时间复杂度

中图分类号: TN 911. 7 **文献标识码:** A

Improvement and Simulation for the ACELP Speech Encoding Algorithm

LIU Hua, OUYANG Hong-zhi, ZUO Wang-xia

(School of Electrical Engineering University of South China Hengyang Hunan 421001, China)

Abstract The algebra code-excited linear prediction (ACELP) is the core algorithm of a lot of low bit rate speech coding standards including the 3G speech standard VSELP, AMR, AMR-NB, AMR-WB. The algorithm is based on code excited linear prediction model. The pitch delay is determined by the effective codebook search. The algorithm's time complexity is $O(n^3)$. Based on the ACELP, this paper proposes an improved algorithm called E-ACELP. E-ACELP reduces the time of codebook search and time complexity through the simulation of eight ratios for AMR. The performance and efficiency of the standards based on ACELP are improved.

Key words algebra code-excited linear prediction; search for adaptive codebook; time complexity

收稿日期: 2010- 01- 17

基金项目: 湖南省教育厅基金资助项目 (09C855)

作者简介: 刘 华 (1979-), 男, 湖南衡阳人, 南华大学电气工程学院讲师, 硕士。主要研究方向: 语音编码及智能控制。

0 引言

目前诸多低速率语音编码标准由于其编码速率低、占用信道资源少、速率转换灵活的优点,得到了广泛应用^[1]. 包括 3G 语音标准 VSELP、AMR、AMR-NB、AMR-WB 中的大部分编码速率均属于低速率语音编码^[2]. 现代低速率语音编码的算法体制主要沿着 4 个方向发展,即码激励线性预测模型、混合激励线性预测模型、正弦激励模型和插值编码模型^[3-4]. CELP 和 ACELP 算法模型是使用最普遍也是最优的. 两种模型源于码激励线性预测模型,同时又是对模型的改进. CELP 是语音标准 FS1016 的算法核心, ACELP 则是 G. 723. 1、G. 729、G. 729A、GSM 06. 60、AMR、AMR-NB、AMR-WB 等语音编码标准的算法核心^[5-7].

代数码激励线性预测算法尽管使用普遍,相对基于波形编码的 PCM 和 ADPCM 语音标准,算法整个过程非常复杂^[8]. ACELP 算法核心是对基音延迟的确定,基音延迟的确定最终依靠自适应码本的有效和快速搜索. 自适应码本搜索占用了 ACELP 的绝大部分执行时间^[9].

1 语音编解码的数学极限

语音编解码体现变换域的思想,从模拟域到数字域再到模拟域;从时域到频域再到时域. 从语音的角度,最基本的单位是音素,世界上语音音素在 128~256 个,取上限 256 个. 根据世界不同语言发音的速度,设定每秒平均发出 10 个音素,以码元 0 和 1 进行编码,依据信息论编码理论,编码后的比特速率为

$$I = \log_2(256)^{10} = 80 \text{ bps} \quad (1)$$

从语言的角度,以英语发音为例,发音的物理意义是以一定速率的语音来发送信息. 参照 ASCII 编码,把每个字符用 7 比特编码,设定每分钟发音可包含英语单词 125~600 个,每个英语单词平均由 7 个字母够成. 此时基于 2 进制的编码,编码后的比特速率为

$$I = 7 \times 7 \times \frac{600}{60} = 490 \text{ bps} \quad (2)$$

由式(1)及(2),可以认为语音编码的极限速率为 80~490 bps. 现在的 G. 722. 2 语音编码的最低比特速率为 6. 6 kbps. 由此可见,工程实际中的编码速率距离理论极限还有很大的空间,低速率语音编码的压缩空间还有很大.

2 自适应码本搜索技术

在语音编码中的混合编码中,可以认为编码器就是一个大型码本. 编码器的主要部分也就是针对大量语音数据的基元选择,包括对自适应码本的搜索. 相应的解码器通过对从码本中选出的基元进行波形拼接合成语音信号. 这样的编码既包含有语音的波形信息又包含语音的参数模型,所以是典型的混合编码.

2.1 自适应码本的作用

基于代数码激励线性预测技术的编码可以分成线性预测编码、自适应码本分析、固定码本分析三部分^[7]. 自适应码本搜索获得自适应码本索引和自适应码本增益. 固定码本搜索获得固定码本索引和固定码本增益^[10].

ACELP 用一个自适应码本中的码字来逼近语音的长时周期性结构即基音;用一个生成后就固定的随机码本中的码字来逼近语音经过短时、长时预测后的余量信号. 搜索出来的 2 个最佳码矢量乘以各自的最佳增益后相加,就得到 ACELP 的激励信号源. 再将激励信号输入 P 阶线性预测综合滤波器,就可以得到合成语音,合成语音与原始语音的误差经过感知加权滤波器得到感知加权误差. ACELP 用感知加权的均方预测误差作为搜索最佳码矢量及其幅度增益的度量准则. 使感知加权误差的平方最小的码矢量即为最佳码矢量. 由于语音信号除了声道共振峰信息以外,在浊音段呈现明显的周期特性,自适应码本搜索就是利用这种周期性来提取语音的基音信息.

2.2 自适应码本搜索过程

CELP 和 ACELP 的码本有一些差异. CELP 的固定码本是预先训练得到并存储好的码表,搜索运算量大较大. ACELP 的代数码本通过直接设计得到,这样算法简单,运算量相对较小,由于是固定数据不需要存储. 但搜索的范围和过程,该算法是可以进一步优化的. CELP 和 ACELP 的码本中码字的数量及维数都有差异. 自适应码本搜索包括了开环基音搜索和闭环基音搜索两部分. 自适应码本搜索是在各个子帧上进行的,经过开环基音搜索和闭环基音搜索之后,再通过基音分数延时处内插过去的激励来计算自适应码本矢量. 自适应码本参数就是基音延时和基音滤波器增益. 自适应码本搜索的目的就是要得到一个最佳的自适应码本索引. 在搜索阶段,LP 残差的扩展激励使闭环搜索简单、快速. 由于多速率编码的算

法核心尽管语音编码速率不同,但算法核心都是一致的,这里主要以目前 3G 中 WCDMA 和 TD-SCDMA 采用的 AMR 标准介绍自适应码本搜索算法。

在 AMR 标准的语音编码算法中,开环基音搜索是基于感知加权语音进行的。开环基音搜索为闭环基音搜索提供大概的范围。开环基音搜索是基于加权语音信号 $s_w(n)$, 该信号不是最初的输入语音信号,而是原始语音输入感知加权滤波器的输出信号。闭环基音搜索是基于线性预测的残差信号,同样需要进行感知加权处理。闭环基音搜索是在开环基音搜索得到的基音延迟值的附近进行。用闭环分析的最小化加权均方误差来决定每一子帧的最佳延时。

2.3 ACELP 算法存在问题

自适应码本的搜索是 ACELP 编码的重要组成部分。搜索的正确与否,搜索的时间消耗将直接影响语音编解码效果、语音质量的好坏。分析 ACELP 基音搜索的过程,基音搜索的范围和搜索过程直接决定算法效率。闭环基音搜索范围针对不同的子帧而不同。第一子帧,基音延时在 $[19\frac{1}{3}, 84\frac{2}{3}]$ 范围,延时分辨率为 $\frac{1}{3}$ 分数延时,在 $[85, 143]$ 范围内是整数延时。在第二、第三、第四子帧,整数基音延时在 $[T_1 - 5, T_1 + 4]$ 范围内,分数基音延时在 $[T_1 - \frac{5}{3}, T_1 + \frac{2}{3}]$ 范围内,分辨率为 $\frac{1}{3}$, 这里 T_1 是前一子帧即第一子帧或第三子帧基音分数延时的最大整数部分,取值为 $[20, 143]$ 。在第一子帧,基音延时范围为 $T_{op} \pm 5$ 取值为 $[20, 143]$, 基音延时用 8 比特进行编码,其他子帧的相关延时用 4 比特进行编码。这样,基音延时既有整数也有分数,明显消耗大量搜索时间。

3 E-ACELP 算法

针对 ACELP 算法中存在的问题,提出 E-ACELP 算法。其过程是:原始语音按帧输入,根据使合成语音与原始语音的加权均方误差最小的准则,从随机码本和固定码本中挑选合适的码矢以代替残差信号,并将码矢地址和增益及各滤波器的参数量化编码后传送到接收端;接收端恢复各滤波器时,采用与发送端相同的码本,按照码矢地址找到该码矢乘上增益,激励合成滤波器,得到合成语音,编码器提取 ACELP 模型参数,解码器再

根据这些参数构成的激励信号合成、重建语音信号。在编码部分需要抽取下列典型参数:线性预测滤波器系数记为 LP,自适应码本和固定码本各自的索引编号以及 2 种码本的增益。

作为对 ACELP 算法的改进, E-ACELP 算法中,基于自适应码本的开环基音搜索在 $K = 20 \sim 143$ 范围内再细分成 3 个部分,实际的搜索只是在 3 个小范围内进行;闭环基音延迟搜索只搜索整数,减少搜索范围。这样相比 ACELP 算法更加高效。

3.1 开环基音搜索

开环基音搜索改进过程如下:

1) $K = 20 \sim 143$ 范围:

$$O_k = \sum_{n=0}^{79} s_w(n) s_w(n-k) \quad (3)$$

2) 在三个范围内搜索各自范围内自相关系数的最大值:

$$\begin{cases} O_{t_1} = \sum_{n=0}^{79} s_w(n) s_w(n-k), k = 20, 21, \dots, 39, i = 3 \\ O_{t_2} = \sum_{n=0}^{79} s_w(n) s_w(n-k), k = 40, 41, \dots, 79, i = 2 \\ O_{t_3} = \sum_{n=0}^{79} s_w(n) s_w(n-k), k = 80, 81, \dots, 143, i = 1 \end{cases} \quad (4)$$

3) 对自相关系数进行归一化为 M_i , 即把三个最大值 $O_{t_i}, i = 1, 2, 3$ 对应式 (4) 中的 k

$$M_i = \frac{O_{t_i}}{\sqrt{\sum_{n=1}^N s_w^2(n-t_i)}} \quad (5)$$

式 (5) 中的 $N = 80$ 对应的 $s_w(n)$ 的范围是上一帧的 $s_w(160 - t_i + 1)$ 开始的 80 个点。3 个归一化的自相关系数及其对应的延迟值表示为: $(M_i, t_i), i = 1, 2, 3$

4) 搜索基音延迟的算法如式 (6):

$$\begin{cases} T_{op} = t_1 \\ M(T_{op}) = M_1 \\ \text{if } M_2 > 0.85M(T_{op}) \\ \quad M(T_{op}) = M_2 \\ \quad T_{op} = t_2 \\ \text{end} \\ \text{if } M_3 > 0.85M(T_{op}) \\ \quad M(T_{op}) = M_3 \\ \quad T_{op} = t_3 \\ \text{end} \end{cases} \quad (6)$$

T_{op} 是变量, 输出结果即开环搜索的基音延迟. 由式(4), 开环基音搜索分为3个范围, 选取小的基音延迟值. 此算法可以防止误将基音周期的整数倍作为基音周期, 减少错误判断为基音高阶整数倍的情况. 这种算法处理类似基音检测时需要的 900 Hz ~ 1 kHz 的低通滤波器.

3.2 闭环基音搜索

闭环基音搜索是基于线性预测的残差信号, 同样需要进行感知加权处理. 闭环基音搜索是在开环基音搜索得到的基音延迟值的附近进行. 这样闭环基音搜索的计算量就降低了.

用闭环分析的最小化加权均方误差来决定每一子帧的最佳延时. 对第一子帧的整数延时, 其上下边界分别为 t_{min} 和 t_{max} 可以通过算法计算得到, 见式(7)即:

$$\begin{cases} t_{min} = T_{op} - 5 \\ \quad \text{if } t_{min} < 18 \text{ then } t_{min} = 18 \\ t_{max} = t_{min} + 10 \\ \quad \text{if } t_{max} > 143 \text{ then } t_{max} = 143 \\ t_{min} = t_{max} - 10 \end{cases} \quad (7)$$

t_{min} 和 t_{max} 作为变量在式(7)、式(8)中被限制了范围. 第二、第三、第四子帧的整数延时其上下边界分别为 t_{min} 和 t_{max} 必须根据 T_1 得到, T_1 是第一子帧的自适应码本搜索最佳码字的整数延时. 式(8)中的 int 是取整符号.

$$\begin{cases} t_{min} = \text{int}(T_1) - 5 \\ \quad \text{if } t_{min} < 18 \text{ then } t_{min} = 18 \\ t_{max} = t_{min} + 9 \\ \quad \text{if } t_{max} > 143 \text{ then } t_{max} = 143 \\ t_{min} = t_{max} - 9 \end{cases} \quad (8)$$

闭环基音搜索准则是使原始语音和重构语音之间均方加权误差最小, 就是使得 $R(k)$ 最大.

$$R_k = \frac{\sum_{n=0}^{39} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{39} y_k(n)y_k(n)}} \quad (9)$$

式(9)中: $x(n)$ 式目标信号; $y_k(n)$ 是在延时 k 的滤波激励.

经过开环和闭环基音搜索两个过程后, 最佳的基音延时得到确定.

4 E-ACELP算法仿真

AMR 标准有8种速率, 组成如表1所示的自适应语音编码器. 如模式 AMR_12 20每帧语音提取出 244 bit 的参数信息, 模式 AMR_10 20每帧语音提取出 204 bit 的参数信息, 模式 AMR_7 95每帧语音提取出 159 bit 的参数信息, 模式 AMR_4 70每帧语音提取 95 bit 信息. 表1中 AMR_SID 是无语音信息传输时的静音编码速率.

表1 AMR 编码器的编码速率

Table 1 Encoding rate of AMR

编码模型	语音速率/每帧比特数	编码模型	语音速率/每帧比特数
AMR_12 20	12 20 kbps/244	AMR_5 90	5 90 kbps/118
AMR_10 20	10 20 kbps/204	AMR_5 15	5 15 kbps/103
AMR_7 95	7 95 kbps/159	AMR_4 75	4 75 kbps/95
AMR_7 40	7 40 kbps/148	AMR_SID	1 80 kbps/36
AMR_6 70	6 70 kbps/134		

MOS方法, 它是由原来 CCITT 建议采用的平均评价得分方法, 可分为五级: 5分表示质量完美 (Excellent); 4分表示高质量 (Good); 3分表示质量尚可 (Fair); 2分表示质量差 (Poor); 1分表示

质量不能接受 (Bad). 通过主观评定方法平均得分意见 MOS (Mean Opinion Score) 判定法, 基于 ACELP 和 E-ACELP 的 AMR 语音编码器各模式的话音质量见表 2

表2 AMR 语音编码器各模式的 MOS 值

Table 2 All AMR modes of MOS

速率 /kbps	12 2	10 2	7 95	7 4	6 7	5 9	5 15	4 75
ACELP MOS 值	4 01	4 06	3 91	3 83	3 77	3 72	3 50	3 50
E-ACELP MOS 值	4 00	4 02	3 78	3 77	3 72	3 70	3 40	3 50

针对 10 段不同长度的语音,使用 C 语言编译平台 Turbo C2.0 及 MATLAB 6.5 用 8 种速率对

语音进行 AMR 语音编码,仿真时间取平均值,见表 3 和表 4

表 3 Turbo C2.0 平台算法仿真时间 (s)

Table 3 Simulation time of algorithm in Turbo C2.0 (s)

速率 /kbps	12.2	10.2	7.95	7.4	6.7	5.9	5.15	4.75
ACELP 算法耗时	15.6	14.7	13.9	13.8	13.7	11.7	12.5	10.8
E-ACELP 算法耗时	9.5	9.0	9.7	9.2	9.3	8.4	8.6	8.5

表 4 MATLAB 6.5 算法仿真时间 (s)

Table 4 Simulation time of algorithm in MATLAB 6.5 (s)

速率 /kbps	12.2	10.2	7.95	7.4	6.7	5.9	5.15	4.75
ACELP 算法耗时	76	78	81	87	67	59	57	48
E-ACELP 算法耗时	68	66	69	71	70	62	63	59

5 结论

本文对低速率语音编码中基于 ACELP 的自适应码本搜索部分进行改进,提高搜索正确性、减少时间消耗.通过 AMR 标准中 8 种速率情况的仿真,E-ACELP 算法与 ACELP 算法编码的语音文件质量基本一致,编码时间消耗下降.

在语音编码领域中,随着传输、处理、存储等各种信息量的巨增,信息的压缩处理已成为迫切的要求,从节省传输频带资源或者保持线路通信的高效率等方面来看,研究基于 ACELP 的各种可变速率语音编码技术对移动通信、互联网中提供丰富稳定的语音业务要求有着重要的工程实际意义.

参考文献:

- [1] Cox R V, Fernaz S, Lamblin C, et al. IU-T coders for wideband, superwideband, and fullband speech communication [J]. Communications Magazine, IEEE, 2009, 47 (10): 106-109
- [2] Neuendorf M, Goumay P, Multus M, et al. Unified speech and audio coding scheme for high quality at low bitrates [C] // IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, 1-4

- [3] 武淑红,张刚,赵哲峰.具有后向基音检测的 8 Kbit/s LD-ACELP 语音编码 [J]. 计算机工程与应用, 2009, 45(17): 119-121.
- [4] 王宏,唐凯.低速率语音编码 [M]. 北京: 国防工业出版社, 2006
- [5] Paolo U sai. 3GPP 3G TS26 090, AMR Speech Codec Transcoding function [DB/OL]. http://www.3gpp.org/ftp/Specs/latest/Rel-9/26_series/26090-900.zip
- [6] Paolo U sai. 3GPP 3G TS26 093, AMR Speech Codec Source Controlled Rate operation [DB/OL]. http://www.3gpp.com/ftp/Specs/latest/Rel-9/26_series/26093-900.zip
- [7] Paolo U sai. 3GPP 3G TS26 094, AMR Speech Codec Voice Activity Detector for AMR Speech Traffic Channels [DB/OL]. http://www.3gpp.com/ftp/Specs/latest/Rel-9/26_series/26094-900.zip
- [8] 贾龙涛,鲍长春.基于 NS 的自适应多速率网络语音通信的性能分析 [J]. 通信学报, 2006, 27(6): 121-125
- [9] 焦传斌,于保华,李治柱.新型宽带语音编解码器 AMR-WB 的研究 [J]. 计算机仿真, 2005, 22(1): 50-52
- [10] 贾龙涛,鲍长春.基于 AMR 的抗分组丢失主次型语音编解码器 [J]. 电子学报, 2005, 33(4): 697-700