

文章编号:1000-582X(2002)01-0080-04

增强型全速率语音编码的原理及实现

钟咏,汪纪锋

(重庆邮电学院通信信息学院,重庆 400065)

摘要:在无线通信系统中,语音压缩编码起着非常重要的作用,因为它在很大程度上决定着合成语音的质量和系统容量。为了提高语音质量,GSM提出了增强型全速率(EFR)语音编码方案。它在LPC声码器的基础上,采用了A-B-S和VQ等技术,编码信息中既包含若干语音特征参量又包括部分波形编码信息。因此,能提供高质量的编码,且比特速率压缩到12.2 kbps,为TD-SCDMA移动通信系统提供了一种可行的语音编码方式。笔者从语音编码的基本概念出发,详细地介绍了EFR语音编码的原理及代数码本搜索实现技术。

关键词:增强型全速率;码本激励线性预测;合成分析;矢量量化

中图分类号:TN 919.3⁺1

文献标识码:A

移动通信系统的带宽是非常宝贵的资源,为了获得有效的频谱利用,容纳更多的用户,低比特率的语音编码无疑是至关重要的。如何在给定编码速率的条件下,得到尽量好的重建语音质量,或称编码质量,同时尽量减少编解码延时以及算法的复杂程度,一直是语音编码研究的重要问题。换一个角度也可以说,语音编码研究的基本问题,就是在给定编码质量、编解码延时及算法复杂程度的条件下,如何降低语音编码所需的比特率。

1 概述

自从30年代末提出脉冲编码调制(PCM)原理及声码器概念后,语音信号编码一直沿着这两个方向发展,它们也称为语音信号的波形编码和参数编码。

波形编码以再现波形为目的,这种方式能保持较好的语音质量,硬件上也容易实现,但比特率较高,而且时延较大。参数编码是将人类语言信息用特定的声源模型表示,其比特率很低,最低可压缩到2 kbit/s左右,但自然度差,语音质量难以提高。尤其在背景噪声较大的环境下,声码器不能正常工作。

混合编码是最适合于数字移动通信系统的编码方式。混合编码把波形编码的高质量 and 参量编码的高压缩性融为一体,编码信息中既包括若干语音特征参

量又包括部分波形编码信息,可在8~16 kb/s的范围内达到良好的语音质量,比特率最终可望降为4 kb/s。数字蜂窝移动通信系统中实用的语音编码技术均采用混合编码。采用的激励源不同,就构成不同的编码方案。泛欧数字蜂窝网(GSM)对数字语音业务信道规定了三种标准,如表1所示。

表1 GSM的语音标准

标准	编码方案	速率(Kb/s)
全速率	RPE-LTP	13
增强型全速率	ACELP	12.2
半速率	CELP	6.5

2 增强型全速率语音编码的原理分析

从表1可以看出,除了全速率以外,其它两种的编码方案都是基于码本激励线性预测CELP技术的。CELP的基本思想是,在编解码器中事先存放相同的一组矢量称为码本,编码时从码本中选择合适的码矢量来取代残差信号,再将该矢量的地址作为编码参数传送到接收端。在接收端,由于采用的是同一码本,只要根据发端传送来的地址就能找到相应的码矢量,再送至长时和短时滤波器就能重建语音。由于传送的是地址号而非矢量本身,因此大大降低了编码速率。

在分析增强型全速率语音编码的原理之前,先分析

• 收稿日期:2001-06-28

作者简介:钟咏(1974-),女,四川人,重庆邮电学院硕士研究生,研究方向:语音编码技术。

一下语音的生成模型,如图1所示。

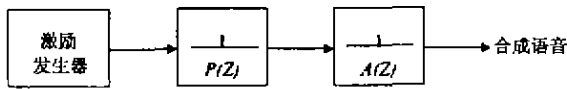


图1 语音生成模型图

如果用预测余量信号作为激励信号源,输入长时预测综合滤波器 $1/P(Z)$,再将其作为短时预测综合滤波器 $1/A(Z)$ 的输入,即可在其输出端得到合成语音信号。

2.1 短时预测

语音信号的短时相关性(谱包络)可以由一个全极点模型来描述,这个模型可称为线性预测(LP)滤波器或短时合成滤波器。它的传输函数 $H(z)$ 为:

$$H(z) = \frac{1}{A(Z)} = \frac{1}{1 + \sum_{i=1}^m a_i z^{-i}}$$

式中, a_i 是量化后的 LP 参数, $i = 1 \sim m$; m 是线性预测滤波器的阶数,对于 8 kHz 采样率, m 的典型值为 8 ~ 12 之间。

2.2 长时预测

语音信号的长时相关性(谱的精细结构)可以用基音合成滤波器 $1/P(z)$ 来表示:

$$\frac{1}{P(Z)} = \frac{1}{1 - g_p Z^{-T}}$$

式中 T 为基音延时, g_p 为基音增益。

2.3 CELP 编码说明

CELP 编码基于合成分析(A-B-S)的搜索过程、矢量量化(VQ)和线性预测(LP)技术,其编码示意图如图2所示。在这个模型中,短时 LP 合成滤波器的输入激励信号是由两个选自自适应和固定(创新)码本的激励矢量乘以最佳增益,相加构成。语音的合成就是使这两个选自这些码本的恰当的码本激励矢量通过短时合成滤波器实现的。

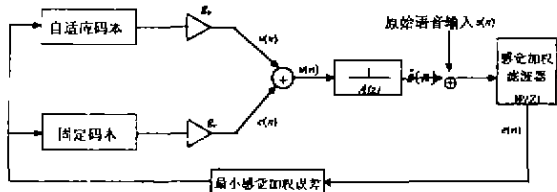


图2 CELP 编码示意图

码本中的最佳激励序列是通过合成分析法

(A-B-S) 搜索进程来选择的。这种方法可以使原始信号和合成信号之间的误差最小,相应的感觉加权误差测度最小。

2.4 感觉加权滤波器

感觉加权滤波器的依据是人耳的掩蔽效应。它的作用是使实际误差信号的谱不再平坦,而是有着与语音信号谱具有相似的包络形状。这就使误差度量的优化过程与感觉上的共振峰对误差的掩蔽效应相吻合,产生较好的主观听觉效果。

EFR 中的感觉加权滤波器由下式给出:

$$W(Z) = \frac{A(Z/\gamma_1)}{A(Z/\gamma_2)}$$

式中, $A(Z)$ 是未经量化的 LP 滤波器, $0 < \gamma_2 < \gamma_1 \leq 1$ 是感觉加权因子,对应的值分别为 $\gamma_1 = 0.9, \gamma_2 = 0.6$ 。感觉加权滤波器使用未经量化的 LP 参数,而共振峰滤波器使用量化的 LP 参数。

3 EFR 语音编码的 CELP 码本搜索算法实现

CELP 编码的关键在于码本搜索,包括对自适应码本和创新码本的搜索。EFR 编码器的创新码本使用代数码本结构,码本容量小,且结构合理,码本搜索按 CELP 创新码本搜索算法进行。以下详细介绍创新码本搜索的算法实现。

3.1 代数码本的结构

代数码本结构是基于交错单脉冲排列(ISPP)设计的。在这个码本中,创新码本包括 10 个非零脉冲,幅度均为 ± 1 。子帧的 40 个位置分为五组,每组中两个脉冲,如表 3 所示。

表3 代数码本中各脉冲的潜在位置

组	脉冲	位置
1	i_0, i_5	0, 5, 10, 15, 20, 25, 30, 35
2	i_1, i_6	1, 6, 11, 16, 21, 26, 31, 36
3	i_2, i_7	2, 7, 12, 17, 22, 27, 32, 37
4	i_3, i_8	3, 8, 13, 18, 23, 28, 33, 38
5	i_4, i_9	4, 9, 14, 19, 24, 29, 34, 39

同一组中的两个脉冲编为 6 bit(每个脉冲的位置编为 3 bit), 每组中的第一个脉冲编为 1 bit(总计 5 bit)。对于同一组中的两个脉冲,仅需要 1 个符号比特,这个比特表示第一个脉冲的符号,第二个脉冲的符号根据与第一个脉冲的相对位置来决定。所有的脉冲位

置都用 3 bit 格雷码来表示,以提高抗信道误码的能力。最终的代数码本编为 35 bit。

3.2 代数码本搜索

代数码本搜索是按加权输入语声与合成语声之间的均方误差最小化的原则实现的。用于闭环基音搜索的目标信号通过减少自适应码本贡献来更新:

$$x_2(n) = x(n) - \beta_p y(n), \quad n = 0, \dots, 39 \quad (1)$$

式中 $y(n) = v(n) * h(n)$, 是通过滤波器后的自适应码本矢量, β_p 是量化后的自适应码本增益。如果令 c_k 是第 k 个代数码矢量, 则代数码本可通过使下式达到最大值来搜索:

$$A_k = \frac{(C_k)^2}{E_{dk}} = \frac{(d'c_k)^2}{c_k' \Phi c_k} \quad (2)$$

式中 $d = H'x_2$ 是目标信号 $x_2(n)$ 与冲激响应 $h(n)$ 之间的相关, 而 Φ 是 $h(n)$ 的自相关矩阵。矢量 d 由下式算出:

$$d(n) = \sum_{i=0}^{39} x_2(i)h(i-n), \quad n = 0, \dots, 39 \quad (3)$$

对称矩阵 Φ 的各元素由下式算出:

$$\Phi(i, j) = \sum_{n=i}^{39} h(n-i)h(n-j), \quad (j \geq i) \quad (4)$$

码本的代数结构允许快速的搜索进程, 因此创新码本中仅包含很少的非零脉冲。等式(2)分子部分的相关函数由下式给出:

$$C = \sum_{i=0}^{N_p-1} \theta_i d(m_i) \quad (5)$$

式中 m_i 表示第 i 个脉冲的位置, θ_i 表示第 i 个脉冲的幅度, N_p 是脉冲的个数 ($N_p = 10$)。等式(2)分母部分的能量由下式给出:

$$E_D = \sum_{i=0}^{N_p-1} \phi(m_i, m_i) + 2 \sum_{i=0}^{N_p-2} \sum_{j=i+1}^{N_p-1} \theta_i \theta_j \phi(m_i, m_j) \quad (6)$$

为简化搜索过程, 脉冲幅度用近似的信号量化值预置, 这个信号为 $b(n)$, 它是归一化的矢量 $d(n)$ 和归一化的长时预测余量 $res_{LTP}^{(n)}$ 之和:

$$b(n) = \frac{res_{LTP}(n)}{\sqrt{\sum_{i=0}^{39} res_{LTP}(i)res_{LTP}(i)}} + \frac{d(n)}{\sqrt{\sum_{i=0}^{39} d(i)d(i)}}, \quad n = 0, \dots, 39 \quad (7)$$

这次简化是将某一位置 $b(n)$ 的符号置为此位置的幅度值。简化步骤如下: 第一步令 $s_0(n) =$

$\text{sing}[b(n)], d'(n) = d(n)s_0(n)$; 第二步, 矩阵 Φ 变成包含符号信息的矩阵, 即 $\phi'(i, j) = s_0(i)s_0(j)\phi(i, j)$ 。此时, 式(5)中的相关可以改写为:

$$C = \sum_{i=0}^{N_p-1} d'(m_i) \quad (8)$$

式(6)中的能量可以改写为:

$$E_D = \sum_{i=0}^{N_p-1} \phi'(m_i, m_i) + 2 \sum_{i=0}^{N_p-2} \sum_{j=i+1}^{N_p-1} \phi'(m_i, m_j) \quad (9)$$

如上所述, 预置好脉冲幅度后, 最佳脉冲位置使用高效的、不完备的 A-B-S 搜索技术来确定。

首先, 搜索每组中对应于 $b(n)$ 绝对值最大的位置, 由此可以搜索到所有脉冲的初始位置。第一次脉冲 i_0 的位置总是对应于 $b(n)$ 的最大值。接着, 执行 4 次循环, 每次循环中, 脉冲 i_1 的位置对应每组中 $b(n)$ 的最大值, 其余脉冲按脉冲对 $\{i_2, i_3\}$, $\{i_4, i_5\}$, $\{i_6, i_7\}$, $\{i_8, i_9\}$ 在内循环中顺序探索。每个脉冲有 8 种可能的位置, 总共有 4 次 8×8 -重循环, 因此每次循环有 256 种不同的脉冲位置组合。

在脉冲位置确定后, 将选定的码矢量通过一个自适应预滤波器 $F_E(z)$, 以提高合成语音的质量。滤波器的表达式为: $F_E(z) = 1/(1 - \beta z^{-T})$ 式中 T 为与子帧闭环分数基音延时最接近的整数基音延时, β 是基音增益。

值得注意的是, 在码本搜索之前, 冲激响应 $h(n)$ 必须包括预滤波器 $F_E(z)$, 即 $h(n) = h(n) + \beta h(n-T)$, $n = T, \dots, 39$ 。固定码本增益从下式中可以得出:

$$\beta_c = \frac{x_2' z}{z' z} \quad (10)$$

其中, x_2 是固定码本搜索的目标矢量, z 是固定码本矢量与 $h(n)$ 的卷积, 如式(11)所示:

$$z(n) = \sum_{i=0}^n c(i)h(n-i), \quad n = 0, \dots, 39 \quad (11)$$

4 小结

从以上分析可以看出, EFR 编码器是基于 CELP 技术实现的。它采用代数码本激励线性预测, 在提供语音质量的同时压缩了语音传输速率, 因此, 在无线通信系统中得到了广泛的应用。

参考文献:

- [1] 杨行峻, 迟惠生. 语音信号数字处理[M], 北京: 电子工业出版社, 1998.
- [2] 易克初, 田斌, 付强. 语音信号处理[M], 北京: 国防工业出版社, 2000.
- [3] 程佩青. 数字信号处理教程[M], 北京: 清华大学出版社, 1995.
- [4] 杨留清, 张闽中, 徐菊英. 数字移动通信系统[M], 北京: 人民邮电出版社, 1995.
- [5] THEODORE S RAPPAPORT. 无线通信原理与应用[M], 北京: 人民邮电出版社, 1995.
- [6] MICHEL MOULY, MARIE-BERNADETTE PARTET. GSM 数字移动通信系统[M], 北京: 电子工业出版社, 1996.
- [7] ETSI GSM06. 60 version 7. 0. 1 Release 1998. Digital cellular telecommunications system (Phase 2 +); Enhanced Full Rate (EFR) speech transcoding[S].
- [8] ETSI GSM06. 53 version 7. 0. 0 Release 1998. Digital cellular telecommunications system(Phase 2 +); ANSI-C code for the GSM Enhanced Full Rate(EFR) speech codec[S].
- [9] ETSI GSM06. 81 version 7. 0. 0 Release 1998. Digital cellular telecommunications system (Phase 2 +); Discontinuous Transmission(DTX) for Enhanced Full Rate(EFR) speech traffic channels[S].

Principle and Algorithm Implement of Enhanced Full-rate Speech Coding

Zhong Yong, Wang Ji-feng

(Department of Communication and Information, Institute of Communication & TTC,
Chongqing University, chongqing 400044, china)

Abstract: Speech coders have assumed considerable importance in radio communication systems as their performance, to a large extent, determines the quality of the recovered speech and the capacity of the system. In order to enhance the quality of speech, GSM puts forward the EFR speech coding algorithm. It bases on the LPC coder, the encoding information includes not only speech feature parameters, but also some wave information, due to the application of A-B-S and VQ. It can afford TD-SCDMA a feasible codec algorithm, for its high quality and the 12.2kbps speech rate. This article starting with the principal concepts of speech coding, introduces the principle of speech coding and ACELP algebraic code book search algorithm in detail.

Key words: Enhanced Full Rate (EFR); Codebook-Excited Linear Prediction (CELP); Analysis-By-synthesis (A-B-S); Vector Quantization (VQ)

(责任编辑 吕赛英)

[简讯]

重庆大学学报(自科版)入选“中国期刊方阵”

根据新闻出版总署“建设‘中国期刊方阵’工作方案”的通知[新出版刊(2001)472号文件]精神,科技期刊的评选推荐工作由科技部负责组织。按照期刊方阵入选要求和比例,经部门推荐、专家评审,国家科技部审定并公布结果。重庆大学学报(自科版)已入选“中国期刊方阵”。重庆市这次有15家科技期刊进入期刊方阵,均评为“双效”期刊。