一种可用于 CDMA 移动通信的变速率语音编码算法 [1]

朱 琦 鄢广增

(南京邮电学院通信工程系 南京 210003)

摘 要 该文提出了一种码率为 0.75–5.4kb/s 可变速率的高质量语音编码算法。该算法对 CELP 的激励进行了改进，根据语音的特征把语音分成 4 类，不同类型的语音采用不同的激励码本。特别是对于浊音，提出了一种基于基音同步的嵌入分裂式激励码本，该码本利用浊音具有准周期性的特点，使该算法在很低的码率下就可很好地恢复浊音信号，克服了 CELP 在 4kb/s 速率以下因码本尺寸小而导致合成语音质量差的缺点。经非正式听音测试，它的主观质量超过了 1~8kb/s 的可变速率 QCELP 系统，并且平均速率大约只有 2kb/s，比 QCELP 的 5kb/s 平均速率低了很多，非常适用于 CDMA 移动通信系统。

关键词 语音编码，变速率，矢量量化，基音周期

中图号 TN929.5, TN912.3

1 引 言

随着语音包交换技术和码分多址 (CDMA) 移动通信系统的迅速发展，为了使语音的质量和系统的容量得到合理的配置，最大限度地利用系统的资源，近年来变速率语音压缩编码的研究越来越受到人们的重视 [1−3]。

变速率语音编码可分为两种：网络控制和信源控制。对于网络控制的变速率编码器，它的语音传输速率是由网络根据它的业务流量决定的。例如 ITU G.727 提出的嵌入式可变速率 ADPCM 编码器 [4]，它的速率可有 40kb/s、32kb/s、24kb/s 和 16kb/s，主要用于 ATM 网络。对于信源控制的变速率编码器，它的语音传输速率是由语音的短时统计特性决定的，其中的代表是用于 IS-95CDMA 移动通信系统的可变速率语音编码系统 QCELP[5]。

在 CDMA 系统中所有的用户共用同样的频率，系统的容量受限于用户产生的干扰。而用户产生的干扰和信号的传输速率是密切相关的，速率越低，用户之间的干扰越小，系统的容量就越大。这样的系统就需要有一种先进的信源控制的可变速率编码器，它可以在满足服务质量的前提下尽可能扩大系统容量，Qualcomm 公司提出的基于码激励线性预测 (CELP) 的 QCELP 就是这样一种算法。该编码器虽然有 4 种速率可以选择，但对有声段来说，主要集中在 8kb/s，因此总的平均速率偏高，平均干扰大，系统容量偏小。

根据语音信号的特征，本文提出了一种基于语音帧分类的可变速率 CELP 编码算法 SC-VR-CELP。对不同类型的语音采用不同的激励码本，特别是浊音，根据其具有准周期性的特点，提出了一种基音周期同步的嵌入式激励码本。该码本能很好地代表浊音的激励信号，在速率降到 2.4kb/s 左右时，仍能恢复出高质量的语音。与 QCELP 相比，在提高语音质量的同时，使编码器总的平均速率大大下降。文章第 2 节分析了新算法的基本结构和原理，第 3 节研究了嵌入分裂式激励码本及其搜索算法，第 4 节给出了比特分配及计算机模拟结果。

2 新算法的基本结构和原理

基于语音分类的变速率编码器，首先把输入语音信号分成 4 类：浊音、清音、过渡音和噪声，然后对不同类型的语音采用各自最优的编码算法和最优的比特分配方案。本文采用文献 [6] 所提出的模式识别的分类方法，选用低通能量、过零率、一阶反射系数、前向基音预测增益和

[1] 2000-09-12 收到，2001-08-16 定稿

反向基音预测增益等 6 个参数形成六维的 Euclid 空间，利用 Fisher 分类器进行分类，其准确率可达 98% 左右。

上面所述的 4 种语音，虽然采用不同的编码结构，但其中的基本组成部分和原理是类似的。结构如图 1 所示，分类器控制了 LPC 系数的分析和量化、激励码本的结构与尺寸的选择等。

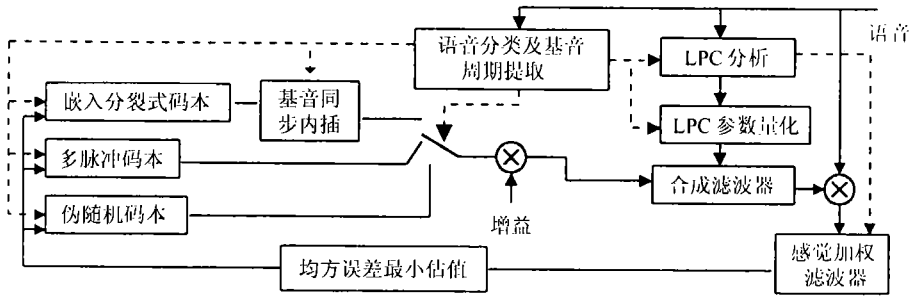


图 1　SC-VR-CELP 工作原理图

对于噪声和清音，由于它们不呈现准周期性，故不需要自适应码本或不需传输有关基音周期的信息，只需要一个伪随机码本即可。由于噪声的相关性较小，故 LPC 预测可降为 6 阶，采用矢量量化算法，每帧只需 14bit，噪声的变化速度亦较慢，所以分析帧长取 40ms，即每隔 40ms 传输一次参数，这样码率可得到进一步压缩。而与噪声相比，清音变化较快，所以帧长取 20ms，LPC 预测阶数取 8。

过渡段语音是非常重要的，它的恢复直接影响到整个编码器的质量。由于过渡音的变化非常快，因此 LPC 分析帧长取 10ms，阶数取 10，也采用矢量量化的算法，具体见第 3 节。由于它是浊音和清音的交界处，其激励信号一部分类似噪声，一部分又呈现一定的准周期性，因此多脉冲码本作为它的激励信号比较合适，由于变化快，故激励码矢量 5ms 更新一次。

语音中有 80% 左右是浊音，因此降低整个编码器的平均速率的关键是如何提高浊音的编码效率。对于一般的 CELP 算法，要想进一步降低码率，就需要减少激励的传输比特，这样激励码本的尺寸就会减小，使得激励码矢量不能很好地代表语音的激励信号，造成合成语音的质量急剧下降。通过对大量短时 LPC 残差信号的观察，我们发现浊音的残差也具有较强的周期性，因此提出了一种基音同步内插的方法对残差信号进行恢复。

如图 2 所示，设一个基音周期内的 LPC 残差信号为 $R_k(n)$，其中 k 为基音序号。由于相邻基音的残差信号相当类似，不需要传输所有的信号，每 r 个基音只传输其中一个基音的信号，而其余 $r-1$ 个基音区间的信号不传输。r 个基音区间构成一帧，每一个基音区间称为子帧，因此该编码器浊音的分析帧长是变化的。由于一帧只需要传输一个子帧的激励，其它子帧的激励可由该子帧恢复得到，所以编码器的速率可以得到大大降低。现在的问题是如何恢复 $r-1$ 个子帧的激励信号，如果简单地重复 $R_k(n)$，则在帧与帧之间会引起跳变。本文采用前后两帧加权平均的方法。

设 $R'_k(n)$ 为 $R_k(n)$ 的重复拓展：

$$R'_k(n)=\begin{cases}R_k(n), & n_k \le n < n_k+p_k\\ R_k(n-p_k), & n_k+p_k \le n < n_k+2p_k\\ 0, & \text{其它}\end{cases} \tag{1}$$

其中 n_k 为第 k 个基音区间内的第一个样点，p_k 表示第 k 个基音周期的长度。同理可得前一帧 $R_{k-r}(n)$ 的重复拓展为 $R'_{k-r}(n)$。现在假设 $r=4$，即每 4 个基音信号保留一个，则令

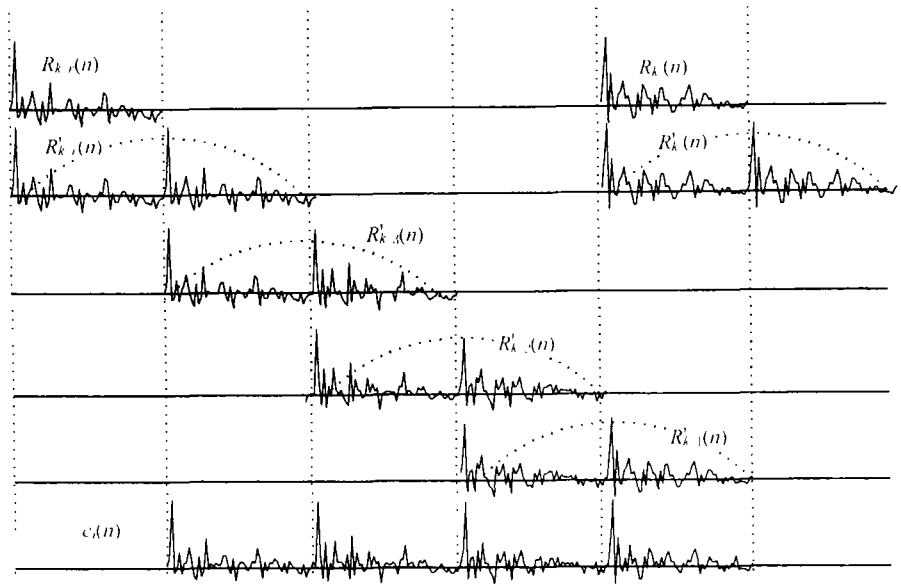


图 2 浊音残差信号的内插恢复 ($r=4$)

$$R'_{k-3}(n)=\frac{3}{4}R'_{k-4}(n-p_k)+\frac{1}{4}R'_k(n+3p_k) \tag{2}$$

$$R'_{k-2}(n)=\frac{1}{2}R'_{k-4}(n-2p_k)+\frac{1}{2}R'_k(n+2p_k) \tag{3}$$

$$R'_{k-1}(n)=\frac{1}{4}R'_{k-4}(n-3p_k)+\frac{3}{4}R'_k(n+p_k) \tag{4}$$

恢复的激励信号为

$$\bar{e}_i(n)=\sum_{j=1}^{r}R'_{k-r+j}(n)w_k(n),\qquad n_{k-r}\le n<n_k \tag{5}$$

其中 i 为帧的序号，$w_k(n)$ 满足下列条件：

$$w_k(n)+w_{k-1}(n)=1,\qquad n_k\le n<n_{k+1} \tag{6}$$

如图 2 中虚线所示。

现在的问题是如何用合理的码本结构来表示某一个基音周期内的 LPC 残差信号。首先我们设计了一种多个不同维数的码本，每个码本的维数对应于一个基音周期长度，然后用合成分析法及感觉加权均方误差最小准则分析出合适的替代残差信号的最佳码矢量 e_j，作为当前帧最后一个基音周期的激励信号 $R_k(n)$。由于基音周期的取值范围是 20~ 147，这就需要存储 128 个维数为 20~147 的码本，由于码本多维数大，所以存储量和搜索量太大，实现起来很困难。本文提出了一种嵌入分裂式矢量量化算法，该算法可大大减少码本的存储量，对于基音周期大于 40 小于 90 的语音，我们将码本分裂为两个，基音周期大于 90 的语音，码本分裂为 3 个，这样可以使搜索码本所需的计算量大大下降。

3 嵌入分裂式自适应码本及其搜索算法

浊音残差信号的基音周期内都有一个很明显的最大正脉冲，如图 2 所示，并且这些脉冲之间的距离就是基音周期的长度，我们称之为基音脉冲，因此为了节省存储量，本文提出了一种嵌入式的码本。在该码本中有 N 个 147 维的自适应矢量 $Y^i=[y_1^i,y_2^i,\cdots,y_{147}^i]$，其中 $1<i<N$。码本中每一个矢量的第一个分量 y_1^i 对应于激励信号中的基音脉冲，即它是所有分量中最大的。设当前语音的基音周期为 40，则只对码本中的前 40 维进行搜索，找到最佳码矢量 $y'=[y_1^i,y_2^i,\cdots,y_{40}^i]$ 后，用当前帧一个基音周期中的残差信号来更新自适应码本，作为下一帧的自适应码本，更新的过程如下。设自适应码本中 y^1 为最新的矢量，y^N 为最老的矢量，当前帧语音一个基音周期的残差信号为 $e=[e(1),e(2),\cdots,e(40)]$，其中 $e(1)$ 为基音脉冲。更新时将最老矢量 y^N 的前 40 维元素移出码本，将 e 移入码本作为最新的码矢量，新的自适应码本用下面的公式生成

$$\left.\begin{aligned}&Y^{i+1}=[y_1^i,y_2^i,\cdots,y_{40}^i,y_{41}^{i+1},y_{42}^{i+1},\cdots,y_{147}^{i+1}],\quad 1<i<N\\&Y^1=[e(1),e(2),\cdots,e(40),y_{41}^1,y_{42}^1,\cdots,y_{147}^1]\end{aligned}\right\}\tag{7}$$

这种嵌入式码本可大大节省储存容量，合成语音的质量基本不受影响。

嵌入式码本解决了储存量的问题，但由于尺寸太大，搜索起来需要大量的计算，无法实时实现，因此我们提出用分裂式矢量量化的算法。将 147 维的码矢量分裂成 3 个小矢量，维数分别为 40，50 和 57，每个小矢量的码本尺寸分别为 N_1, N_2, N_3。对于基音周期 $P\le 40$ 的语音，只需搜索第一个维数为 40 的码本，并且只需传输 $R=\log_2 N_1$ 个比特来表示激励信号；对于 $40<P\le 90$ 的语音，需要搜索第 1 个码本和第 2 个维数为 50 的码本，需要传输 $R=\log_2 N_1+\log_2 N_2$ 个比特来表示激励信号；对于 $P>90$ 的语音，则要对 3 个码本进行搜索，需要传输 $R=\log_2 N_1+\log_2 N_2+\log_2 N_3$ 个比特来表示激励信号。通过实验发现对合成语音主观质量影响较大的激励信号主要集中在基音脉冲周围，因此我们将基音脉冲周围的 40 个样点放在码本 1。不管基音周期为多少，激励信号的主要部分分布在码本 1 中，因此码本 1 必须分配较多的比特对这 40 维的矢量进行量化，码本 2 和码本 3 只需分配少量的比特数进行矢量量化。为了将基音脉冲周围的主要部分放在码本 1 中，则基音脉冲不能再对应于每一个矢量的第一个分量，而应放在 40 维矢量的中间，对应于第 20 个分量。

本文首先采用开环的方法求出最佳基音周期 P，然后再用闭环的方法对维数为 $P-3\sim P+3$ 的码本进行搜索，得到一个最佳激励矢量，同时更新基音周期。实验表明，这种开环 - 闭环分两步自适应码本的搜索方法，可以使计算量降低很多，而合成语音质量基本不受影响。

4 比特分配和模拟结果

在进行计算机模拟之前，我们首先要对编码器进行比特分配。该编码器是基于语音分类基础之上，因此不同类型的语音编码方案不一样，比特分配不一样，从而形成的速率也不一样。通过实验，比特分配如表 1, 2 所示。

表 1　噪声、清音和过渡音比特分配

	噪声	清音	过渡音
帧长 (ms)	40	20	10
LPC 分析阶数	6	8	10
LSP 参数编码所需比特数	14	20	24
激励编码所需比特数	14	18	28
分类所需比特数	2	2	2

对于噪声、清音和过渡音，比特分配如表 1 所示。噪声变化慢，故分析帧长取 40ms， LPC 阶数为 6，只需 14 个比特进行矢量量化即可，一帧内分成两个子帧，第 1 个子帧的激励码本和增益需比特数为 14，第 2 个子帧的激励不传输，由第 1 个子帧重复得到，再加上 2bit 的分类表示，一帧共需 30bit，故噪声的传输速率为 0.75kb/s。相对于噪声，清音的变化较快，分析帧长取 20ms， LPC 分析阶数为 8，需分配 20bit，激励需 18bit 进行矢量量化，故一帧总比特数为 40，速率为 2.0kb/s。过渡音是变化最快的一种语音，对合成语音的质量影响较大，所以分析帧长取 10ms， LPC 分析为 10 阶，需 24bit 进行矢量量化，一帧内再分成两个子帧，每个子帧的激励用 14bit 进行量化，则一帧总比特数为 54，速率为 5.4kb/s，是所有类型中速率最高的。

对于浊音，比特分配比较复杂。由于浊音的激励信号我们采用基音同步抽取的方法，即每 r 个基音周期的激励信号只传输其中一个基音，其它基音区间内信号由前后两帧保留的信号内插得到，分析帧即由 r 个基音信号组成，为变帧长。所以首先需要确定的是 r 为多少，分析帧是多长。根据语音基音周期的不同， r 取不同的值。通过实验， r 的取值如下：

当 $20 \le P \le 30, r = 5, 100 \le$ 帧长 ≤ 150；当 $30 \le P \le 60, r = 4, 120 <$ 帧长 ≤ 240；

当 $60 < P \le 80, r = 3, 180 <$ 帧长 ≤ 240；当 $80 < P \le 120, r = 2, 160 <$ 帧长 ≤ 240；

当 $120 < P \le 147, r = 1, 120 <$ 帧长 ≤ 147。

r 确定后，我们对浊音进行比特分配。每帧 LPC 分析阶数为 10，为确保 LPC 参数量化质量，我们采用了 G.723.1 语音编码标准中，基于 LSP(线谱对) 分裂矢量量化码本的 LPC 参数编码方案 [7]，用 24bit 对它进行矢量量化。由于激励信号的编码采用嵌入分裂式矢量量化，故不同的基音周期，对激励信号编码所需的比特数不同，具体比特分配如表 2 所示，语音的采样速率为 8kHz。

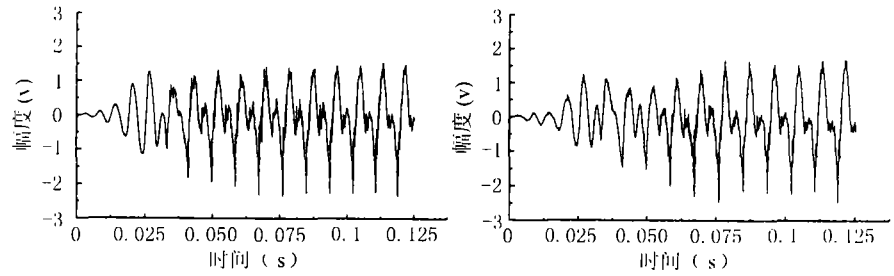
表 2 浊音的比特分配

基音周期 P	20~40	41~90	91~147
帧长 (ms)	12.5~20	20~30	15~30
LSP(bit/ 帧)	24	24	24
分类器 (bit/ 帧)	2	2	2
基音 P(bit/ 帧)	7	7	7
自适应码本 (bit/ 帧)	8	14(8+6)	19(8+6+5)
码本增益 (bit/ 帧)	7	7	7
总计 (bit/ 帧)	48	54	59
比特率 (kb/s)	2.4~3.84	1.8~2.7	2.0~3.9

通过观察大量浊音，我们发现基音周期主要集中在 30~90 之间，则浊音的平均速率只在 2.5kb/s 左右。使得该编码器在保证质量的前提下平均码率很低，非常适合于 CDMA 移动通信系统。

我们采用以上比特分配原则对一段语音信号进行编码和解码，图 3(a) 为这段语音的原始波形，图 3(b) 为编码后的恢复语音波形。从图中可看出，由于过渡带分配了较多的比特，所以它能较好地跟踪语音信号的变化。

最后我们对该编码器的质量进行了主观测试，测试语音来自 5 个男声、 5 个女声和 2 个童声，有 15 人参加非正式听音测试。同时选择 6.7kb/s 的 VSELP 、 1~8kb/s 可变速率的 QCELP 、 8kb/s 的 G.729 作为参照进行比较，测试结果如表 3 所示，恢复语音的主观质量由 MOS 分来表示。



(a) 原始语音　　(b) 编码后恢复语音

图 3

表 3　测试结果

编码器	比特率 (kb/s)	MOS
VSELP	6.7	3.78
QCELP	1~8(平均约 5)	3.61
G.729	8	3.89
SC-VR-CELP	0.75~5.4(平均约 2)	3.67

5 结　论

本文提出了一种基于语音分类的可变速率 CELP 算法 SC-VR-CELP，该算法针对语音的特征，对不同类型的语音采用不同的激励码本，分配不同的比特，特别是利用浊音准周期性的特点，提出了基于基音同步的嵌入分裂式激励码本。通过试验可以看出，虽然 SC-VR-CELP 的平均速率比 QCELP 减少了很多，但合成语音的质量却略有提高，非常适合于 CDMA 移动通信系统。

参　考　文　献

[1] A. S. Spanias, Speech coding: A tutorial review, Proc. IEEE, 1994, 82(10), 1541–1582.

[2] W. B. Kleijn, K. K. Paliwal, Speech Coding and Synthesis, Amsterdam, The Netherlands: Elsevier, 1995, 15–40.

[3] R. V. Cox, Three new speech coders from the ITU cover a rang of application. IEEE Comm. Mag., 1997, 35(9), 40–47.

[4] K. Kondo, M. Ohno, Variable rate embedded ADPCM coding scheme for packet speech on ATM networks. Proc. of ICASSP, New Mexico, USA, 1990, 405. 3.1–405. 3.4.

[5] Speech service option standard for wideband spread spectrum digital cellular system, TIA/EIA/IS-96, Feb. 1996.

[6] B. Adil, S. Eyal, A silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications, IEEE Comm. Mag., 1997, 35(9), 64–73.

[7] ITU-T Draft G.723, Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s, July 1995.

A VARIABLE RATE SPEECH CODING ALGORITHM FOR CDMA MOBILE COMMUNICATIONS

Zhu Qi Feng Guangzeng

(*Dept. of Comm. Eng., Nanjing Institute of Posts and Telecomm., Nanjing* 210003, *China*)

Abstract This paper focuses on the design of high quality speech coding with variable rate at 0.75~5.4kb/s. The new algorithm classifies input speech signals as noise, unvoiced speech, transitional speech and voiced speech, and it uses different codebooks as excited impulses according to different types of speech frames. Especially, an embedded splitting vector quantization method based on pitch synchronization to synthesize voiced speech is proposed. The speech can be recovered very well at low bit rate by making use of the pitch periodicity for voiced speech frames. This new algorithm can overcome the disadvantage of CELP whose recovered speech quality will degrade quickly when the bit rate is below 4kb/s because the codebook size is too small. The informal listening test results show that the subject quality of the algorithm exceeds that of QCELP while the average bit rate of the algorithm is only about 2kb/s, which is much lower than that of QCELP whose average bit rate is about 5kb/s. This new speech coding algorithm is therefore very apt to CDMA mobile communication systems.

Key words Speech coding, Variable rate, Vector quantization, Pitch

朱 琦: 女, 1965 年生, 副教授, 硕士, 目前主要研究方向是: 数字移动通信和语音信号处理.

郑广增: 男, 1943 年生, 教授, 博士生导师, 系主任, 中国通信学会学术工作委员会委员, 中国通信学会无线通信委员会委员. 目前主要研究方向是: 数字移动通信、个人通信及信号处理.