第 25 卷第 6 期　　声 学 学 报　　Vol. 25, No. 6
2000 年 11 月　　ACTA ACUSTICA　　Nov., 2000

基音同步特征波形内插语音编码算法

徐金标*　杜利民**
(* 安捷伦科技软件有限公司　北京　100044)
(** 中国科学院声学研究所语音交互信息研究中心　北京　100080)

1999 年 6 月 7 日收到
1999 年 9 月 20 日定稿

摘要　研究了在特征波形语音编码算法中的特征波形分解算法，提出了一种基于基音同步的特征波形内插语音编码算法。特征波形的量化采用变维矢量量化 (VDVQ)。通过实现的 2.4 kb/s 的语音质量表明，这种语音压缩算法在低码率时能得到高通信质量的重建语音。

PACS 数：　43.70

Characteristic waveform interpolation speech coding based on pitch synchronization

XU Jinbiao*　DU Limin**
(* *Agilent China Software Design Center* Beijing 100044)
(** *Research Center of Speech Interactive Information, Institute of Acoustics, The Chinese Academy of Sciences* Beijing 100080)

Received Jun. 7, 1999
Revised Sept. 20, 1999

Abstract　The decomposition algorithm of the characteristic waveform interpolation(WI) speech coding is investigated. A novel WI speech coding algorithm based on pitch synchronization is proposed. In the new speech coding algorithm, the quantization of the characteristic waveform is the variable dimension vector quantization(VDVQ). The decoding speech of 2.4 kb/s shows it can deliver decoded speech with high communication quality.

引言

最近几年，在 4 kb/s 以上码率的语音压缩算法主要是码本激励线性预测算法 (CELP)。然而，当码率低于 4 kb/s 时，由于没有足够多的比特数来精确地描述一个波形，CELP 算法会带来很大的量化噪声，重建语音的质量将严重下降[1]。因此，目前在低码率语音压缩算法的研究主要集中在谐波编码的频域的语音压缩算法[1]，如正弦变换编码 (STC)[5]、多带激励编码 (MBE)[6] 和特征波形内插编码 (WI)[1,3,4]，其中的 WI 编码算法则是当前研究的主要方向[1]。

Kleijn 提出的典型波形内插 (PWI)[2] 语音编码算法在 3 ～ 4 kb/s 编码速率能够产生高质量的浊音语音。对于浊音语音帧，通过内插当前帧和原来帧的残差域典型波形产生激励信号。然而，当为了降低比特率而增加内插帧长或语音信号的基音周期较短时，由于过强的基音周期性和混迭失真，降低了 PWI 编码算法的性能。其次，PWI 编码算法要与 CELP 编码算法相组合来重建清音语音的激励信号，这样在过渡语音段产生了语音信号的不连续性，影响了重建语音信号的质量。所以，PWI 编码算法仅在 3 kb/s 以上获得高质量浊音语音，当码率低于 3 kb/s 时，重建语音质量将下降。

针对 PWI 的缺点，Kleijn 提出了一种特征波形内插 (WI) 语音编码算法[1,3,4]，这种方法是将残差信号域的提取的特征波形分解为慢渐变和快渐变两部

分，然后，对这两部分分别编码，这种方法有效的降低了码率；并且该方法不分清浊音判决，因此，这种算法受到了研究人员的普遍关注，目前的低码率编码算法的研究主要集中在这类算法的研究。

本文提出了一种新的波形分解方法，该方法保证每帧有两个精确表示的特征波形，首先对提取的特征波形 (CW) 进行傅里叶变换，然后求一帧内的 CW 的傅里叶级数的均值，这个均值就是本文中提出的慢渐变波形 (SEW)，而快渐变波形 (REW) 由原始信号和 SEW 之差得到。一帧内传输上述傅里叶级数的均值和两个 REW 谱矢量。该算法的线性预测 (LP) 激励由 SEW 和 REW 两部分组成， REW 由所传输的两个 REW 谱矢量通过线性内插得到，再加上傅里叶级数的均值 (SEW) 可重建 CW, SEW 和 REW 由变维矢量量化方法量化。实验表明，基于该模型的 2.4 kb/s 语音编码算法的语音质量接近或超过 4.8 kb/s 的 FS1016 标准。

1 帧内基音同步特征波形内插语音编码算法

本文提出的帧内基音同步特征波形内插语音编码算法的编码 / 译码原理见图 1 和图 2 。该算法处理的语音帧长 L_f 为 200 个样点。该算法与一般的 WI 算法一样，是在残差域进行分析，提取特征波形，然后进行特征波形的分解。我们在本文的后面部分分别介绍新算法特有的几个部分：特征波形的提取、特征波形的分解和特征波形的两个部分 SEW 和 REW 的量化。

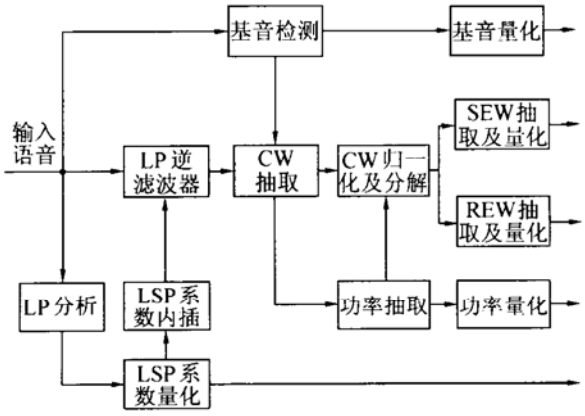


图 1 编码器原理框图

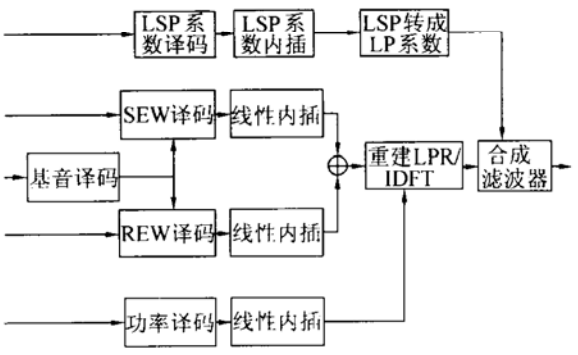


图 2 译码器原理框图

2 残差信号上特征波形的提取和分解

2.1 特征波形的提取

在线性预测残差域 (LPR) 中以基音周期长度为间隔抽取特征波形。假定经基音检测后第 k 帧语音信号的基音周期为 N，第 k 帧的残差信号可由 M 个特征波形来表示，这 M 个 CW 可用如下矩阵表示：

$$\boldsymbol{R} = (P_0, P_1, \cdots, P_{M-1}), \tag{1}$$

$$\begin{cases} P_i = [r(i,0), r(i,1), \cdots, r(i,N-1)]^T, \\ \qquad (0 \le i \le M-1), \end{cases} \tag{2}$$

其中 M 由下式决定：

$$M = \begin{cases} \dfrac{L_f}{N}, & \text{如果 } L_f \bmod N = 0, \\ \left\lfloor \dfrac{L_f}{N} \right\rfloor, & \text{如果 } L_f \bmod N \neq 0. \end{cases} \tag{3}$$

最早的波形 P_0 也许包含前一帧的样本，这个数据叠接能够改善帧间平滑，图 3 说明了 $M = 3$ 时一帧中的特征波形的抽取过程。因此，矩阵 $\boldsymbol{R}$ 为：

$$\boldsymbol{R} = \begin{bmatrix} r(0,0) & r(0,1) & \cdots & r(0,M-1) \\ r(1,0) & r(1,1) & \cdots & r(1,M-1) \\ \vdots & \vdots & \vdots & \vdots \\ r(N-1,0) & r(N-1,1) & \cdots & r(N-1,M-1) \end{bmatrix}. \tag{4}$$

矩阵 (4) 与两种类型的傅里叶变换谱有关，一个为按列的傅里叶变换，即在给定时间 t 处，表示一个与该时刻相联系的短时谱；另一个为按行的傅里叶变换，即对一个给定的相位值表示一个渐变频率谱，它与 CW 的渐变率有关[1] 。为了更好地重建语音信号，语音信号应在每个基音周期采样一次，若基音周期为 N，则 CW 的提取率和渐变带宽分别至少为 $1/N$ 和 $1/2N$，即提取率随基音周期而变。当语音信号的采样率为 8 kHz 时，基音周期的变化范围为 20 ～ 147 个样点，则 CW 的抽取和渐变带宽分别为 400 ～ 60 Hz 和 200 ～ 30 Hz 。对于 200 个样点的语音帧 (25 ms)，每帧包含的特征波形个数为 2 ～ 10 个。这种提取 CW 的方法与[7] 中提出的波形提取的方法类似。

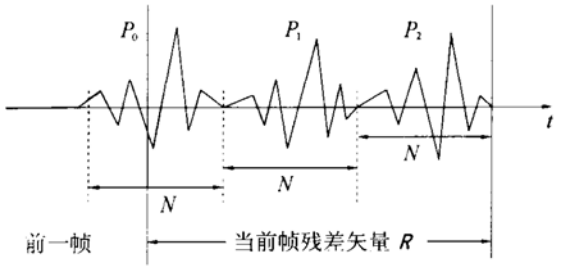


图 3　基音特征波形抽取

2.2　特征波形的分解

Kleijn 提出对特征波形分解成周期性较强的慢渐变部分和非周期性的快渐变部分，然后对这两部分分别量化，这种方法是目前对 CW 量化的最有效方法[1,3,4]。目前，有两种方法，一种是 CW 分别通过一个线性低通滤波器和一个线性高通滤波器[1,3,4]；另一种方法是利用小波变换对 CW 进行分解[8]。另外[7]中将 DFT 谱中的高频分量直接除去的方法也是一种较为常见的方法，但是该方法只是针对浊音段。新算法的分析是不分清音 / 浊音处理的。在本文中我们提出了一种有效的特征波形分解方法。

对矩阵 $\boldsymbol{R}$ 的每一列分别作 DFT 变换，得到矩阵 $\boldsymbol{R}_1$:

$$\boldsymbol{R}_1=\begin{bmatrix} r_1(0,0) & r_1(0,1) & \cdots & r_1(0,M-1) \\ r_1(1,0) & r_1(1,1) & \cdots & r_1(1,M-1) \\ \vdots & \vdots & \vdots & \vdots \\ r_1(N/2-1,0) & r_1(N/2-1,1) & \cdots & r_1(N/2-1,M-1) \end{bmatrix}. \tag{5}$$

在对 $\boldsymbol{R}_1$ 的每个列向量作分解之前，首先对 $\boldsymbol{R}_1$ 的每个列向量作归一化处理，使得每个 CW 的能量为 1，得到矩阵 $\boldsymbol{R}_1'$

$$\boldsymbol{R}_1'=\begin{bmatrix} r_1'(0,0) & r_1'(0,1) & \cdots & r_1'(0,M-1) \\ r_1'(1,0) & r_1'(1,1) & \cdots & r_1'(1,M-1) \\ \vdots & \vdots & \vdots & \vdots \\ r_1'(N/2-1,0) & r_1'(N/2-1,1) & \cdots & r_1'(N/2-1,M-1) \end{bmatrix}. \tag{6}$$

通过研究表明，在每一帧内对的每一个 CW 的 DFT 系数求均值，即将矩阵 $\boldsymbol{R}_1'$ 的列向量求均值，通过这个求均值得到的向量恢复的语音信号变化缓慢，它对应于 PCW 的慢渐变部分 (SEW)，本文中的 SEW 如下得到：

$$SEW=(sew_0,sew_1,\cdots,sew_{N/2-1})^T, \tag{7}$$

$$\begin{cases} sew_i=\dfrac{1}{M}\displaystyle\sum_{l=0}^{M-1} r_1'(i,l), \\ \qquad i=0,\cdots,\dfrac{N}{2}-1. \end{cases} \tag{8}$$

每一帧中的 CW 的快渐变谱 REW 可由原语音波形谱减去其均值得到，因此，得到矩阵 REW:

$$REW=(REW_0,REW_1,\cdots,REW_{M-1})=$$
$$\begin{bmatrix} rew(0,0) & rew(0,1) & \cdots & rew(0,M-1) \\ rew(1,0) & rew(1,1) & \cdots & rew(1,M-1) \\ \vdots & \vdots & \vdots & \vdots \\ rew(N/2-1,0) & rew(N/2-1,1) & \cdots & rew(N/2-1,M-1) \end{bmatrix}. \tag{9}$$

$$\begin{cases} rew(i,j)=r_1(i,j)-sew_i, \quad i=0,1,\cdots,N/2, \\ \qquad\qquad j=0,1,\cdots,M-1. \end{cases} \tag{10}$$

通过 (8) 和 (9) 两式将特征波形分解成两部分。由于 SEW 每帧只有一个，并且随时间缓慢变化。所以，其提取率为每帧抽取一次。

由于 REW 变化迅速，所以，其提取率相对 SEW 的提取要高，针对不同码率声码器的要求，可以决定每帧提取几个 REW。在后面要介绍的 2.4 kb/s 声码器中，以 80 Hz 作为快渐变波形的提取率，每帧提取两个 REW；即矩阵 (9) 中的第一列和第 M 列。

2.3　PCW 谱的变维矢量量化

原理上，借助于一个固定维数的矢量量化码本对变维矢量进行优化矢量量化是可行的。当前，VDVQ 主要应用在频域参数语音编码方案中，如正弦变换编

码 (STC)[5] 和多带激励编码 (MBE)[6]. 这些编码算法中最困难的问题就是语音谱幅度的有效量化. 由于语音短时谱随时变化, 需要重建的谱特性的谐波个数与基音周期有关, 而基频是随时间变化的, 所以, 每个谱所需要的幅度样点数是个变量.

文献 9 首先提出了一种 VDVQ 方法, 在 MBE 类编码器中对变维谱量化问题提供了较好的解决方法. 在本文中, 对 SEW 和 REW 量化也采用 [9] 中介绍的 VDVQ 方法, 即用一个具有固定维数, 覆盖了所考虑的输入矢量全部范围的通用码本量化 SEW 和 REW 谱矢量, 通用码本和结构化的 VDVQ 相结合, 减少了存储和计算复杂度, 产生了高的量化效率.

3 2.4 kb/s 码率的语音压缩算法

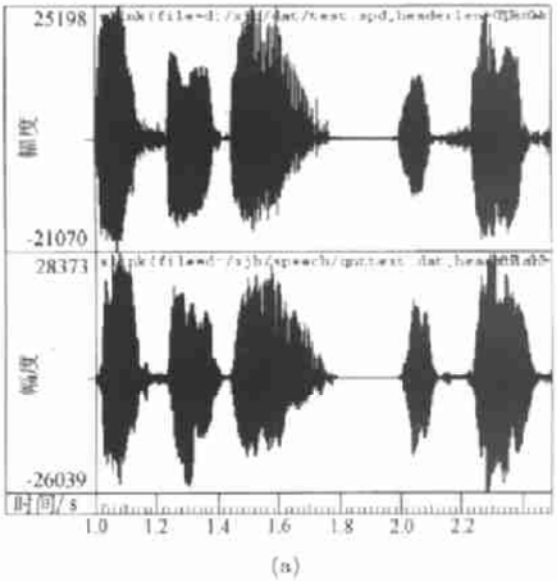
2.4k b/s 码率的语音压缩算法的原理见图 1, 编码器首先估计 10 阶线性预测 (LP) 器系数, 将 LP 系数转换成线谱对系数 (LSF), 然后对 LSF 系数进行内插, LSF 系数以 40 Hz 提取, 用 24 比特进行分裂矢量量化 (分裂成维数分别为 3、3 和 4, 每一个子矢量用 8 比特量化); 基音检测算法则采用文献 1 中介绍的基本的归一化互相关函数法. 语音信号通过 LP 逆滤波器得到残差域 (LPR) 信号, PCW 以 $1/N$(Hz) 的速率在 LPR 上提取, 然后对提取的 CW 作 DFT 变换, 提取功率, 并归一化 CW 的谱矢量; 将归一化的 CW 谱矢量进行分解, 得到一个 SEW 和 M 个 REW 谱矢量, 并提取这个 SEW 谱矢量和 2 个 REW 谱矢量 (矩阵 (9) 中的第 1 列和第 M 列), 与两个 REW 的提取位置相对应, 在谱矢量的归一化处理过程中以 80 Hz 速率提取 CW 的功率 (即第一个和第 M 个 CW 的能量). 比特分配如表 1 所示.

表 1 2.4 kb/s 语音编码算法比特分配表

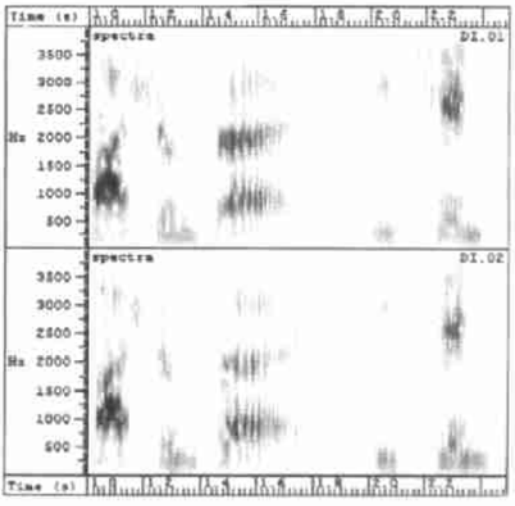
参数	LPC	基音	功率	SEW	REW	总比特数
比特数	24	7	6	7	4	59
提取率 /Hz	40	40	80	40	80	40

译码器的原理框图如图 2 所示. 对 LP 系数进行译码得到量化的 LSF, 对得到的 LSF 进行内插, 将通过内插得到的 LSF 转换成 LP 系数. 对基音周期译码, 得到量化的基音周期. 同时通过 SEW 和 REW 的序号得到一个量化 SEW 谱和两个量化的 REW 谱, 通过线性内插, 得到量化的 M 个 PCW 的谱矢量; 通过功率译码, 得到量化的平均功率, 然后通过线性内插得到 M 个量化的平均功率; 然后, 通过 M 个 IDFT 重建量化的残差信号 (LPR); 将得到的量化的残差信号通过合成滤波器, 就得到重建的合成语音信号.

用于训练 SEW 和 REW 谱矢量码本的语音数据库为汉语语音, 长度约 30 min, 这些语音来源于 30 多名不同年龄段的男女讲话录音. 语音库经 100 ~ 3400 Hz 带通滤波器, 以 8 kHz 采样数字化, 每个样点用 16 比特线性码表示. 经广泛的非正式主观试听表明, 该 2.4 kb/s 算法的重建语音质量明显优于 2.4 kb/s 的 LPC10e, 并且接近或超过 4.8 kb/s 的 FS1016 标准. 预示该算法在 2.4 kb/s 码率提供语音通信具有非常大的潜力. 图 4 是原始语音与经以上介绍的编码 / 解码器之后得到的重建语音、对应的语谱图及其残差信号的比较. 限于篇幅, 我们在图中不与 LPC10e 或 FS1016 CELP 算法比较.

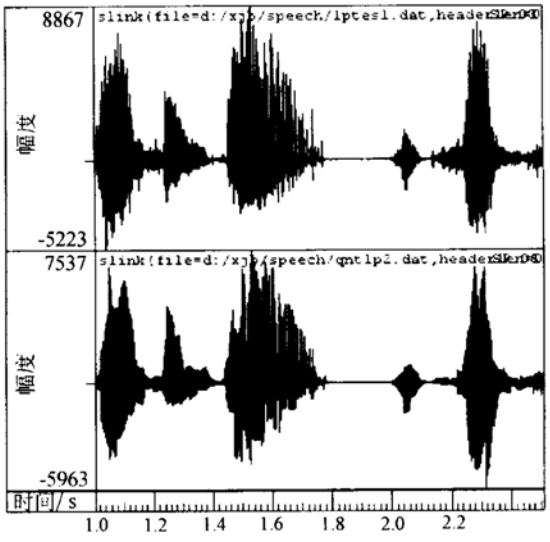


(a)



(b)

图 4



(c)

图 4　一段原始语音与经以上介绍的编码 / 解码器之后得到的重建语音、对应残差信号的语谱图及其的比较；其中 (a) 是原始语音与重建语音的比较，(b) 是 (a) 中对应的信号的语谱图，(c) 是 (a) 信号的残差信号；并且 (a)、(b) 和 (c) 中对应的中上面的图中的信号是对应原始语音的，下面的图中的信号对应的信号是重建语音的。

4　结论

本文提出了一种新的特征波形内插语音压缩算法，它通过帧内特征波形内插重建语音信号。2.4 kb/s 码率的声码器的语音质量表明，该算法是一种非常有效的低码率语音压缩算法，能够得到良好的压缩效率，具有一定的实用价值。

参　考　文　献

1　Kleijn K B, Paliwal K K *et al.* Speech coding and synthesis. Elsevier, 1995

2　Kleijn W B. Encoding speech using prototype waveforms. *IEEE Trans. Speech and Audio Processing*, 1993; **1**(4): 386—399

3　Kleijn W B, Haagen J. Transformation and decomposition of the speech signal for coding. *IEEE Signal Processing Letter*, 1994; **1**(9): 136—138

4　Kleijn W B, Haagen J. A speech coder based on decomposition of characteristic waveforms. in Proc. Int. Conf. Acoustic Speech Signal Processing, 1995: 508—511

5　McAulary R J, Quatieri T F. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans. Acoustic Speech Signal Processing*, 1986; **34**: 744—754

6　Griffin D W, Lim J S. Multiband excitation vocoder. *IEEE Trans. Acoustic Speech Signal Processing*, 1988; **36**(8): 1223—1235

7　Tanaka Y, Kimura H. Low-bit-rate speech coding using a two-dimensional transform of residual signals and waveform interpolation. in Proc. Int. Conf. Acoustic Speech Signal Processing, 1994: I173—I176

8　Nicola R Chong, Ian S Burnett, Joe F Chicharo, Mark M Thomson. Use of the pitch synchronous wavelet transform as a new decomposition method for WI. in Proc. Int. Conf. Acoustic Speech Signal Processing, 1998: I513—I516

9　Das A, Gersho A. Variable dimension vector quantization of speech spectra for low-rate vocoders. IEEE Proc. Of the Data Commpression Conference, 1994: 420—429