

宽带 ISF 参数的非等系数帧间预测分裂矢量量化方法

李海婷, 鲍长春

(北京工业大学电子信息与控制工程学院语音与音频信号处理研究室, 北京 100124)

摘 要: 本文提出了一种新的适用于宽带语音编码 ISF 参数量化的非等系数帧间预测分裂矢量量化方案。该量化方案利用 ISF 参数的帧间相关性, 基于预测分裂矢量量化原理, 首先对待量化的 ISF 参数矢量进行去均值和非等系数帧间预测, 然后对去均值后的 ISF 参数的预测残差进行分裂矢量量化。实验表明, 该算法在每帧编码比特数为 46bits 时达到了透明量化, 且平均谱失真比 G. 722. 2 中 ISF 参数量化的平均谱失真小。

关键词: 语音编码; 矢量量化; 导抗谱频率; 非等系数帧间预测

中图分类号: TN912. 3 **文献标识码:** A **文章编号:** 0372-2112 (2008) 06-1214-04

Unequal Coefficients Interframe Predictive Split Vector Quantizer for Wideband ISF Parameters

LI Hai-ting, BAO Chang-chun

(Speech and Audio Signal Processing Lab, School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: An unequal coefficients interframe predictive split vector quantizer is proposed. This quantizer is used for quantizing the ISF (Immittance Spectral Frequencies) parameters of wideband speech. This quantizer is based on interframe correlation and split vector quantization. In this quantizer, the unequal coefficients interframe prediction is firstly applied on the mean-removed ISF parameters vector to be quantized, and then the ISF residual vector is quantized using split vector quantizer. Experimental results show that this quantization scheme achieves transparent coding at 46bits/frame, and the average spectral distortion is lower than that of AMR-WB (ITU-T-G. 722. 2) speech coding standard.

Key words: speech coding; vector quantization; immittance spectral frequencies; unequal coefficients interframe prediction

1 引言

在宽带 (50 ~ 7000 Hz) 语音编码领域, 如何用有限的比特数达到尽可能高的量化效率, 同时具有较强的错误鲁棒性一直是困扰着人们的问题。宽带语音通常需要 16 阶 ISF (Immittance Spectral Frequencies) 参数^[1] 表征语音的谱包络。如果直接对这 16 阶参数进行矢量量化势必会造成较大的计算复杂度和存储复杂度, 而且需要消耗较多的比特数。为此, 相继有学者提出预测矢量量化^[2]、分裂矢量量化^[3]、多级矢量量化^[4]、乘积码锥形矢量量化^[5] 或几种方法相结合的矢量量化^[6~8] 方案。其中, 预测分裂矢量量化^[9] 已被广泛运用到宽带语音编码标准中。现有的预测方法如一步差值预测^[10]、等系数 MA 预测^[11] 等, 往往不能同时兼顾有丢失帧和无丢失帧时的量化性能。因此, 本文提出了一种宽带语音编码 ISF 参数的非等系数帧间预测分裂矢量量化方案, 旨在平衡有丢失帧和无丢失帧时 ISF 参数量化的性能。

本量化方案基于预测分裂矢量量化原理, 利用了 ISF 参数的帧间相关性。首先, 对待量化的 ISF 参数矢量去均值; 然后, 根据训练出的非等系数帧间预测系数, 得到 ISF 参数的预测矢量; 接下来, 对去均值后的 ISF 参数的预测残差矢量进行分裂矢量量化。本量化器在每帧编码比特数为 46bits 时, 实现了对 16 阶 ISF 系数的透明量化^[3]。

2 本文提出的算法

本文提出的宽带语音编码 ISF 参数的非等系数帧间预测分裂矢量量化方案, 先通过非等系数帧间预测的方法求出当前帧去均值后 ISF 参数的预测残差矢量, 再利用分裂矢量量化 (Split Vector Quantization) 的方法量化 ISF 参数的预测残差矢量。非等系数帧间预测, 在一定程度上去除了 ISF 参数的帧间相关性, 减小了量化所需的比特数。以下将根据图 1、图 2 对本文提出的宽带语音 ISF 参数的非等系数帧间预测分裂矢量量化方法作详细介绍。

收稿日期: 2007-04-06; 修回日期: 2008-03-24

基金项目: 北京市自然科学基金 (No. 4082006); 华为技术有限公司合作资助项目

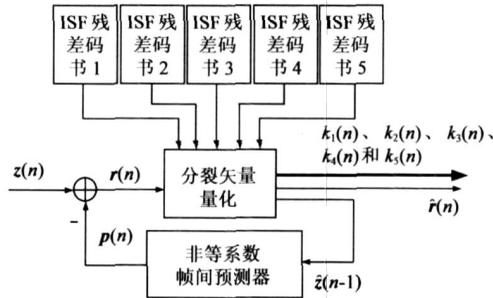


图 1 非等系数帧间预测分裂矢量量化的编码器原理框图

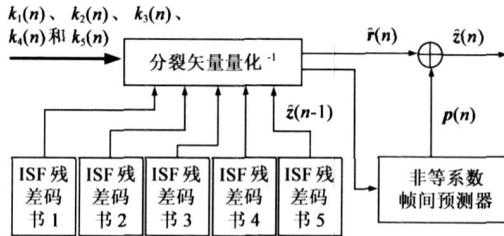


图 2 非等系数帧间预测分裂矢量量化的译码器原理框图

图 1 和图 2 中, $z(n)$ 为去均值后的第 n 帧 ISF 矢量, $\hat{z}(n)$ 为去均值后的第 n 帧 ISF 参数的量化矢量, $p(n)$ 为第 n 帧 $z(n)$ 矢量的预测矢量, $r(n)$ 为第 n 帧 ISF 的预测残差矢量, $\hat{r}(n)$ 为第 n 帧量化的 ISF 预测残差矢量, $\hat{z}(n-1)$ 为去均值后的第 $n-1$ 帧 ISF 参数的量化矢量, $k_1(n)$ 、 $k_2(n)$ 、 $k_3(n)$ 、 $k_4(n)$ 和 $k_5(n)$ 分别表示第 n 帧五个 ISF 残差码书的码矢量索引值。

非等系数帧间预测分裂矢量量化器编码端的工作原理如图 1 所示:设 $\hat{z}_i(n)$, $i=0,1, \dots, m-1$ 为第 n 帧的第 i 个 ISF 频率, m 为 ISF 参数的阶数. 为了减少计算和量化对象的动态范围,通常将第 i 个 ISF 频率的平均值 $\bar{z}_i(n)$, $i=0,1, \dots, m-1$ 从 $\hat{z}_i(n)$ 中减去,得到当前帧的第 i 个去均值后的 ISF 频率 $z_i(n)$:

$$z_i(n) = \hat{z}_i(n) - \bar{z}_i, \quad i=0,1, \dots, m-1 \quad (1)$$

首先,在编码端,将第 $n-1$ 帧量化的去均值后的第 i 个 ISF 频率与第 i 个非等系数帧间预测器系数的乘积作为第 n 帧的第 i 个 ISF 频率的预测值,即 $p_i(n)$

$$p_i(n) = \hat{z}_i(n-1), \quad i=0, \dots, m-1 \quad (2)$$

式中, m 为 ISF 参数的阶数, $\hat{z}_i(n-1)$, $i=0 \dots m-1$ 为第 $n-1$ 帧量化后的去均值的第 i 个 ISF 频率, \bar{z}_i , $i=0 \dots m-1$ 为非等系数帧间预测器的预测系数. \bar{z}_i 可以通过使平方预测误差最小的方法,从训练语音序列中估计出来,并在量化器中保持恒定. \bar{z}_i 的具体估计方法为:

令第 i 个 ISF 参数的平方预测误差为:

$$E_i = \sum_{n=1}^{N_f} [r_i(n)]^2 = \sum_{n=1}^{N_f} [z_i(n) - p_i(n)]^2 \quad (3)$$

式中, N_f 为训练数据的总帧数. 令 $\partial E_i / \partial \hat{z}_i = 0$, 预测系数 \hat{z}_i 被确定的过程如下:

$$E_i = \sum_{n=1}^{N_f} [z_i(n) - p_i(n)]^2 = \sum_{n=1}^{N_f} [z_i(n) - \hat{z}_i(n-1)]^2$$

$$= \sum_{n=1}^{N_f} [z_i^2(n) + \hat{z}_i^2(n-1) - 2 z_i(n) \hat{z}_i(n-1)] \quad (4)$$

$$\partial E_i / \partial \hat{z}_i = 2 \sum_{n=1}^{N_f} \hat{z}_i(n-1) - 2 \sum_{n=1}^{N_f} z_i(n) \hat{z}_i(n-1) = 0 \quad (5)$$

$$\hat{z}_i = \frac{\sum_{n=1}^{N_f} z_i(n) \hat{z}_i(n-1)}{\sum_{n=1}^{N_f} \hat{z}_i^2(n-1)}, \quad i=0,1, \dots, m-1 \quad (6)$$

计算时为方便起见,在估计中用原始数据 $\hat{z}_i(n-1)$ 代替量化数据 $\hat{z}_i(n-1)$. 但是为了平衡 ISF 参数量化在无丢失帧和有丢失帧时的量化性能,这里对由原始训练数据估计出的预测系数作了一定的处理. 具体的做法是将估计出的预测系数乘以 0.5 作为最后的非等系数帧间预测系数. 这样做的目的是为在发生丢帧的情况时可以有更好的恢复出 ISF 参数.

第二,将输入量化器的无偏 ISF 参数矢量 $z(n)$ 与其预测值 $p(n)$ 相减,得到 ISF 预测残差矢量 $r(n)$,即:

$$r(n) = z(n) - p(n) \quad (7)$$

预测残差矢量的动态范围比 $z(n)$ 的动态范围减小很多,因而非常适合作为量化的对象以提高量化的效率.

第三,将预测后得到的 ISF 预测残差矢量 $r(n) = [r_0(n), r_1(n), \dots, r_{15}(n)]$ 进行分裂矢量量化(SVQ). $r(n)$ 矢量将被分裂为五个子矢量 $r^{(1)}(n)$ 、 $r^{(2)}(n)$ 、 $r^{(3)}(n)$ 、 $r^{(4)}(n)$ 和 $r^{(5)}(n)$. 其中 $r^{(1)}(n)$ 、 $r^{(2)}(n)$ 、 $r^{(3)}(n)$ 和 $r^{(4)}(n)$ 均是 3 维矢量, $r^{(5)}(n)$ 是 4 维矢量,对这五个子矢量分别用表 1 中分配的比特数进行矢量量化.

这里用量化误差平方作为量化过程中的失真测度. 通常,对于一个输入的 ISF 预测残差子矢量 $\{r_i\}$ 和码书中的第 k 个码字 $\{\hat{r}_i^k\}$ 来说,量化的过程就是在码书中寻找使下式的值最小的那个码字的索引号:

$$E = \sum_{i=m}^n [r_i, \hat{r}_i^k]^2 \quad (8)$$

这里 m 和 n 表示子矢量中的第一个和最后一个元素在整个预测残差 ISF 矢量中的元素序号. 五个子矢量的量化分别对应不同的残差码书,搜索得到的五个残差码书的码矢量索引值分别记作 $k_1(n)$ 、 $k_2(n)$ 、 $k_3(n)$ 、 $k_4(n)$ 和 $k_5(n)$. 将其进行二进制编码后写入码流.

表 1 矢量量化比特分配表

未量化的 16 维 ISF 预测残差矢量 $r(n)$	
$r^{(1)}(n) = [r_0(n), r_1(n), r_2(n)]$	10bits
$r^{(2)}(n) = [r_3(n), r_4(n), r_5(n)]$	10bits
$r^{(3)}(n) = [r_6(n), r_7(n), r_8(n)]$	9bits
$r^{(4)}(n) = [r_9(n), r_{10}(n), r_{11}(n)]$	9bits
$r^{(5)}(n) = [r_{12}(n), r_{13}(n), r_{14}(n), r_{15}(n)]$	8bits

第五,根据非等系数帧间预测分裂矢量量化器译码端的工作原理,重构 ISF 预测残差矢量、无偏 ISF 矢量及 ISF 矢量。

第六,用重构的无偏 ISF 矢量更新预测器中 $\hat{z}(n-1)$ 的值,以用于下一帧的非等系数帧间预测。

ISF 残差码书均采用传统的 LBG 算法^[12]训练得到,训练数据为包括汉语、英式英语、美式英语、芬兰语、日语、法语在内的多语种数据库,长度约为 2 个小时,采样率为 16kHz,精度为 16 位线性 PCM。

非等系数帧间预测分裂矢量量化器译码端的工作原理如图 2 所示:首先,译码端根据接收到的码书索引值 $k_1(n)$ 、 $k_2(n)$ 、 $k_3(n)$ 、 $k_4(n)$ 和 $k_5(n)$,通过查表从五个 ISF 残差码书中找到对应的重构预测残差子矢量的第 i 个元素分别为 $\hat{r}_i^{k_1(n)}$, $i=0,1,2$ 、 $\hat{r}_i^{k_2(n)}$, $i=3,4,5$ 、 $\hat{r}_i^{k_3(n)}$, $i=6,7,8$ 、 $\hat{r}_i^{k_4(n)}$, $i=9,10,11$ 和 $\hat{r}_i^{k_5(n)}$, $i=12,13,14,15$ 。将这五个重构的预测残差子矢量合并,得到重建的 ISF 预测残差矢量 $\hat{r}(n)$,其第 i 个元素为:

$$r_i(n) = \begin{cases} \hat{r}_i^{k_1(n)}, & i=0,1,2 \\ \hat{r}_i^{k_2(n)}, & i=3,4,5 \\ \hat{r}_i^{k_3(n)}, & i=6,7,8 \\ \hat{r}_i^{k_4(n)}, & i=9,10,11 \\ \hat{r}_i^{k_5(n)}, & i=12,13,14,15 \end{cases} \quad (9)$$

第二,用非等系数帧间预测过程得到 ISF 参数的预测矢量 $p(n)$,第 i 个元素为:

$$p_i(n) = a_i \hat{z}_i(n-1), \quad i=0, \dots, m-1 \quad (10)$$

这里,阶数 $m=16$, $\hat{z}(n-1)$ 为前一帧量化后的无偏 ISF 矢量, a_i 为非等系数帧间预测系数。

第三,将 $\hat{r}(n)$ 与 $p(n)$ 各对应分量相加得到量化后的无偏 ISF 参数子矢量第 i 个元素为:

$$\hat{z}_i^{k_1(n)}(n) = p_i(n) + \hat{r}_i^{k_1(n)}(n), \quad i=0,1,2 \quad (11)$$

$$\hat{z}_i^{k_2(n)}(n) = p_i(n) + \hat{r}_i^{k_2(n)}(n), \quad i=3,4,5 \quad (12)$$

$$\hat{z}_i^{k_3(n)}(n) = p_i(n) + \hat{r}_i^{k_3(n)}(n), \quad i=6,7,8 \quad (13)$$

$$\hat{z}_i^{k_4(n)}(n) = p_i(n) + \hat{r}_i^{k_4(n)}(n), \quad i=9,10,11 \quad (14)$$

$$\hat{z}_i^{k_5(n)}(n) = p_i(n) + \hat{r}_i^{k_5(n)}(n), \quad i=12,13,14,15 \quad (15)$$

第四,将无偏 ISF 参数矢量加上 ISF 的均值 \bar{z} ,最后得到量化后的整个 ISF 参数矢量 $\hat{z}(n)$,第 i 个元素为:

$$\hat{z}_i(n) = \begin{cases} \bar{z}_i + p_i(n) + \hat{r}_i^{k_1(n)}(n), & i=0,1,2 \\ \bar{z}_i + p_i(n) + \hat{r}_i^{k_2(n)}(n), & i=3,4,5 \\ \bar{z}_i + p_i(n) + \hat{r}_i^{k_3(n)}(n), & i=6,7,8 \\ \bar{z}_i + p_i(n) + \hat{r}_i^{k_4(n)}(n), & i=9,10,11 \\ \bar{z}_i + p_i(n) + \hat{r}_i^{k_5(n)}(n), & i=12,13,14,15 \end{cases} \quad (16)$$

其中, \bar{z}_i 为 ISF 参数各分量的平均值。该均值是事先从所有 ISF 训练矢量集中计算得到的,在量化器中为常数。

第五,用重构的无偏 ISF 矢量更新预测器中 $\hat{z}(n-1)$ 的值,以用于下一帧的非等系数帧间预测。

3 实验结果及性能分析

为了验证本文所提方法的有效性,本文进行了 2 个仿真试验,并进行了性能分析,具体如下所述。

在仿真实验 1 中,需要用训练语音外的 345 秒(17250 帧)汉语语音产生 17250 个 16 阶的 ISF 矢量,具体做法如下:首先,对输入语音进行预处理,将采样率调整到 12.8kHz,并进行 50Hz 的低通滤波。第二,对预处理后的输入语音信号进行线性预测分析。分析窗长为 30ms,其中中间的 20ms(256 个样点)来自当前帧,前面的 5ms(64 个样点)来自上一帧,后面的 5ms(64 个样点)来自下一帧。窗函数采用集中于当前帧第四子帧处的非对称窗,该窗由两部分组成,第一部分是半个汉明窗(256 点),第二部分是四分之一余弦函数(128 点)。第三,对加窗后的语音进行自相关估计,用滞后窗乘以自相关系数使其具有 60Hz 的带宽扩展。第四,采用莱文逊-杜宾递归算法获取 $m=16$ 阶 LP 系数。第五,将 LP 系数 $i(n)$, $i=0,1, \dots, m-1$ 转化为 ISF 参数 $z_i(n)$, $i=0,1, \dots, m-1$ 。

在同一训练集下,分别用 ITU-T G 722.2 中 46bits/frame 的 ISF 参数量化器和本文提出的 46bits/frame 的量化器对这 17250 帧 ISF 矢量进行量化,并按国际通用的方法计算平均谱失真^[11],实验结果如表 2 所示。

表 2 16 阶 ISF 参数矢量的量化性能对比

测试项	ITU-T G. 722.2 中的 ISF 参数量化方案	本文提出的量化方案
(0~7kHz 内)平均谱失真	0.807256dB	0.787173dB
2dB~4dB	0.875%	0.968%
>4dB	0%	0%

仿真实验 1 的测试结果表明,在无丢失帧的情况下,本文提出的 46bits/frame 的量化方案得到的平均谱失真为 0.787173dB,谱失真介于 2dB 和 4dB 之间的帧的百分比仅为 0.968% (<2%),而平均谱失真大于 4dB 的帧的百分比为零,其量化性能达到了透明量化的标准。与 ITU-T G 722.2 中的 ISF 参数量化方案相比较,虽然本文提出的算法得到的谱失真介于 2dB 和 4dB 之间的百分比要略大于 G 722.2,但是平均谱失真要小于 G 722.2。

在仿真实验 2 中,调整本算法中非等系数帧间预测器系数估计时原始系数所乘的比例因子,用仿真实验 1 中的数据重新进行测试,测验结果如表 3 所示。

仿真实验 2 的测试结果表明,当比例因子为 1 时,即直接用原始估计出的系数作为非等系数帧间预测器

系数时,得到的平均谱失真最小,且远远小于 G 722.2 的值.但是在本文提出的算法中并没有采用直接将原始估计出的系数作为非等系数帧间预测器系数的方法,而是将比例因子设定为 0.5.这样做主要是为了在有丢失帧的情况下能更好地恢复出丢失帧的 ISF 参数,尽量减小错误繁殖.同样由前一接收帧的信息恢复丢失帧的 ISF 参数,在帧丢失率为 3%的情况下,本文提出的算法恢复出来的语音没有明显的厌恶声,而采用比例因子为 1 时得到的合成语音可以明显听到不连续的声音.

表 3 调整预测系数估计时所乘比例因子后的量化性能对比

比例因子	1	2/3	0.5
(0 ~ 7kHz 内) 平均谱失真	0.7300dB	0.7543dB	0.7871dB
2dB ~ 4dB	1.484 %	1.11 %	0.968 %
> 4dB	0 %	0 %	0 %

4 结论

本文提出了一种适用于宽带 ISF 参数量化的非等系数帧间预测分裂矢量量化方案.该量化方案基于预测分裂矢量量化原理,利用了 ISF 参数的帧间相关性,采用非等系数预测的方法去除帧间冗余,实现了平衡有丢失帧和无丢失帧时 ISF 参数量化器量化性能的目的.在无丢失帧的情况下,每帧用 46bits 进行量化,得到的平均谱失真 < 1dB,谱失真介于 2dB 和 4dB 之间的帧的百分比仅为 0.968% (< 2%),而平均谱失真大于 4dB 的帧的百分比为零,达到了透明量化的标准.在有丢失帧的情况下,只用前一接收帧的信息就能恢复出丢失帧的 ISF 参数,并且错误的延续性能保持保持在几帧之内.从听觉上讲,恢复出来的语音没有明显的厌恶声.

参考文献:

- [1] Y Bistriz, S Peller. Immittance spectral pairs (ISP) for speech encoding[A]. Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 2) [C]. Minneapolis, Minnesota: IEEE, 1993. 9 - 12.
- [2] J Skoglund, J Linden. Predictive VQ for noisy channel spectrum coding: AR or MA[A]. Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 2) [C]. Munich, Germany: IEEE, 1997. 1351 - 1354.
- [3] Paliwal K K, Atal B S. Efficient vector quantization of LPC parameters at 24 bits/frame[J]. IEEE Trans on Speech and Audio Processing, 1993, 1(1): 3 - 14.
- [4] 鲍长春. 低比特率数字语音编码基础[M]. 北京: 工业大学出版社, 2001. 96 - 148.
- [5] 李海婷, 鲍长春. 宽带 ISF 参数的转换分类乘积码锥形矢量量化[J]. 电子学报, 2008, 36(2): 362 - 366.
Li Hai-ting and Bao Chang-chun. Switched product code pyra-

mid vector quantizer for wideband ISF Parameters [J]. ACTA ELECTRONICA SINICA, 2008, 36 (2) : 362 - 366. (in Chinese)

- [6] Y Shin, S Kang, T R Fischer, C Son and Y Lee. Low complexity predictive trellis coded quantization of wideband speech LSF parameters [A]. Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 1) [C]. Montreal, Quebec, Canada: IEEE, 2004. 145 - 148.
- [7] W P LeBlanc, B Bhattacharya, S A Mahmoud, V Cuperman. Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding [J]. IEEE Trans on Speech and Audio Processing, 1993, 1(4) : 373 - 385.
- [8] Stephane Ragot, Roch Lefebvre, Redwan Salami, Jean-Pierre Adoul. Stochastic algebraic wideband LSF quantization [A]. Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 2) [C]. Istanbul, Turkey: IEEE, 2000. 1169 - 1172.
- [9] G Biundo, S Grassi, M Ansorge, F Pellandini and P A Farine. Design techniques for spectral quantization in wideband speech coding [A]. Proc of 3rd COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication [C]. Budapest, Hungary, 2002. 114 - 119.
- [10] 鲍长春, 戴逸松. 线谱对参数的一步差值预测矢量量化 [J]. 长春邮电学院学报, 1995, 13(4) : 1 - 7.
- [11] B Bessette, R Salami, R Lefebvre, M Jelinek, J Rotola Pukilla, J Vainio, H Mikkola, K Jarvinen. The adaptive multirate wideband speech codec (AMR-WB) [J]. IEEE Trans on Speech and Audio Processing, 2002, 10(8) : 620 - 636.
- [12] 孙圣和, 陆哲明. 矢量量化技术及应用 [M]. 北京: 科学出版社, 2002. 221 - 231.

作者简介:



李海婷 女, 1982年6月出生于北京, 硕士研究生, 研究方向为语音信号处理、宽带变速率语音编码等.

E-mail: lihaiting @emails. bjut. edu. cn



鲍长春 男, 1965年6月出生于内蒙古赤峰市, 博士, 教授、博士生导师, 国际语音通信学会 (ISCA) 会员, 中国电子学会理事, 信号处理学会委员, 《通信学报》编委会副主任委员, 《信号处理学报》与《数据采集与处理》编委. 主要研究领域为语音与音频编码.

E-mail: chchbao @bjut. edu. cn