

SELP 声码器基音周期参数量化合成改进算法^①计 哲^② 李 晔 崔慧娟 唐 昆

(清华大学电子工程系微波与数字通信技术国家重点实验室 北京 100084)

摘 要 针对正弦激励线性预测(SELP)声码器在解码端使用整型基音周期会引入额外舍入误差的问题,提出了一种码本整型化修正算法,该算法通过对码本中的各个码字在线性域进行取整,达到与解码端反量化过程一致的效果。测试结果表明,该算法能够有效地降低基音周期对数域矢量量化的量化误差,并且能够将合成语音的平均意见得分(MOS)提高 0.05 左右,明显改善合成语音质量。

关键词 语音编码,基音周期,矢量量化,特征参数

0 引 言

为了提高压缩效率,现有的低速声码器主要采用参数编码。它利用语音产生的数学模型,提取语音信号的特征参数,在解码端按照模型参数重构语音信号。其中,特征参数的提取和量化是影响声码器质量的关键因素^[1]。如何在保证语音质量的前提下进一步降低比特率是当今语音编码研究的主要焦点。因此,如何对特征参数进行有效地量化显得至关重要。基音周期是表征语音特征参数之一,它的倒数是声带振动频率,体现在听觉上就是语音音调的高低^[2]。在 2.4kb/s 正弦激励线性预测(sinusoidal excitation linear prediction, SELP)声码器算法^[3]中,基音周期参数采用逐帧的标量量化方式,在 1.2kb/s 及 0.6kb/s 等更低速率的 SELP 算法中,则采用超帧联合的矢量量化,且码本的训练和搜索在对数域进行。本论文在 1.2kb/s 及更低速率的 SELP 声码器算法的基础上,针对基音周期矢量量化中存在的问题,提出了码本的整型化修正算法。测试表明,该算法能够明显改善合成语音的质量。

本文首先分析了基音周期的参数特性和提取方法,对 SELP 声码器中基音周期的量化算法进行简单的介绍,随后着重介绍本文提出的码本整型化修正算法,并根据测试结果对算法的性能进行分析和评价。

1 基音周期的参数特征和提取方法

语音段根据是否具有周期性可以分为清音段和浊音段。发浊音时声带振动,基音周期即为发浊音时声带振动频率的倒数。它的检测和估计是语音处理中一个非常重要的问题,尤其在汉语语音处理中更是如此。因为汉语是一种有调语音,基音的变换模式称为音调,它携带着非常重要的具有辨意作用的信息^[4]。在低速率语音编码中,准确的基音估计非常关键,如波形差值(waveform interpolation, WI)和混合激励线性预测(mixed excitation linear prediction, MELP)算法中,基音周期的提取将直接影响到整个系统的性能^[5]。

SELP 算法采用自相关法提取基音周期参数。它用截止频率为 800Hz 的低通滤波器对信号进行滤波,并将语音信号经过二阶逆滤波后计算信号的自相关函数,最后采用前瞻一帧的动态规划算法对基音周期参数进行分析平滑,这里不做详细说明,具体方法见文献[6]。

自相关函数的计算公式如下:

$$r(m) = \frac{\sum_{n=0}^{N-1} e_{LP}(n) e_{LP}(n+m)}{[\sum_{n=0}^{N-1} e_{LP}(n)^2]^{1/2} [\sum_{n=0}^{N-1} e_{LP}(n+m)^2]^{1/2}} \quad m = 18 \sim 144 \quad (1)$$

其中, $e_{LP}(n)$ 为二阶低通逆滤波后的输出信号, N

① 国家自然科学基金(60572081)资助项目。

② 女,1984 年生,博士生;研究方向:低速率语音压缩编码;联系人, E-mail: jiz06@mails.tsinghua.edu.cn (收稿日期:2008-10-27)

为每一帧的采样点个数,这里取 200。基音周期将在自相关函数序列上若干个极大值点内选取^[6]。

2 SELP 声码器中基音周期量化算法

2.1 标量量化

SELP 声码器中基音周期的取值范围是 18 ~ 144 之间的整数,在 2.4kb/s 速率下采用标量量化,分配给该部分 7bit。量化方法是将区间 [18, 144] 线性映射到区间 [0, 127], 把 0 ~ 127 共 128 个数用 7bit 表示,计算关系如下

$$\text{index}_p = \lfloor (p - 18) \cdot \frac{127}{144 - 18} \rfloor \quad (2)$$

其中 p 为基音周期, $\lfloor \cdot \rfloor$ 为取整符号,由于 7bit 表示范围大于参数分布范围,所以基音量化可以将基音信息无失真传送^[6]。

2.2 矢量量化

在 1.2kb/s 及更低速率下, SELP 声码器基音周期参数是一个 M 维矢量,记为 $\mathbf{p} = [p_1, p_2, \dots, p_M]$, 各维的取值范围都是 [18, 144] 区间内的整数。为符合听觉对其进行对数域量化,相应对数域矢量记为 $\mathbf{p}_{\lg} = [\lg(p_1), \lg(p_2), \dots, \lg(p_M)]$ 。对数域基音周期采用矢量量化,码本搜索采用全搜索算法,失真度量采用加权欧式距离,表示如下:

$$d(\mathbf{p}_{\lg}, \hat{\mathbf{p}}_{\lg}) = \sum_{i=1}^M w_i (\lg(p_i) - \hat{p}_{\lg i})^2 \quad (3)$$

其中, $\hat{\mathbf{p}}_{\lg}$ 为码本中矢量, $\hat{p}_{\lg i}$ 为其第 i 个分量, M 为超帧中子帧的个数, w_i 为加权系数,由下式给出:

$$w_i = \begin{cases} 1, & \text{如果该子帧为浊音} \\ 0.1, & \text{如果该子帧为清音} \end{cases} \quad i = 1, 2, \dots, M \quad (4)$$

本文在对数域矢量量化的基础上,提出了基音周期码本整型化修正算法,接下来介绍该算法的主要内容。

3 码本整型化修正算法

由于基音周期的对数域失真度量更符合人耳听觉特性,所以被广泛应用于基音周期矢量量化中。然而, SELP 算法最终进行语音合成时使用整型基音周期,在解码端直接将反量化后基音周期转化为整型数,这将带来额外误差。为此,本文提出了基音周期的码本整型化修正算法,以尽可能消除该额外误差。

在 SELP 算法中,原始的基音周期矢量量化方法是:首先对基音周期 $\mathbf{p}$ 取对数 $\mathbf{p}_{\lg}$,然后在对数域进行码字搜索,按照公式(3)选取加权欧式距离最小的码字 $\mathbf{p}_i$,将码字的索引号 i 编码传输给解码端。解码端根据接收到的索引号找到对应的码字 $\hat{\mathbf{p}}_{\lg} = \mathbf{p}_i$,并反变换到线性域 $\hat{\mathbf{p}}$,最后对求取的数值进行取整操作。

但是,上述量化方法存在的问题。码字搜索算法是对于给定的输入矢量,在码本中搜索与输入矢量之间失真最小的码字,即找到码字 $\mathbf{p}_i$,使得 $d(\mathbf{p}_{\lg}, \mathbf{p}_i) = \min_{0 \leq j \leq N-1} d(\mathbf{p}_{\lg}, \mathbf{p}_j)$, N 为码本的尺寸^[7]。根据该码字搜索准则,易知对于任意 $j \neq i$, 都有 $d(\mathbf{p}_{\lg}, \mathbf{p}_i) \leq d(\mathbf{p}_{\lg}, \mathbf{p}_j)$ 。由于指数函数是单调的,因此可以保证对于任意 $j \neq i$, $d(2^{\mathbf{p}_{\lg}}, 2^{\mathbf{p}_i}) \leq d(2^{\mathbf{p}_{\lg}}, 2^{\mathbf{p}_j})$ 。但是,由于在线性域进行了取整操作,单调性被破坏,有可能存在 $j \neq i$, 使得 $d(2^{\mathbf{p}_{\lg}}, \lfloor 2^{\mathbf{p}_i} \rfloor) \geq d(2^{\mathbf{p}_{\lg}}, \lfloor 2^{\mathbf{p}_j} \rfloor)$, 即 $d(\mathbf{p}_{\lg}, \lg \lfloor 2^{\mathbf{p}_i} \rfloor) \geq d(\mathbf{p}_{\lg}, \lg \lfloor 2^{\mathbf{p}_j} \rfloor)$, 因此不能保证解码后的码字失真最小。为了解决这个问题,可以对码本进行整型化修正。它的原理是在码本训练结束后,再对码本中的各个码字进行线性域的整型化修正,即指数取整再取对数。然后将修正后的码本作为新的码本,与待量化的基音周期参数在指数域计算失真度量,从而保证了反量化后的参数与原始参数的失真最小。优化后的失真度量计算公式为

$$d(\mathbf{p}_{\lg}, \hat{\mathbf{p}}_{\lg}) = \sum_{i=1}^M w_i (\lg(p_i) - \lg \lfloor 2^{\hat{p}_{\lg i}} \rfloor)^2 \quad (5)$$

为了检验算法的性能,以 SELP 声码器算法为基础,在 0.6kb/s、0.8kb/s 以及 1.2kb/s 三个速率下,采用 4 个不同的测试文件分别对本文提出的算法进行了测试。其中,0.6 和 0.8kb/s 声码器将 3 个 25ms 的语音帧联合构成一个超帧,分别用 9bit 和 8bit 对超帧的基音周期参数进行矢量量化,而 1.2kb/s 声码器则采用两帧联合的方法,每个超帧的基音周期矢量用 9bit 量化。测试文件采用 8kHz 采样,16bit 量化,每个测试文件大约 3min,内容包含男女声普通话以及方言。测试指标采用平均意见得分 (mean opinion score, MOS),测试过程采用国际电信联盟 (International Telecommunication Union, ITU) 建议的 P. 862 MOS 分测试软件,测试结果如表 1 所示。从表中可以看出,采用整型化修正算法后,语音文件的 MOS 值平均提高了 0.05 左右,这些改善在低速声码器中是相当可观的。

表 1 算法改进前后 MOS 值比较

编码速率 (bps)	是否采用整 型闭环搜索	MOS 值			
		语音 1	语音 2	语音 3	语音 4
1200	否	3.024	2.967	3.023	2.993
1200	是	3.073	3.008	3.057	3.003
800	否	2.839	2.761	2.819	2.798
800	是	2.917	2.834	2.885	2.854
600	否	2.706	2.627	2.670	2.664
600	是	2.769	2.679	2.757	2.724

为了直观地反映出算法的性能,选取一段 20 帧测试序列,在 0.6kb/s 速率下,计算算法改进前后得到的基音轨迹与实际基音轨迹的比较,如图 1 所示。其中,码字误搜索标志位表示当前帧算法改进前后的基音周期码字索引是否一致,若码本整型化后搜索到的码字索引与原始算法码字索引不一致,码字误搜索标志位为 1,反之为 0。可以看出,在码字搜索正确的帧数范围内,算法改进前后的基音轨迹相同。在码字误搜索的帧数范围内,改进后的基音轨迹与实际的基音轨迹更加接近,量化误差相对较小。由于采用的是三帧联合的矢量量化模式,因此有个别帧数的量化基音周期与实际基音周期误差较大,但整体量化效果仍优于原始算法。

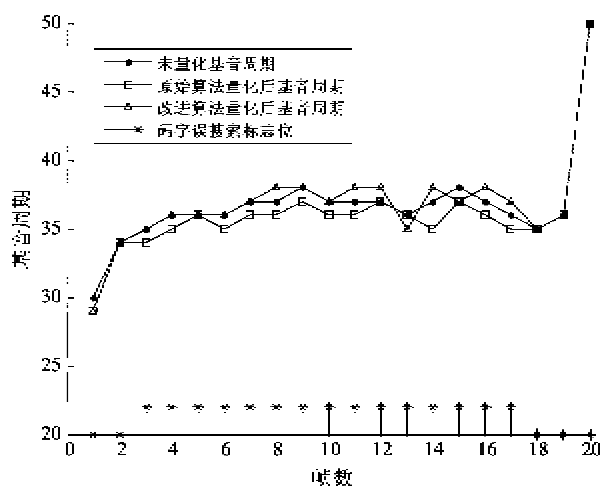


图 1 算法改进前后得到的基音轨迹与实际基音轨迹的比较

为了比较码本整型化修正算法在不同量化精度下的效果,以 0.8kb/s SELP 声码器为基础,计算当基音周期的量化比特数分别为 8、9、10bit 时,4 个语音测试文件基音周期量化的对数域误差,测试结果如表 2、3、4 所示。按式(3)计算量化误差如下:

$$err = \sum_{i=1}^N w_i (\lg(p_i) - \lg(\hat{p}_i))^2 \quad (6)$$

其中, p_i 为实际的基音周期, $\hat{p}_i$ 为量化后的基音周期, w_i 为式(4)所示的加权系数, N 为总帧数。

表 2 算法改进前后量化误差比较(基音周期以 8bit 量化)

	语音 1	语音 2	语音 3	语音 4
原始算法量化误差	18.314	16.557	18.142	16.848
改进算法量化误差	15.838	14.540	15.828	14.829
量化误差减小比率(%)	13.52	12.81	12.76	11.98
码字误搜索概率(%)	16.16	15.91	16.14	15.85

表 3 算法改进前后量化误差比较(基音周期以 9bit 量化)

	语音 1	语音 2	语音 3	语音 4
原始算法量化误差	12.279	11.697	12.397	11.441
改进算法量化误差	10.493	9.889	10.549	9.756
量化误差减小比率(%)	14.54	15.46	14.91	14.73
码字误搜索概率(%)	18.35	19.56	19.20	18.24

表 4 算法改进前后量化误差比较(基音周期以 10bit 量化)

	语音 1	语音 2	语音 3	语音 4
原始算法量化误差	9.592	8.504	9.323	8.399
改进算法量化误差	7.615	6.696	7.318	6.582
量化误差减小比率(%)	20.61	21.25	21.50	21.64
码字误搜索概率(%)	57.53	56.37	56.25	56.39

对比以上 3 个表格可以看出,随着量化比特数的增加,原始码本搜索算法的误搜索概率逐渐提高,采用改进算法后的量化误差减小比率逐渐增加。分析其原因主要有:(1)对于量化比较精细的码本,各个码字间的码距较小,码字整型化带来的码距的改变相对较大,因此误搜索概率也随之提高。(2)对于量化比特数较大的码本,其整体的量化误差较小,误搜索带来的失真占总失真的比重较大,因此算法改进后量化误差的减小相对较大。

4 结 论

本文针对 SELP 声码器基音周期量化过程中存在的不足,提出了一种码本整型化修正算法。它通过对码本在线性域进行取整达到了与反量化过程一致的效果,有效地减小了解码端取整带来的舍入误差。由于码本的整型化修正是离线操作的,因此不会对声码器算法造成运算量、存储量或延时等方面的额外开销。通过对算法改进前后的合成语音质量、基音周期轨迹以及量化误差三个方面的测试,可以验证本文提出的码本整型化修正算法能够明显改

善原有算法的不足,减小基音周期的量化误差,从而提高合成语音的质量。进一步分析表明,对于相同维数的基音周期参数,分配的矢量量化比特数越多,量化误差减小比率越大,改进效果越明显。本文提出的码本修正算法和码本训练过程是分离开来的,也可以考虑将码本训练和码本修正相结合,在码本训练时对失真度量进行相应的修改,从而达到最小化舍入误差的目的,这一方面有待更加深入的研究。

参考文献

- [1] Kondo A M. Digital Speech. Chichester: John Wiley & Sons Ltd, 2004
- [2] Barbara R, Mattias N, Anders E. Estimation of the instantaneous pitch of speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 2007, 15(3): 813-822
- [3] 江灏. 低速率语音编码的研究:[博士学位论文]. 北京:清华大学电子工程系, 1998
- [4] 鲍长春. 低比特率数字语音编码基础. 北京:北京工业大学出版社, 2001. 44-73
- [5] Hassan F. A novel method to improve pitch and voicing strength estimation for low bit rate speech coding. *Information and Communication Technologies*, 2006, 1: 1281-1286
- [6] 张建伟. 甚低速率语音编码算法研究:[博士学位论文]. 北京:清华大学电子工程系, 2002. 15-31
- [7] 孙圣和, 陆哲明. 矢量量化技术及应用. 北京:科学出版社, 2002. 260-297

An improved algorithm for pitch quantization in SELP vocoders

Ji Zhe, Li Ye, Cui Huijuan, Tang Kun

(State Key Laboratory of Microwave and Digital Communication, Department of Electronic Engineering,
Tsinghua University, Beijing 100084)

Abstract

To resolve the problem of the extra rounding error in sinusoidal excitation linear prediction (SELP) vocoders caused by turning the quantized pitch parameters into integers in the decoding end, this paper proposes an integer codebook modified method. It turns each codeword into integers in the linear-field, making the quantization consist with the decoding process. The simulation results show that the proposed algorithm can reduce the vector quantization error of the pitch parameters in the logarithm field, and increase the objective mean opinion scores (MOS) of the synthesized speech by 0.05, which obviously improves the quality of the speech.

Key words: speech coding, pitch, vector quantization, characteristic parameter