

低速率 WI 编码器中 4~6bit 基音量化算法研究

罗亚飞 鲍长春

(北京工业大学电子信息与控制工程学院 北京 100022)

摘要: 基音在语音编码中通常采用 7bit 无失真均匀量化。由于浊音段语音的基音普遍具有缓慢渐变的特点, 为了更有效地去除前后帧基音之间存在的相关性, 该文基于 Eriksson 和 Kang 提出的 4bit 基音量化算法, 针对汉语语音进行研究, 实现了一套 4-6bit 基音量化算法。该算法计算简单, 无需码书存储。将此基音量化方案应用于 WI 模型和 WI 编码器, 主观 A/B 听力测试结果表明, 该方案在高效量化基音的同时保证了合成语音质量几乎没有损失, 完全满足低速率 WI 编码器对量化基音的要求。

关键词: 语音编码; 基音量化; 波形内插

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2007)11-2669-03

Research on 4~6bit Pitch Quantization Algorithm in Low-Rate WI Coder

Luo Ya-fei Bao Chang-chun

(School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100022, China)

Abstract: Pitch parameter is often quantized uniformly with no distortion using 7 bit in speech coding applications. As pitch changes slowly in voiced segments, in order to remove the correlation between two consecutive pitch more effectively, based on Eriksson and Kang's 4 bit pitch quantization algorithm, this paper implements a pitch quantization algorithm using 4-6 bit for Chinese. This algorithm has a small complexity and no codebook is needed. This pitch quantization algorithm is applied to WI model and WI coder, the results from the subjective A/B listening test indicate that this algorithm not only quantized pitch effectively, but also kept the quality of synthesized speech with no obvious hearing distortion. Therefore, it satisfies the accuracy of the quantized pitch of WI coder completely.

Key words: Speech coding; Pitch quantization; Waveform interpolation

1 引言

基音周期在低速率语音编码中非常重要, 它的准确与否直接关系到合成语音的质量。基音周期一般为 60~400Hz, 8000Hz 采样率下即 20~147 个采样点^[1]。

基音周期的量化一般分为均匀量化与非均匀量化两种。由于基音周期极其重要, 低速率语音编码中通常采用 7bit 无失真均匀量化 128 个基音周期值。2.4kb/s MELP 声码器在对数域采用 7bit 均匀量化对数基音^[2]。另有一些低速率声码器结合自身情况采用 6~8bit 进行基音量化。比如, 一种非均匀基音量化方案利用“较大的基音量化时需要较低的绝对分辨率”的先验知识, 在规定的基音周期范围为 18~114 个样点的情况下, 用 7bit 量化基音。其中, 18~48 的基音值用 0.5 个样点精度量化, 49~114 的基音值用 1 个样点精度量化^[3]。本课题的波形内插(WI)编码器规定基音周期范围为 20~120 个样点, 在采用 7bit 无失真均匀量化方案的前提下得到了高

质量的合成语音^[4]。

由于浊音段语音的基音普遍具有缓慢变化的特点, 前后帧基音之间存在很大相关性。因此, 采用差分编码的思想对基音进行更有效的量化是合理可行的。为此, Eriksson 与 Kang 提出了一种 4bit 的基音量化算法^[5], 并在英语实验语音下得到了与 7bit 无失真均匀基音量化非常接近的性能。

本文围绕此问题, 基于 Eriksson 与 Kang 的思想, 针对汉语语音进行研究, 实现了一套 4-6bit 的基音量化算法。将此基音量化方案应用与 WI 模型及 WI 编码器后的实验结果表明, 该方案在高效量化基音的同时, 保证了合成语音质量几乎没有损失。

本文在第 2 节介绍 Eriksson 与 Kang 提出的 4bit 基音量化算法, 第 3 节介绍汉语语音基音量化参数的确定, 第 4 节介绍针对汉语的 4-6bit 基音量化方案及实验结果, 最后是结束语。

2 针对英语语音的 4 bit 基音量化

在绝大多数情况下, 浊音段基音缓慢渐变, 前后帧基音之间存在很大相关性, 如图 1 所示。并且研究表明, 人耳对浊音段基音的突变非常敏感^[6]。基于这样的先验知识, Eriksson 与 Kang 设计了一种 4bit 差分与对数基音量化器。其中, 差分

2006-05-08 收到, 2006-09-27 改回

国家自然科学基金(603720263), 北京市自然科学基金(4042009)和北京市教委科技发展计划项目(KM200710005001)资助课题

量化器用来量化缓慢变化的基音,同时,由于相对基音变化比绝对基音变化更重要,采用对数值进行量化更为有利。

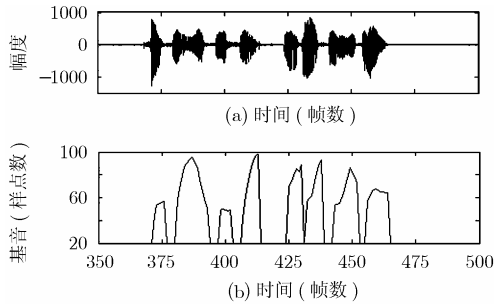


图 1 缓慢渐变的基音轨迹

为了进行更为有效的基音量化,该基音量化算法首先确定人耳对量化基音的敏感程度。然后,将基音分为高度相关与低度相关两种。在保证合成语音质量的前提下,用一个由差分量化器与无记忆均匀量化器组成的基音量化器对高度相关与低度相关基音在不同分辨率下分别进行量化。图 2 为该基音量化算法框图。

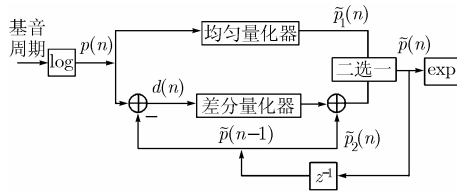


图 2 基音量化算法框图

量化过程如下:首先求取待量化基音的对数值 $p(n)$,然后将 $p(n)$ 分别送入均匀量化器与差分量化器。 $p(n)$ 在均匀量化器中被直接量化为 $\tilde{p}_1(n)$ 。在差分量化器中, $p(n)$ 先减去前一帧已量化的基音 $\tilde{p}(n-1)$ 形成差分 $d(n)$ 。然后对 $d(n)$ 进行量化并加上 $\tilde{p}(n-1)$ 形成差分量化器的输出 $\tilde{p}_2(n)$ 。比较 $\tilde{p}_1(n)$ 、 $\tilde{p}_2(n)$ 两个量化结果,取更接近待量化对数基音 $p(n)$ 的一个为最优的量化对数基音 $\tilde{p}(n)$ 。最后,对 $\tilde{p}(n)$ 做指数运算得到最终的基音量化值。

由于基音轨迹在经过一段平缓渐变后总是在幅度上有一个微小的上下跳动(如图 1 的 430 帧附近),这种跳动可能会被该基音量化算法放大到几个样点导致量化错误。因此,小的跳动消除技术被应用于该基音量化器的输出 $\tilde{p}(n)$ 。

3 汉语语音基音量化参数的确定

由于汉语是一种有调语音,因此,汉语语音的基音变化有其独有的特点。本节基于第 2 节的 4bit 基音量化算法,针对汉语语音进行分析实验,确定了针对汉语语音的基音量化参数。其中,基音估计均采用基于归一化互相关函数(Normalized Cross-Correlation Function, NCCF)的基音检测算法^[7]。该算法性能优异,很大程度上减少了清浊误判与基音加倍减

半现象。实验结果表明,NCCF 基音检测算法在信噪比较高的环境下能得到平滑准确的基音检测轨迹。

为了正确反映汉语语音的基音特征,本文选取约 80min 的汉语语音实验句子,其中包括成年男性 80498 帧、成年女性 79879 帧、男性儿童 10984 帧、女性儿童 11211 帧,全部数据为 8kHz 采样,16bit 量化的单声道语音。为了正确检测基音,实验采用 MCAMDF-NCCF 基音检测算法计算 80min 语音的基音周期。

图 3 为连续两帧基音周期的离散图,图中横纵坐标分别对应连续两帧的基音周期。由于浊音段基音高度相关,因此连续两帧基音在离散图上很大程度上聚集于直线 $y = x$ 附近的区域。

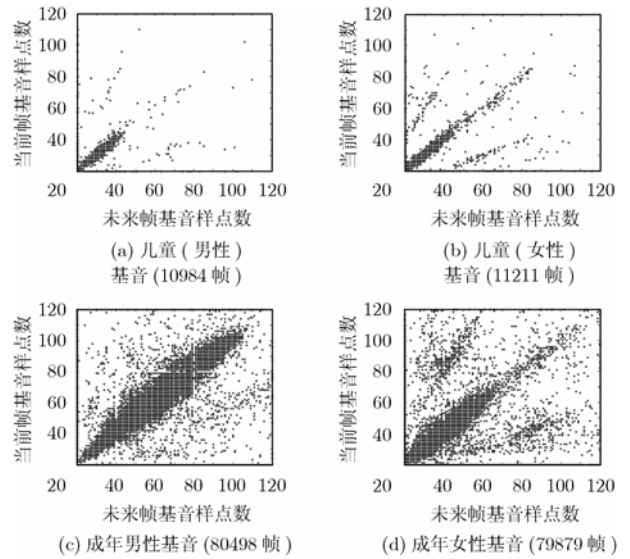


图 3 汉语连续两帧基音周期的离散图

图 4 给出了成年男性、成年女性基音周期统计直方图。实线与虚线分别对应男性与女性基音统计结果。从图中看出,在横轴为线性坐标的图 4(a)中,男性基音统计峰宽于女性基音统计峰。然而,基音的相对变化对于男性女性基音是一致的,这在横轴为对数坐标的图 4(b)中可以看出,因为此时男性与女性的基音统计峰宽度相当。此时男性与女性语音的平均基音周期分别为 54.2 与 37 样点,均值比英语语音偏低。

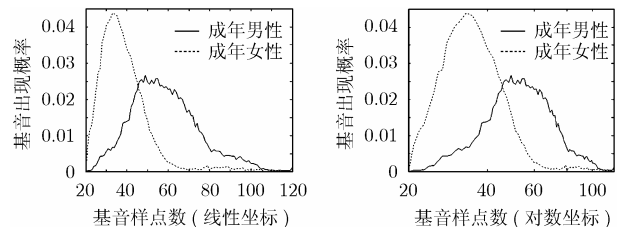


图 4 男性女性基音周期统计直方图

为了对高度相关基音高效量化,需确定高度相关基音的相关度范围。图 5 给出了成年男性、女性各约 80000 帧基音比率统计直方图。可以看出,男性、女性基音比率统计曲线

几乎重合。因此,本文结合实际情况,将高度相关基音相关度范围定为 0.84~1.2,高于英语基音相关度范围 0.9~1.1,图中 A, B 点为高度相关基音的相关度范围边界。

为了确定汉语语音中男性、女性基音量化的基音分辨率,本文在 WI 分析合成层进行实验。在 WI 分析器中分别将基音分辨率调至 2,3,⋯,8 样点精度,并在 WI 合成器测试不同基音分辨率下的合成语音质量,选取没有明显听觉失真时的最低基音分辨率做为基音量化的基音分辨率。大量实验结果表明,对于汉语而言,男性、女性基音量化的基音分辨率仍为 5 和 3 个样点,因此量化阶仍选为 $\log 1.06$ 。同样地,低度相关基音量化的基音分辨率为 10~15 个样点,与英语情况基本相同。

4 针对汉语的 4-6bit 基音量化方案及实验结果

综合以上针对汉语语音的基音量化参数分析,现将 4-6bit 基音量化方案介绍如下:首先,仍然采用 Eriksson 与 Kang 提出的差分与对数基音量化器。由于汉语高度相关基音的相关度范围为 0.84~1.2,明显高于英语基音的 0.9~1.1,在量化阶没有改变的前提下,需要增加量化码字的数量。

对于本文提出的 4bit 基音量化方案,7 个量化阶 $\log 1.06 \cdot \{-3, -2, -1, 0, 1, 2, 3\}$ 正好可以量化高度相关基音的相关度范围 $\log 0.84 \sim \log 1.2$,其余的 9 个码字均匀量化 20~120 个样点的基音完全合适。

对于 5bit 基音量化方案,将量化精度加倍,即用 13 个量化阶 $\log 1.03 \cdot \{-6, -5, \dots, 5, 6\}$ 量化高度相关基音的相关度范围 $\log 0.84 \sim \log 1.2$,其余的 19 个码字均匀量化 20~120 个样点的基音。

对于 6bit 基音量化方案,将量化精度继续加倍,即用 25 个量化阶 $\log 1.015 \cdot \{-12, -11, \dots, 11, 12\}$ 量化高度相关基音的相关度范围 $\log 0.84 \sim \log 1.2$,其余的 39 个码字均匀量化 20~120 个样点的基音。

本文选取男女各 8 句汉语语音实验句子,组织 10 名测试者进行主观 A/B 听力测试。将 4-6bit 汉语基音量化方案应用于 WI 分析合成层(即只量化基音,不量化其他参数),对合成语音进行主观 A/B 测试的结果表明,采用本文提出的 4-6bit 基音量化方案,与采用 7bit 无失真基音量化方案所得到的合成语音质量非常相近,没有明显的听觉失真。图 6 给出了基音量化器的失真曲线。

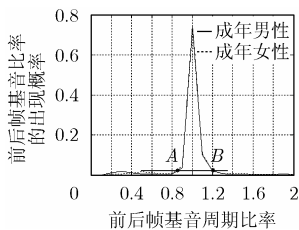


图 5 成年男性女性前后帧基音周期比率统计直方图

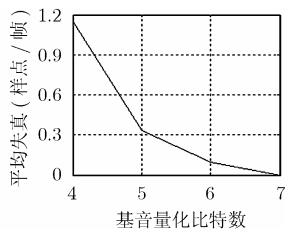


图 6 基音量化器的失真曲线

将 4bit 基音量化方案应用于 WI 编码器,与 7bit 无失真量化基音的原始 2kb/s WI 编码器^[4]合成语音质量进行主观 A/B 听力测试。从表 1 的测试结果可以看出,偏爱改进后 WI 算法的人数与偏爱原始 WI 算法的人数接近。同时,听不出 WI 算法改进前后差别的无偏爱人数也占有相当的比例,这充分说明 4-6bit 基音量化误差极小,足以满足低速率 WI 编码器对量化基音精度的要求。需要特别指出的一点是,该 4-6bit 基音量化方案还具有计算简单,无需码书存储的优势,完全可以应用于绝大多数低速率声码器中。

表 1 主观 A/B 听力测试结果(%)

纯净语音	改进 WI	原始 WI	无偏爱
男性语音	34.2	33.7	32.2
女性语音	28.2	28.5	43.3
所有语音	31.2	31.1	37.8

5 结束语

本文利用浊音段基音缓慢渐变、人耳对浊音段基音的突变非常敏感的特点,基于 Eriksson 与 Kang 提出的英语语音基音量化算法,针对汉语语音进行研究,实现了一套 4-6bit 基音量化算法。该算法计算简单,无需码书存储。将此基音量化方案应用与 WI 模型与 WI 编码器,主观 A/B 听力测试结果表明,该方案在高效量化基音的同时保证了合成语音质量几乎没有损失,更有效地去除了前后帧基音之间存在的相关性,完全满足低速率 WI 编码器对量化基音的要求。

参考文献

- [1] 鲍长春. 低比特率数字语音编码基础. 第 1 版. 北京: 北京工业大学出版社, 2001: 44-73.
- [2] Supplee L M, Cohn R P, and Collura J S. MELP: The new federal standard at 2400 bps. IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany, 1997, (2): 1591-1594.
- [3] 胡剑凌, 徐盛. 1.8kb/s 多带线性预测编码算法设计. 上海交通大学学报, 2002, 36(12): 1774-1777.
- [4] 李靛. 高质量的 2kb/s 波形内插语音编码算法研究. [博士论文], 北京工业大学, 2005: 14-114.
- [5] Eriksson T and Kang Hong-Goo. Pitch quantization in low bit-rate speech coding. IEEE International Conference on Acoustics, Speech, and Signal Processing, Phoenix, AZ, USA, 1999, (1): 489-492.
- [6] Zwicker E and Fastl H. Pitch and Pitch Strength. in Psychoacoustics, Springer-Verlag, New York, 1990: 103-132.
- [7] 鲍长春, 樊昌信. 基于归一化互相关函数的基音检测算法. 通信学报, 1998, 19(10): 27-31.

罗亚飞: 男, 1980 年生, 硕士生, 研究方向为低速率语音编码、语音邮寄系统等。

鲍长春: 男, 1965 年生, 博士, 教授, 博士生导师, 主要研究领域为数字信号处理与语音编码。