

基于 MELPe 模型的 600 BPS 声码器算法设计

邹峰, 郭英, 陈新富, 刘燕

(空军工程大学 电讯工程学院, 陕西 西安 710077)

摘要:在增强型混和激励线性预测(MELPe)模型的基础上,提出了一种高音质的600 bps声码器算法。保持MELPe算法特征的同时,利用帧间参数冗余,进行多帧联合量化;运用基于预测的分级矢量量化(PMSVQ)算法对线谱频率(LSF)参数进行量化。在非正式的主观语音质量测试中,合成语音质量优于传统的LPC10e声码器,接近2400 bps MELP标准的合成语音。

关键词:语音编码;线性预测;矢量量化

中图分类号: TN912.3 **文献标识码:** A **文章编号:** 1009-3516(2007)03-0070-04

1996年3月,美国国防部语音信号数字处理协会(DDVPC)选定混和激励线性预测(MELP)声码器取代原有的线性预测编码算法LPC10,作为新一代的2400 bps声码器标准。增强型混和激励线性预测(MELPe)在MELP的基础上增加了噪声预处理算法,使其能在不同的背景噪声的环境下得到高音质的合成语音^[1-3]。MELPe支持2400 bps和1200 bps两种速率,并被北大西洋公约组织选为语音编码标准STANAG 4591。

本文研究的600 bps声码器基于MELPe标准算法,重点从参数的量化编码入手,研究有效的码率压缩算法。4个连续语音帧组成一个超级帧进行联合矢量量化,将编码速率大幅度压低而保持较高清晰度。线谱频率(LSF)是线性预测声码器编码传输的主要参数,利用其帧间相关性,提出基于预测的分级矢量量化(PMSVQ)算法。两种模式的滑动平均(MA)预测器^[4-5]用来去除LSF参数帧间相关,采用分级矢量量化^[6-7]方法对预测残差进行量化编码。MA预测器能有效地减小LSF参数的动态范围,在相同比特率下获得更小的量化失真,同时控制信道误码扩散。PMSVQ算法能在极低的码率下获得接近“透明”的量化效果^[7],为高品质低码率声码器打下了基础。根据超帧中不同的清浊音判决组合运用不同的编码方案,进一步消除帧间冗余。

1 声码器概述

基于MELPe算法的600 bps声码器,采用与MELPe基本相同的语音合成和分析方法。语音信号采样率为8000 Hz,帧长为25 ms,4帧组成一个超级帧,算法时延为100 ms。在600 bps速率下,平均每帧用15 bit来量化,每一超级帧量化60 bit。

在标准的MELPe算法中,LSF参数、基音周期、清浊判决、增益、傅立叶级数和非周期脉冲标志6种参数被编码传输。由于比特率的限制,600 bps声码器选择前4种参数进行编码量化,傅立叶级数和非周期脉冲标志这两种参数则通过在译码端得到的参数进行一定程度的恢复。傅立叶级数用两个矢量表示,如果合成语音段为浊音,采用浊音残差的一个典型矢量来表示傅立叶级数;若合成语音段为清音,则采用一个平坦的矢量来表示。语音的非周期脉冲一般出现在语音信号的过渡区域,在译码端当出现清音到浊音,或浊音到清音的过渡区时,设置非周期脉冲标志。

收稿日期:2006-09-29

基金项目:陕西省自然科学基金项目资助(2006F40)

作者简介:邹峰(1982-),男,湖南浏阳人,硕士生,主要从事语音信号处理和自适应信号处理;

郭英(1961-),女,山西大同人,教授,博士生导师,主要从事通信信号处理和自适应信号处理。

2 多帧联合矢量量化和比特分配

2.1 基于 PMSVQ 算法的 LSF 参数量化

采用 Levinson - Durbin 算法计算哈密窗加权信号的 10 阶线性预测系数,再用 Chebyshev 多项式递推算法将线性预测系数转化为线谱频率参数。LSF 参数本身具有的特征使其成为量化线性预测系数最常用的形式,在 2 400 bps MELP 算法中每一帧用 25 bit 对其进行矢量量化,取得了良好的效果。在 600 bps 声码器算法中,平均每帧只能用 15 bit 来量化,用于量化 LSF 参数的比特数就更少了。提出一种基于预测的分级矢量量化(PMSVQ)算法,将 4 帧连续的 10 维 LSF 参数组合成一个参数矩阵,通过帧间 MA 预测器利用前一矩阵对当前矩阵参数进行预测,再将预测残差矩阵进行多级矢量量化传输。使用 MA 预测器一方面减小 LSF 参数的动态范围,有利于之后的量化编码;另一方面克服了传统预测器所存在的信道误码扩散问题,控制信道误码影响在其预测器阶数之内。MA 预测器阶数为 2,具有两组预测系数,一组常用在语音帧的过渡区;另一组则多用于语音信号的稳定区域。预测残差矩阵采用级数为 4 的矢量量化算法进行量化。

设输入的 LSF 参数为 $\omega_{i,j}$, $\hat{\omega}_{i,j}$ 为量化输出的 LSF 参数, i 表示超帧序号, j 为超帧中子帧的序号。量化输出的 LSF 参数由下式得到。

$$\hat{\omega}_{i,j} = P_{0,j} \hat{r}_{i,j} + \sum_{k=1}^2 P_{k,j} \hat{r}_{i-1,5-k} \quad j = 1, 2, 3, 4, \quad k = 0, 1, 2 \quad (1)$$

其中 \hat{r}_i 和 \hat{r}_{i-1} 分别为当前第 i 超帧和前一超帧的量化 LSF 预测残差; $P_{k,j}$ 为预测系数矩阵,由式(2) 给出。从式(1) 可以看出,当前超帧 LSF 参数由前一超帧 LSF 预测残差对当前帧的预测值,和当前超帧的 LSF 预测残差两部分组成。MA 预测器阶数为 2,所以前一超帧最后两子帧的 LSF 预测残差被用来预测当前超帧。预测系数矩阵有两组,预测误差最小的那组系数将用来对当前超帧进行预测。用 1 bit 来选择预测系数矩阵,每一组预测系数均满足下式:

$$P_{k,j} = \text{diag}\{p_{k,j}^1, p_{k,j}^2, \dots, p_{k,j}^{10}\} \quad (2)$$

$$\sum_{k=1}^2 P_{k,j} = I \quad j = 1, 2, 3, 4, \quad k = 0, 1, 2 \quad (3)$$

其中预测系数矩阵 $P_{k,j}$ 为对角矩阵,用来对当前超帧中第 j 子帧进行预测; I 为单位矩阵; k 表示预测器阶数。使用 Generalized Lloyd 算法^[8] 来训练这两组预测系数矩阵,对应每一组预测系数训练一组码本。谱失真作为失真测度,定义式如下:

$$SD = \left[\int_0^\pi \left(10 \lg \frac{S_{i,j}(\omega)}{\hat{S}_{i,j}(\omega)} \right)^2 \frac{d\omega}{\pi} \right]^{1/2} \quad (4)$$

式中 $S_{i,j}(\omega)$ 和 $\hat{S}_{i,j}(\omega)$ 分别为量化前后第 i 超帧中第 j 子帧的功率谱函数。使谱失真最小的 LSF 预测残差码本序号将被编码传输。

选用尺寸为 1 024, 512, 512, 256 的码本,对预测残差进行级数为 4 的分级矢量量化,另用 1 bit 标志所选 MA 预测器组,对应每一组预测系数有相互独立的一组码本,一共 37 bit 量化 LSF 系数。码本搜索采用 M - Best 算法^[7], $M = 8$ 的多路径搜索算法逼近全局最优。

表 1 列出了基于 PMSVQ 算法的 LSF 系数量化失真情况。通过去除 LSF 参数帧间冗余,能在极低的码率下获得接近“透明”的量化,为取得高质量的合成语音打下了基础。

表 1 基于 PMSVQ 算法的 LSF 参数量化失真

平均 SD/dB	4 > SD > 2	SD > 4
1.18	5.14%	0.1%

2.2 多带清浊音参数联合矢量量化

在 2 400 bps MELPe 标准算法中,每一语音帧提取 5 个子带的清浊音判决,基带判决通过基音参数传输,另用 4 bit 表示剩余的子带判决。提出的 600 bps 声码器把 5 个子带的清浊音参数当成 5 维矢量,4 个连续帧组成一矩阵。 $b_{i,j}$ 表示第 i 超帧,第 j 个子带的清浊音判决。 $b_{i,j} = 0$ 表示为清音, $b_{i,j} = 1$ 表示为浊音。采用加权欧氏距离为失真测度:

$$d = \sum_{i=1}^4 \sum_{j=1}^5 w_j (b_{i,j} - \hat{b}_{i,j})^2 \quad (5)$$

其中 w_j 为加权量,通过训练得到,根据语音特性,频带越低权重越大。采用上述失真测度,选取尺寸为 16 的码本,即用 4 bit 矢量量化超帧的清浊判决参数。

2.3 基于清浊音判决的基音周期联合矢量量化

根据清浊音判决情况,4 个连续语音帧的基音周期参数选用不同的编码方案。当 4 个连续语音帧的基带清浊音判决均为清音时,不对基音周期参数编码传输,其比特用来做信道纠错编码;若 4 个连续语音帧的基带清浊音判决中仅有一帧为浊音时,采用 2 400 bps MELPe 标准相同的基音周期编码算法,用 7 bit 量化浊音的基音参数,其他清音帧则不进行编码量化;如果 4 个连续语音帧的基带清浊音判决有两帧或两帧以上浊音时,则将 4 帧基音参数组成 4 维矢量,用文献[9] 提出失真测度进行联合矢量量化。失真测度定义如下:

$$d = \sum_{i=1}^4 w_i |p_i - \hat{p}_i|^2 + \delta \sum_{i=1}^4 |\Delta p_i - \Delta \hat{p}_i|^2 \quad (6)$$

其中 i 表示超帧中的第 i 个子帧; p_i 和 \hat{p}_i 分别为量化前后的对数域基音值; w_j 为加权量,由式(7)给出; Δp_i 和 $\Delta \hat{p}_i$ 表示量化前后浊音帧之间的对数基音值的变化量,式(8)给出了其定义:

$$w_i = \begin{cases} 1 & \text{浊音帧} \\ 0.1 & \text{清音帧} \end{cases} \quad (7) \quad \Delta p_i = \begin{cases} p_i - p_{i-1} & i \text{ 和 } i-1 \text{ 帧均为浊音时} \\ 0 & \text{其它} \end{cases} \quad (8)$$

p_0 定义为前一超帧最后一子帧的对数基音值; δ 在算法中设置为 1。这一测度既考虑了整个超帧的量化失真,又保留帧间变化信息。对数基音值选取码本尺寸为 512,即 9 bit 来联合矢量量化,对于前两种情况未使用的比特则用来进行信道纠错编码。

2.4 增益联合矢量量化

每一语音帧提取 2 个增益参数,4 个连续帧增益参数组成 8 维矢量 $G = \{G_1, G_2, \dots, G_8\}$; 首先用 5 bit 标量量化其能量均值 $\bar{G} = (G_1 + G_2 + \dots + G_8)/8$,然后将能量与能量均值的比值 $\Delta G = G/\bar{G}$,用 4 bit 进行矢量量化,以欧氏距离作为失真测度,一共用 9 bit。算法兼顾了动态范围和变化特性,取得了好的量化效果。

2.5 量化比特分配

表 2 列出了 600 bps 声码器超帧的比特分配情况。每一超帧持续时间为 100 ms,量化 60 bit,即速率为 600 bit/s。

表 2 600 bps 声码器比特分配

	LSF 参数	MA 模式标志	基音周期	清浊参数	增益	同步	共计
比特数	36	1	9	4	9	1	60

3 仿真测试及分析

选择诊断押韵测试(DRT - Diagnostic Rhyme Test)和诊断可接受度测试(DAM - Diagnostic Acceptability Measure)来做语音质量主观听觉测试。DRT 主要用来评价合成语音的可懂度,DAM 则用来评价语音质量。实验室条件下录制的相对纯净语音作为测试语音,其中男生 4 名,女生 3 名。在安静背景和 1% 随机误码信道的仿真条件下做了仿真测试,表 3 和表 4 分别给出了其在非正式主观听觉测试的结果,为了比较给出了标准 MELPe 的测试结果。

表 3 安静背景主观听觉测试结果

测试项目	DRT	DAM
原始语音	95.4	84.5
标准 2 400 bps MELPe	93.2	66.9
提议的 600 bps 声码器	90.8	55.4

表 4 1% 随机误码信道主观听觉测试结果

测试项目	DRT	DAM
原始语音	95.4	84.5
标准 2 400 bps MELPe	91.0	52.7
提议的 600 bps 声码器	86.5	45.6

从测试结果可以看出,通过将连续语音帧组成超帧进行联合量化,利用帧间参数固有的冗余,在大幅度压缩编码速率的条件下,获得高质量的合成语音。

4 结束语

本文提出一种基于 MELPe 模型的 600 bps 声码器,主要阐述了参数的量化编码算法。LSF 参数量化运用了 PMSVQ 算法,在极低的码率下获得接近“透明”的量化效果;基于基带清浊音判决的基音周期量化算法,根据清浊音分布选择不同的量化方案;通过编码端参数部分地恢复在编码端未编码的参数。这一系列新技术的运用,使其在 600 bps 速率下获得高音质的合成语音。仿真结果显示,其非正式主观听觉测试效果优于传统的 LPC10e 声码器,接近 2 400 bps MELP。

参考文献:

- [1] Collura J S, Brandt D F. The 1.2kps/2.4kpbs MELP Speech Coding Suite with Integrated Noise Pre - Processing[A]. IEEE Mil Comm[C]. Atlantic, 1999. 144 - 145.
- [2] Tremain T E. The Government Standard Linear Predictive Coding Algorithm: LPC - 10[J]. Speech Technology, 1982, 1(2): 40 - 49.
- [3] McCree A, Truong K. A 2.4 kbit/s MELP Coder Candidate for The New U. S. Federal Standard[A]. Proceedings of IEEE ICASSP 1996[C]. Piscataway, New Jersey, 1996. 200 - 203.
- [4] Salami R, Laflamme C. Design and Description of CS - ACELP: A Toll Quality 8kb/s Speech Coder[J]. IEEE Transactions on Speech and Audio Processing, 1998, 6(2): 116 - 130.
- [5] Koishida K, Cuperman V. Enhancing MPEG - 4 CELP by jointly optimized inter/intra - frame LSP predictors[A]. IEEE Workshop on Speech Coding[C]. Delavan, USA, 2000. 90 - 92.
- [6] Chan W Y, Gupta S. Enhanced Multistage Vector Quantization by Joint Codebook Design[J]. IEEE Transactions on Communications, 1992, 40(11): 1693 - 1697.
- [7] LeBlanc W P, Bhattacharya B. Efficient Search and Design Procedures for Robust Multi - Stage VQ of LPC Parameters for 4 kb/s Speech Coding[J]. IEEE Transactions on Speech and Audio Processing, 1993, 1(4): 373 - 385.
- [8] Lloyd S P. Least Squares Quantization in PCM[J]. IEEE Trans. Inform Theory, 1982, 28(2): 129 - 137.
- [9] Wang Tian, Koishida K. A 1 200 bps Speech Coder Based on MELP[A]. IEEE. ICASSP 2000[C]. Piscataway, New Jersey, 2000. 1375 - 1378.
- [10] 王博, 郭英, 李宏伟, 等. 基于倒谱距离的语音端点检测改进算法[J]. 空军工程大学学报: 自然科学版, 2006, 7(1): 59 - 63.

(编辑: 门向生)

A Design of 600 bps Speech Coding Algorithm Based on MELPE

ZOU Feng, GUO Ying, CHEN Xin - fu, LIU Yan

(The Telecommunication Engineering Institute, Air Force Engineering University, Xi'an 710077, China)

Abstract: This paper describes a new speech coder at 600 bps based on the enhanced mixed excitation linear prediction (MELPE) algorithm, which is the new NATO standard STANAG 4591. The algorithm of this speech coder includes features of MELPE. Multi - frame joint vector quantification is performed, which takes advantage of inherent inter - frame redundancy. The predicted multi - stage vector quantization (PMSVQ) algorithm is designed to quantize the line spectrum frequency (LSF) parameters. The informal subjective quality tests show that the speech quality of the proposed coder is better than that of LPC10e and approximates to 2.4 kbps MELP standard.

Key words: speech coding; linear prediction; vector quantization