

800/920/1200bps改进型多带激励语音编码器的DSP实现

王都生

(大连民族学院机电与控制工程研究所 大连 116600)

摘要: 该文介绍了基于多带激励(MBE) 语音模型的改进型800/920/1200bps 语音编码器。该编码器采用二级离散余弦变换编码方案对逐帧变化的谱幅度参数进行自适应动态量化编码,大大减小了编码比特率,使得800~1200bps速率的MBE语音编码器的实现成为可能。该文重点介绍二级离散余弦变换编码方案及DSP硬件实现。

关键词: 语音编码; 多带激励; 离散余弦变换

中图分类号: TN912.32

文献标识码: A

文章编号: 1009-5896(2007)05-1106-04

Implementation of the 800/920/1200bps Modified Multi-band Excitation Speech Coder Based on DSP Processor

Wang Du-sheng

(Institute of Electromechanical & Control Engineering, Dalian Nationalities University, Dalian 116600, China)

Abstract: An 800/920/1200bps modified speech coder based on the Multi-Band Excitation (MBE) speech model is presented. The two stage discrete cosine transformation coding scheme has been used to quantize the spectral magnitude parameters changing frame-by-frame adaptively in the speech coder. The coded bit rate of speech is largely reduced and that make it possible the implementation of 800~1200bps MBE speech coder. This paper focuses on the design of two stage discrete cosine transformation coding scheme and hardware implementation based on DSP processor.

Key words: Speech coding; Multi-band excitation; Discrete cosine transformation

1 引言

低比特率语音编码系统有很多应用。传统的编码方法在传输比特率较低并逐步减小时,语音质量下降很快。尽管人们在激励信号的描述和量化上作了很多工作并且改善了语音质量,但也相应地增加了这些编码算法的复杂度。

Lim 提出的多带激励(MBE)语音模型^[1]是从频域角度来描述激励信号的。它将一帧语音的频谱按基频分成多个谐波频带,在每个频带上进行清浊判别,从而更准确地表达了语音特征信息,其结果在很大范围内改善了语音质量,而算法的复杂度并没有明显增加。本文在实现2400 bps多带激励声码器的基础上,设计了一个二级离散余弦变换(DCT)的编码方案。采用该方案对逐帧变化的谱幅度参数进行自适应动态量化编码,大大减小了编码比特率,使得800~1200bps速率的MBE语音编码器的实现成为可能。下面介绍MBE 语音分析合成算法,重点介绍二级离散余弦变换编码方案及DSP硬件实现。

2 MBE 语音分析合成算法

MBE 语音编码模型所用到的参数包括基音频率、每一频带上的U/V 信息、谱包络参数(浊音频带谱幅度及相位、

清音频带谱幅度)。

2.1 语音分析

MBE分析端的框图如图1所示。在估计这些参数时,文中采用文献[2]中所提到的一种次优的算法分两步来提取。第1步采用与分析端合成(A-B-S)类似的方法,在使原始语音谱和合成语音谱之间的加权误差为最小的准则下,确定基音频率和每个分带的谱包络参数;第2步再利用每个谐波处合成谱与原始语音谱拟合的程度来对每个分带进行U/V 判决。其具体步骤如下:

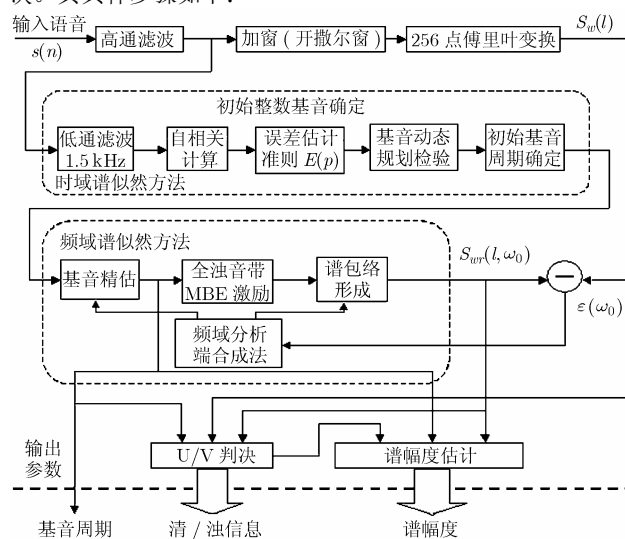


图1 MBE分析端框图

2005-10-08 收到, 2006-03-14 改回

国家民委科研项目“低速率语音编码器的 DSP 设计与实现”资助课题

(1)加开撒尔(Kaiser)窗取一帧语音,在时域按归一化无偏拟合误差准则^[3]粗略估计基音周期 P 。

(2)参考前两帧的 P_0 值及相应的误差值,采用动态规划方法^[4]对 P 进行非线性平滑,以更准确地选择当前帧的初始基音周期 P_I 值。

(3)在 $P_0 = P_I \pm 1/4, P_I \pm 2/4, P_I \pm 3/4$ 等点上用频域谱加权的总拟合误差公式:

$$\varepsilon(\omega_0) = \sum_{l=1}^{127} G(l) |S_w(l) - S_{wr}(l, \omega_0)|^2 \quad (1)$$

进行细搜索,确定更精确的基音周期 P_0 。式(1)中 $S_w(l)$ 为加窗原始语音信号的256点DFT谱, $S_{wr}(l, \omega_0)$ 为合成信号谱, $G(l)$ 为频率加权函数,在低频端取0,高频端取1。最佳基音角频率 $\omega_0 = 2\pi/P_0$ 应使加权总拟合误差 $\varepsilon(\omega_0)$ 为最小。

(4)利用在每个谐波处合成谱与加窗语音谱拟合的程度来确定U/V的判决。如归一化拟合误差小于一定的门限值,则判该谐波频带为浊音区,否则为清音区。由于在低速率语音编码中没有足够的比特来表示每一个谐波频带的U/V,所以可将几个相邻的谐波频带合并为一个带,根据该带总的拟合误差做出该带的U/V判决。在800~1200bps编码方案中,8个相邻的谐波频带合并为一个带,最多分成3个带(24个谐波)进行U/V判决,超过24个谐波的频带包含在最高带内一起进行U/V判决。U/V确定之后,浊音带的谱幅度和相位及清音带的谱幅度也就确定了。

2.2 语音合成

MBE合成端的框图如图2所示。本文对于浊音部分的语音采用时域合成法^[5],该方法具有允许帧与帧之间的基频连续变化的特点,从而使合成的波形效果更好。而对于清音部分的语音,采用频域短时傅氏变换(STFT)法合成,因为带通滤波器在频域中易于实现,用FFT实现,计算复杂度更低。

浊音部分的语音合成:首先对各谐波频率处的幅度加以修正,若第 m 次谐波属浊音频带,则合成语音模值 $M_{mq} = x_{mq}$ (x_{mq} 为其量化后的幅度模值);否则, $M_{mq} = 0$;然后进行插值计算:

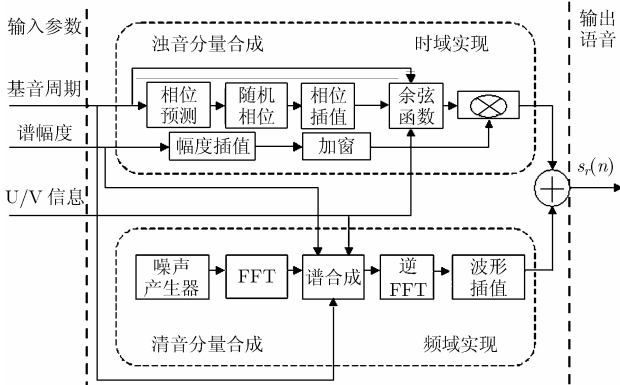


图2 MBE合成端框图

$$M_{mq}(n, 0) = \frac{N-n}{N} M_{mq}(-1) + \frac{n}{N} M_{mq}(0), \quad 0 \leq n \leq N \quad (2)$$

式(2)中(-1)表示前一帧的参数,(0)表示当前帧参数, N 为帧长。这样合成语音的浊音部分 $s_{rv}(n)$ 为

$$s_{rv}(n) = \sum_m M_{mq}(n) \cos \theta_m(n), \quad 0 \leq n \leq N \quad (3)$$

式(3)中的相位函数 $\theta_m(n)$ 由初始相位以及插值角频率计算得到^[5]。

清音部分的语音合成:首先产生一个均值为零、频谱密度值为1的白噪声序列 $u(n)$,对它加窗后进行256点的DFT,得到 $U_w(l)$ 。然后根据谐波频带 m 是处于浊音区还是清音区,得到修正后的 $U_{wc}(l)$ 如下:

若 m 频带为浊音频带,则

$$U_{wc}(l) = 0, \quad a_m \leq |l| \leq b_m \quad (4)$$

否则

$$U_{wc}(l) = x_{mq}(0) U_w(l) / \sqrt{\sum_{a_m}^{b_m} |U_w(k)|}, \quad a_m \leq |l| \leq b_m \quad (5)$$

式(4),式(5)中 a_m 和 b_m 是第 m 次谐波的频带下限和上限。对 $U_{wc}(l)$ 进行IDFT产生时间序列 $u_{wc}(n)$ 。为了使合成语音连续,还要与前一帧的清音序列做叠接处理。若本帧的清音序列用 $u_{wc}(n, 0)$ 表示,前一帧的清音序列用 $u_{wc}(n, -1)$ 表示,那么当前帧合成语音的清音部分为

$$s_{ru}(n) = \frac{w(n)u_{wc}(n, -1) + w(n-N)u_{wc}(n-N, 0)}{w^2(n) + w^2(n-N)}, \quad 0 \leq n \leq N \quad (6)$$

其中 $w(n)$ 为窗函数, N 为帧长。

在分别求出浊音部分和清音部分之后,最后的合成语音 $s_r(n)$ 为

$$s_r(n) = s_{rv}(n) + s_{ru}(n), \quad 0 \leq n \leq N \quad (7)$$

3 800/920/1200 bps 改进型MBE参数编码方案

在MBE模型中,谐波数目是逐帧变化的,所以每帧的谱幅度参数也在变化,因此谱幅度参数占用比特数较多,也是最量化的。本文设计了一个二级离散余弦变换(DCT)自适应动态比特分配的量化编码方案,用该方案可动态地对逐帧变化的谱幅度参数进行量化编码。在800/920/1200bps编码速率中,由于没有更多的比特数来量化浊音频带的相位信息,因此在浊音语音合成时,文中采用了相位预测、随机相位、相位插值等处理技术来模拟浊音相位的轨迹。表1是800/920/1200bps编码速率的比特分配。下面介绍各参数的量化编码。

表1 800/920/1200bps编码速率的比特分配

参数	800bps的量化比特数	920bps的量化比特数	1200bps的量化比特数
基音频率	6	7	7
U/V信息	$K(\leq 3)$	$K(\leq 3)$	$K(\leq 3)$
谱幅度	动态分配	动态分配	动态分配
帧编码数	20	23	30

3.1 基音频率量化编码

由频域谱似然法确定的最佳基音频率是用小数表示的,在量化时利用下式:

$$P_q = \left\lfloor \frac{4.0}{f_0} - 78.0 \right\rfloor \times 2^{-2} \tag{8}$$

变换为整数,再用7bit直接量化编码。6bit基音量化取中间值。基音频率解码用下式进行:

$$f_0 = \frac{1.0}{P_q + 19.0 + 0.5} \tag{9}$$

3.2 U/V信息编码

从MBE 语音分析中我们知道,在800~1200bps 编码方案中,8个相邻的谐波频带合并为一个带,最多分成3个带(24个谐波)进行U/V判决,超过24个谐波的频带包含在最高带内一起进行U/V判决。所以分带数目 $K \leq 3$,用1表示浊音带,0表示清音带,直接编码。

3.3 二级离散余弦变换的谱幅度量化编码

首先将 M (在语音分析时限定在11~59之间)个谐波谱幅度值变换到对数域,然后进行二级DCT变换,最后对变换后的DCT系数进行自适应动态比特分配再量化编码。

将每帧的 M 个谐波对数谱幅度值按下式约等分为6块:

$$M = (6 - M\%6) \times \lfloor M/6 \rfloor + M\%6 \times (\lfloor M/6 \rfloor + 1) \tag{10}$$

式(10)中 $M\%6$ 表示取余数, $\lfloor M/6 \rfloor$ 表示下取整。这样可将 M 个谱幅度参数分成 $(6 - M\%6)$ 块,每块含 $\lfloor M/6 \rfloor$ 个参数和 $M\%6$ 块,每块含 $(\lfloor M/6 \rfloor + 1)$ 个参数。第1级对每块分别进行 $\lfloor M/6 \rfloor \times 1$ 和 $(\lfloor M/6 \rfloor + 1) \times 1$ 长度的DCT变换;每块变换后的第1个系数再组合起来,进行第2级固定长度的 6×1 DCT 变换。二级变换后的系数进行比特分配及量化编码。

以 $M=26$ 为例,这时余数为2,商为4。这样26个谱幅度参数可以分成4块(每块4个谱幅度参数)和2块(每块5个谱幅度参数)大约相等的6块参数。其量化编码示意图如图3所示。

设谐波对数谱幅度为 $A_i, i = 1, 2, \dots, 26$, 第1级DCT变换后的系数为 $A_{1,i}, i = 1, 2, \dots, 26$, 第2级DCT变换后的系数为 $A_{2,i}, i = 1, 2, \dots, 6$ 。

对 $A_{2,1}$ 固定分配6bit,用64阶标量量化器量化编码。对

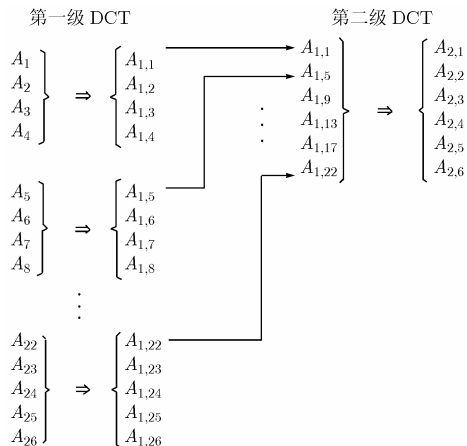


图3 谱幅度参数量化编码示意图($M=26$)

$A_{2,i}, i = 2, \dots, 6$ 采用动态分配比特数的方案。根据DCT变换的性质,变换后的能量都集中在较低序号的系数上,因此应给较低序号的系数分配更多的比特数。

在1200 bps中,给 $A_{2,i}, i = 2, \dots, 6$ 5个系数预先分别分配 2, 2, 1, 1, 1bit, 再将 $\lfloor 90/M \rfloor$ 个比特数按 $A_{2,i}, i = 4, 5, 6, 2, 3$ 的次序依次增加1bit, 增完为止。这时剩余比特数为

$$NUM = 30 - 7 - 6 - 7 - \lfloor 90/M \rfloor - K \tag{11}$$

这些比特数按0.9: 0.9: 0.7: 0.5: 0.4: 0.4的比例依次分配给第1级DCT变换后的每块系数,即第一块分得的比特数为

$$NUM_1 = NUM \times [(0.9 \times m_1) / (0.9 \times m_1 + 0.9 \times m_2 + 0.7 \times m_3 + 0.5 \times m_4 + 0.4 \times m_5 + 0.4 \times m_6)] \tag{12}$$

式中 m_1, \dots, m_6 分别为每块的谐波个数。第1块分配后的剩余比特数为 $NUM = NUM - NUM_1$, 那么第2块分得的比特数为

$$NUM_2 = NUM \times [(0.9 \times m_2) / (0.9 \times m_2 + 0.7 \times m_3 + 0.5 \times m_4 + 0.4 \times m_5 + 0.4 \times m_6)] \tag{13}$$

依次类推...

每块内的系数再按照分得的比特数和块内的谐波个数进行自适应动态比特分配,以确定第1级DCT变换后的每个系数的比特分配,这个过程就不详述了。

在920 bps中,给 $A_{2,i}, i = 2, \dots, 6$ 5个系数分别分配2, 2, 1, 1, 1bit, 基音用7bit量化, $A_{2,1}$ 用6bit量化,其它系数的比特分配与1200 bps的相同。

在800 bps中,给 $A_{2,i}, i = 2, \dots, 6$ 5个系数分别分配2, 2, 1, 1, 1bit, 基音用6bit量化, $A_{2,1}$ 用5bit量化,其它系数的比特分配与1200 bps的相同。

分配好比特数后,每个系数按照所得的比特数乘以不同的量化参考值后,再进行均匀量化。量化表的范围及量化参考值是经大量的语音数据模拟得到的。

在参数解码时,先对基音频率 f_0 进行解码,那么在0~4000Hz范围内有 $M = \lfloor 1/(2f_0) \rfloor$ 个谐波谱幅度参数。根据 M 值在解码端进行与编码端相同的自适应动态比特分配,确定各个DCT系数的编码比特率,依次查表解码,没有分配到比特数编码的DCT系数用零填补;然后进行第2级逆离散余弦变换(IDCT),再进行第1级IDCT;最后进行对数扩张,将对数域的谐波谱幅度变换为实际的谱幅度值。对U/V信息进行解码,依次确定各个谐波频带属于清音或浊音频带,然后根据语音合成介绍的方法分别对清音分量和浊音分量进行合成,最后得到总的合成语音。

4 800/920/1200bps 改进型MBE语音编码器的硬件实现

前面几节内容从实际语音处理的角度介绍了MBE模型的语音分析及合成,并针对800/920/1200bps编码速率详细具体地描述了各参数的量化编码及解码过程。在此基础上,本文设计了800/920/1200bps 3种速率的MBE语音编码方案,用Borlandc C++ 语言在PC机上对该方案进行了模拟,

并在TI公司的DSP芯片上实现了该方案。

图4示出了800/920/1200bps改进型MBE语音编码器的硬件结构框图。该语音编码器用一片数字信号处理器TMS320C31完成语音的分析和合成处理,一片程序存储器EPROM(128K×8)和4片高速RAM(32K×8)作为程序及数据的存储、交换,一片TLC32044完成A/D及D/A的转换,自己设计的专用集成电路(ASIC)完成编码数据的发送与接收等功能。语音采样率为8kHz,窗函数长度为221样点,分析帧长度为25ms(200样点),每秒40帧,800bps,920bps和1200bps采用了相同的分析合成子程序,仅在各参数编码比特分配上有所不同,各参数编码比特的具体分配详见第3节。

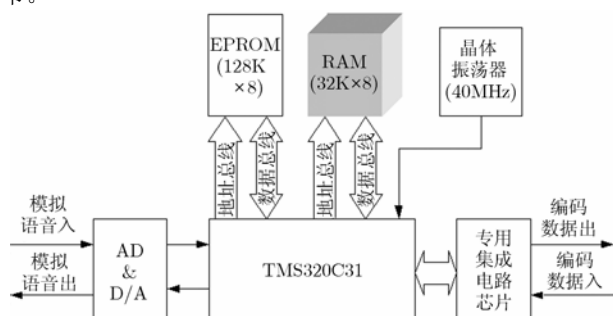


图4 800/920/1200bps改进型MBE语音编码器硬件框图

5 结束语

从C++语言的模拟及硬件系统的实时仿真统计得到,本文所设计的800~1200bps改进型MBE语音编码算法方案的复杂度为7.63MIPS,仅为目前混合激励线性预测(MELP)编码算法^[6]复杂度(20MIPS)的1/3。其中编码部分占

4.17MIPS,解码部分占3.46MIPS。整个软件占用EPROM空间约14K×32字。非正式的语音试听实验结果表明,3种编码速率合成音质的可懂度较高,但自然度还不能令人满意。

参考文献

- [1] Griffin D W. Multi-band Excitation Vocoder[D]. E E C S. Department, M I T, 1987.
- [2] 王都生, 樊昌信. 1200/2400bps改进型多带激励声码器的实时实现[J]. 电子学报, 1999, 27(1): 125-127.
Wang Du-sheng, Fan Chang-xin. A real-time implementation of the 1200/2400bps modified multi-band excitation vocoder. *Acta Electronica Sinica*, 1999, 27(1): 125-127.
- [3] 王都生, 樊昌信. 4.8kbps多带激励语音编码器的模拟实现[J]. 西安电子科技大学学报(自然科学报), 1995, 22(3): 227-234.
- [4] 王都生, 铁满霞, 樊昌信. 一种实用的双向跟踪基音周期平滑算法[J]. 电子学报, 1999, 27(10): 108-110.
Wang Du-sheng, Tie Man-xia, and Fan Chang-xin. A practical look-back and look-ahead pitch tracking smoothing algorithm. *Acta Electronica Sinica*, 1999, 27(10): 108-110.
- [5] 王都生. 多带混合激励低速率语音编码的研究[D]. 西安: 西安电子科技大学, 1998.
- [6] Alan McCree and Kwan Truong, *et al.*. A 2.4kbit/s MELP coder candidate for the new U. S. federal standard[C]. Proc. ICASSP会议, 1996: 200-203.

王都生: 男, 1962年生, 博士, 教授, 主要研究方向为多媒体信号处理、语音编码及计算机图形学。