一种采用定点 DSP 实现的 1.8kbps MBE-LPC 声码器*

李永明　陈弘毅　朱益厅

(清华大学微电子学研究所, 北京　100084)

摘要: 提出了一种结合 MBE(多带激励) 模型和 LPC(线性预测编码) 模型的 1.8kbps 声码器. 在这种声码器中, 采用 LPC 特征参数来代表语音帧的频谱, 利用 LPC 残差进行基音提取和多带清浊音判决, 采用 MBE 模型合成语音, 并在高频浊音带的语音合成中混以清音. 在定点 Motorola DSP56002 EVM 上, 可以在 1.8kbps 的码率下对语音进行实时的编解码处理, 具有存储量和计算量较小的特点. 其合成语音质量超过了 LPC-10e.

关键词: MBE-LPC; 语音分析; 语音合成; DSP 实现

EEACC: B6120B; C5260S; B6130; B6210L; C5260; C5135

中图分类号: TN912.32　　**文献标识码**: A　　**文章编号**: 0253-4177(2000)08-0803-07

1.8kbps MBE-LPC Vocoder Implemented by Fixed-Point DSP*

LI Yong-ming, CHEN Hong-yi and ZHU Yi-ting

(*Institute of Microelectronics, Tsinghua University, Beijing* 100084, *China*)

Received 16 June 1999, revised manuscript received 11 October 1999

Abstract: A 1.8kbps MBE-LPC vocoder is presented, which combines MBE model and LPC model. LPC characteristic parameters are used to represent the spectrum of speech frame; Pith-abstraction and multi-band U/V decision are carried out by analysing the LPC residue. With MBE model, the synthesised speech is obtained. And unvoice is mixed in the speech synthesis of high frequency voice band. At the bit rate of

* 国家自然科学基金资助项目(69636030) [Project Supported by National Natural Science Foundation of China Under Grant No. 69636030].

LI Yong-ming(李永明)　male, born in 1945, Associate Professor, working with design and teach on speech signal processing and analog LSI.

CHEN Hong-yi(陈弘毅)　male, born in 1942, professor, vice director of Institute of Microelectronics, Tsinghua University, senior member of China Electronics Engineers, working with VLSI design, digital signal processing and hardware architecture of algorithms.

ZHU Yi-ting(朱益厅)　received the Master degree from Institute of Microelectronics, Tsinghua University, in 1999, working with digital signal processing and VLSI. From 1996 to 1999 he was involved in speech signal processing and working on low bit-rate vocoder.

1999-06-16 收到, 1999-10-11 定稿

1.8kbps it is implemented by the fixed-point motorola's DSP560002 EVM board real-timely. The vocoder is of smaller memory storage and simplify computation, with quality better than LPC-10e.

Key words: Multi-Band Excitation-Linear-Prediction Coding; speech analysis; speech synthesis; DSP implementation

EEACC: B6120B; C5260S; B6130; B6210L; C5260; C5135

Article ID: 0253-4177(2000)08-0803-07

1 引言

应用参数编码技术,构造和改进 2.4kbps 及以下码率的语音编码算法是目前低码率声码器的研究重点.由于多带激励(MBE)模型[1,2]对频谱的描述更加接近实际语音,因此自从它被提出以来,得到了人们的广泛关注.以这种模型为基础的一种 4.15kbps 声码器被定为 Inmarsat-M 的标准[3],但是这种 MBE 声码器关于谱幅度系数的编码是利用 DCT 变换并需要较多的码元,不利于低码率传输.而线性预测编码模型[4]的 LPC 特征参数可以描述语音帧的频谱包络,将 LPC 特征参数转化为等价的线谱对频率(LSF)[5]后可以用较少的码元量化传输.近年来结合 MBE 和 LPC 的 MBE-LPC 算法[6]在 2.4kbps 左右的传输码率下得到了较高质量的合成语音.

这里,从进一步降低码率和减少存储量、计算量的角度,提出一种 1.8kbps MBE-LPC 语音处理算法.对短时语音帧作 LPC 分析可以计算出该帧语音的能量和谱包络,对 LPC 残差作自相关处理则可以实现基音提取,帧间的动态基音平滑可以明显改善基音提取的准确性;在基音及其谐波频率处对 LPC 谱作抽样,就可以方便地求得峰值拟合较好的谱幅度系数,这种谱幅度系数与通过 MBE 标准算法得到的谱幅度系数基本一致;进一步对 LPC 残差作傅里叶分析,根据残差的频域波形则可以很简洁地判定每个带是清音带还是浊音带.在语音合成端,浊音由规则的余弦波激励在时域合成,而清音则由白噪声谱经带通滤波在频域合成,最后将两者结果相加得到合成语音.声码器的量化传输参数为:LSF 系数、能量系数、基音和清浊音判决.

2 MBE-LPC 语音处理算法

2.1 MBE-LPC 语音分析

语音分析的框图如图 1 所示.输入语音 $S(n)$ 先通过高通滤波,滤除 60Hz 以下低频成份并加强语音的高频部分,以便更好地进行 LPC 分析.对于每一帧加窗的语音信号,利用 Durbin 算法进行 LPC 分析,得到语音帧的能量 G 和 LPC 参数,这种参数表示了语音帧的频谱包络,它可以转化为等价的 LSF 系数予以编码传输.同时,也得到了 LPC 残差信号.

基音提取算法分两步进行,第一步,对低通残差信号作归一化的自相关,自相关值的局部最大处被选为基音候选值.第二步,采用基音平滑算法从候选值中提取出合适的基音.实际的语音信号,基音周期的变化比较缓慢,帧间的相关性较大,表现为一定的连续性.从前一帧基音和紧接着的连续二帧基音候选值中,可以选择不同的路径来作为连续三帧的基音,如果为路径选择合适的权重,则可以将基音平滑转化为无环图的最短路径问题,用动态平滑算法来解决.由于基音平滑要求帧间基音连续且基音处自相关值要尽可能大,因此平滑的权重应该包括两部分,即自相关值权重和基音权重.自相关值权重反映了自相关函数值对基音提取的影响;自相关函数值越大,自相关值权重就越小,接近准确基音周期的可能性就越大.基音权重体现了基音平滑的要求,基音权重越小,相邻两帧的基音变化就越小,基音变化越接近自然.这样,可以将最短路径的当前帧节点视为基音平滑结果,作为最后的基音值.

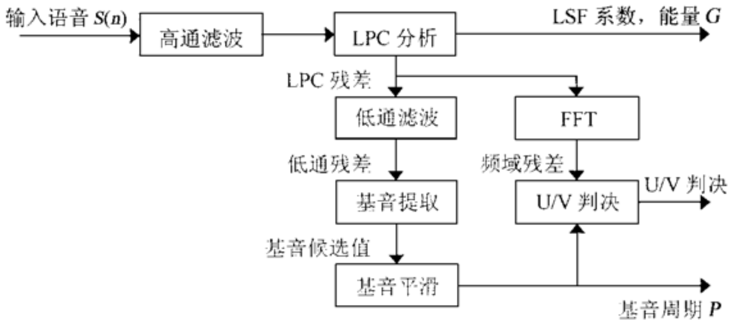

图 1　语音分析框图

FIG. 1　Frame of Speech Analysing

根据清浊音的物理含义，若某一子带是浊音带，其残差频谱必定有周期性的峰起，峰与峰的距离就是基音频率；若是清音带，其残差频谱则表现为噪声谱，在这种谱中很难体现基音频率，而清浊音之间的过渡带则两种性质兼有，但都不明显. 这里的 U/V 判决方法根据 LPC 残差信号在频域的子带波形，对每一个子带计算其自相关函数，进而分析其周期性和非周期性，若周期性占优势就被判为浊音，非周期性占优势则被判为清音，这种判决原则与全带语音 U/V 判决的原理一致. 每个子带的谐波数 l 的选择也相当重要，首先自相关计算本身要求 $l \geq 2$，其次，l 越大，U/V 判决的编码码元数越少，有利于低码率传输，但是，一般还要求 $l \leq 5$，因为 U/V 判决是对带内各谐波的平均判决，所以每带的谐波数越多，U/V 判决的可信度越低，当 $l \geq 8$ 后，其合成语音的音质已经比较接近标准 LPC 声码器.

2.2　MBE-LPC 语音合成

语音合成需要的参数包括：LSF 谱包络信息、能量系数、基音频率和各个带的 U/V 判决. 为了保证基音频率等参数的连续，MBE-LPC 语音合成对浊音带采用时域合成法，因为频域中比较容易实现带通滤波，清音带则采用频域合成法；语音合成的框图如图 2 所示.

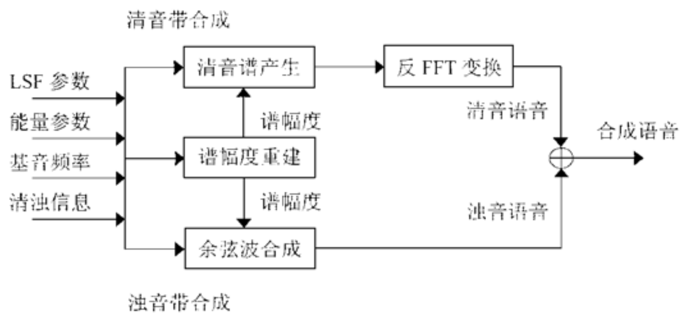

图 2　语音合成框图

FIG. 2　Frame of Speech Synthesis

重建谱幅度系数可以从重建的 LSF 系数、能量系数、基音频率经计算得到. 由于 LSF 系数通过全极点模型表示了一帧语音的频谱包络，能量系数则表示了该帧语音的能量，即频谱的总体绝对大小，因此，根据基音频率在每一个谐波处对频谱进行抽样就可以得到重建的谱幅度系数.

这里的 LSF 系数是对全带频谱的整体拟合，不是对谐波频率处谱幅度的离散点拟合，因此，这种方法得到的谱幅度系数与 4.15kbps MBE 标准算法得到的谱幅度系数相比，其精确度有所下降，但实验表明，这种下降对合成语音的质量影响并不大，而且由于这种方法在语音分析端的计算量和存储量都相对较小，因此，在 MBE-LPC 语音算法中，采用了这种折衷方法.

MBE-LPC 在频域对清音带进行语音合成，采用白噪声谱结合谱幅度系数来产生激励谱. 首先产生一个均值为零的白噪声序列，对这种噪声序列加窗进行傅里叶变换，得到白噪声谱，清音激励谱由白噪声谱和谱幅度系数共同确定. 然后，利用反向傅里叶变换求出清音时域信号.

若高频端是浊音,由于在时域合成浊音时全部采用余弦波激励,会造成合成语音的合成感加重;因此,高频端浊音带混入适当清音,提高合成语音的自然度.

MBE-LPC 在时域对浊音带进行语音合成,可以尽量保持合成语音的帧间连续性.它根据每个谐波的谱幅度系数,利用余弦波直接产生语音波形,最后将所有谐波的语音波形相加得到浊音合成信号.在分别求出清音信号和浊音信号后,在每一个样点位置上把它们相加,产生最后的合成语音.

2.3 参数编码

MBE-LPC 语音分析得到的参数为:LPC 特征参数、能量系数、基音周期和 U/V 判决系数,这些参数的总体量化编码方案为:LPC 特征参数转化为 LSF 系数后采用分裂矢量量化进行编码,能量系数在压缩动态范围后采用均匀标量量化的方法进行编码,基音周期则采用非均匀标量量化进行编码,U/V 判决系数的结果是若干位的二进制码,可以直接传输.

语音抽样率为 8kHz,即 25ms 为一帧, 1.8kbps 的数据传输率,每帧编码的比特数为 45 比特,比特数分配如表 1.

表 1 比特数分配

Table 1 Bit Rate Distribution

参数名称	LSF 系数	能量系数	基音周期	U/V 判决	总计
比特数	25	5	6	9	45

2.3.1 基音周期的量化

基音提取得到的结果为[18, 124]范围内的整数样点,分辨率为 1 个样点,语音分析中谱幅度系数和 U/V 判决对基音分辨率的要求并不是很高,从该语音合成器的角度来看,细微的基音频率的移动对听觉的影响可以忽略.

基音出现于低端的概率远大于出现于高端的概率,同时,对听觉直接产生影响的是基音频率的误差大小而不是基音样点,如在低端附近有 1 个样点的误差反映到频率上是 20Hz 左右,而在高端附近有 1 个样点的误差则只是 0.52Hz,其听觉效应远小于低端,对高端的量化可以相对比较粗糙,因此采用 6 比特非均匀标量量化的方法对基音周期进行量化,其量化表如表 2.

表 2 基音量化表

Table 2 Quantization of Base

索引号/基音值	00/18	01/19	02/20	03/21	04/22	05/23	06/24	07/25
…	…	…	…	…	…	…	…	…
索引号/基音值	48/74	49/76	50/79	51/82	52/85	53/88	54/91	55/94
索引号/基音值	56/97	57/100	58/104	59/108	60/112	61/116	62/120	63/124

2.3.2 U/V 判决的量化

每带包括若干个谐波数,用 1 比特来表示它的 U/V 情况,若浊音带为 1,清音带则为 0. 一般情形,设某帧语音有 M 个谐波,且每 N_1 个谐波组成一个带,最多可有 N_2 个带,则实际带的数目 K 可以由下列公式得到:

$$K = \begin{cases} \left\lfloor \dfrac{M + N_1 - 1}{N_1} \right\rfloor, & M < N_1N_2 \\ N_2, & M \geqslant N_1N_2 \end{cases} \tag{1}$$

每帧语音的 U/V 信息需要使用 K 比特传输.

这里,为有利于简化算法,N_1 取 4,N_2 取 9,K 取为固定值 9,用 9 比特传输 U/V 判决信息.

2.3.3　能量系数的量化

在 MBE-LPC 算法中,能量系数的动态范围是比较大的. 加哈明窗且采用 Durbin 算法作 LPC 分析为前提条件的情况下,其动态范围可以达到(0, 5000).

人的听觉对语音能量的对数域敏感,从电路集成和 DSP 实现考虑,对数运算比较复杂,因此,采用下述量化方案:先对能量系数开平方,然后将结果线性映射到范围[0, 31]中,取这个范围内与映射结果最近的整数作为量化值,量化比特数为 5 比特.

2.3.4　谱幅度系数的量化

短时语音帧 LPC 特征参数$\{a_k\}$,其动态范围很大,不适合量化和传输,与其等价的部分相关系数$\{k_i\}$也没有很好地解决这个问题. 目前,语音编码算法一般都采用 LSF 系数来进行量化和传输.

实验表明,如果要求平均频谱失真低于 1dB,标量量化至少需要 32 至 40 比特用于每帧的 LSF 特征矢量编码,显然,当编码速率低于 2.4kbps 时,这种方法已经不再适合.

在相同比特数的情况下,矢量量化的频谱失真比标量量化小的多. 但要分析,例如,每帧 20 比特,其相应的码本就包含 1048576 个不同的矢量. 这对声码器的设计和硬件实现会造成很大的困难.

在 MBE-LPC 算法中,采用分裂矢量量化的方法来缩小码本,从而减少存储空间和码本搜索的计算量.

2.3.5　LSF 分裂矢量量化

LSF 特征矢量$\{\omega_i\}$的各分量虽然属于一个整体,但是它们单个分量的误差对谱幅度的影响只局限于各自的频率区域内,因此可以把一个维数比较高的特征矢量分成若干个维数较低的矢量[7,8],用几个很小的矢量码本代替单个的大码本. 降低存储量和搜索计算量. MBE-LPC 算法中采用 25 比特来量化特征矢量$\{\omega_i\}$,其分裂矢量量化过程如图 3.

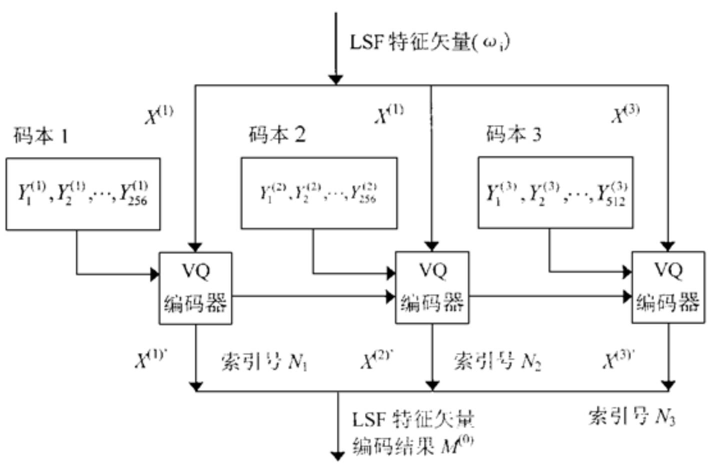

图 3　分裂矢量量化

FIG.3　Quantization of Spliting Vector

经语音分析得到的 LSF 特征矢量$\{\omega_i\}$首先被分裂成三个矢量 $X^{(1)}$、$X^{(2)}$ 和 $X^{(3)}$:

$$\begin{cases} X^{(1)} = (\omega_1, \omega_2, \omega_3) \\ X^{(2)} = (\omega_4, \omega_5, \omega_6) \\ X^{(3)} = (\omega_7, \omega_8, \omega_9, \omega_{10}) \end{cases} \tag{2}$$

对于人的听觉来说,高端频率的误差对合成语音的影响小于低端频率的影响,因此,应该分配相对较多的比特数给矢量 $X^{(1)}$ 和 $X^{(2)}$,分配相对较少的比特数给矢量 $X^{(3)}$,$X^{(1)}$、$X^{(2)}$ 和 $X^{(3)}$ 分别采用 8 比特、8 比

特和 9 比特来编码.

矢量码本$\{Y_i^{(1)}\}$、$\{Y_i^{(2)}\}$和$\{Y_i^{(3)}\}$通过训练得到, 训练中采用欧氏距离来衡量畸变. 三个码本$\{Y_i^{(1)}\}$、$\{Y_i^{(2)}\}$和$\{Y_i^{(3)}\}$的容量应分别是 256 个矢量、256 个矢量和 512 个矢量.

首先, 对 $X^{(1)}$ 进行量化, 接着对 $X^{(2)}$ 和 $X^{(3)}$ 进行量化, 保证最后总的量化特征矢量(ω)满足各分量按顺序递增的 LSF 性质, 得到量化矢量 $X^{(i)'}$和它的索引号 N_i. 权矩阵 $W^{(3)}$ 与 $W^{(1)}$、$W^{(2)}$ 有所区别, 因为人的听觉对高频误差的不敏感性, 减轻 ω_9、ω_{10}的权重, 相对提高 ω_7、ω_8 对应的权重, 对改善合成语音的质量有利.

根据索引号 N_1、N_2 和 N_3, 得到最后的编码结果 $M^{(0)}$, 它是一个 25 位的二进制数, 由 N_1、N_2 和 N_3 均转化为二进制数后连接得到. 由于只使用三个小容量的码本, 而且搜索算法也比较简洁, 只需较少的存储空间和较小的计算量, 有效地降低了 DSP 和集成电路实现的难度.

3 系统结构

采用定点 DSP[9—11]作为核心构件实现系统, 如图 4 所示. 语音信号的处理过程如下: 先在微机上完成处理算法的汇编语言程序或 C 语言程序, 编译生成可执行文件, 通过 OnCE 接口下载到程序 SRAM 中, 或者将可执行文件固化到 32K×24 的 EEPROM. 运行过程中从 EEPROM 取得指令, 下载后就可以运行处理程序. 模拟语音信号通过 A/D 转换成 16 位 PCM 数字信号给 DSP56002, DSP56002 可对输入语音进行语音分析、量化编码及解码、语音合成, 语音合成的结果送回, 经 D/A 转换成模拟信号输出. 模拟信号的接收和发送采用中断的方式, 以保证语音处理程序的正确运行. 在语音处理过程中, 微机可以通过 OnCE 接口经 RS-232 协议对 DSP56002 的程序运行进行监控, 量化编码的结果, 也可以通过 OnCE 接口经 RS-232 协议回传到微机, 实现语音处理的中间结果在微机上的存储.

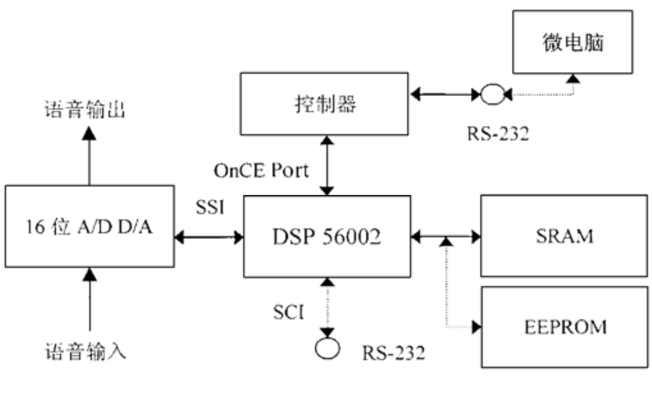

图 4 Motorola DSP56002 EVM 结构图

FIG. 4 Structure Diagram of Motorola DSP56002 EVM

4 实验结果

因为采用存储量和计算量都比较小的算法方案, 所以, 只使用了 DSP56002 EVM 中较少的一部分资源. 表 3 列出了 MBE-LPC 算法的存储器需求, 表 4 列出了该算法的计算量.

表 3 MBE-LPC 算法的存储器分配

Table 3 Memory Arrangement of MBE-LPC Algorithm

程序存储区(容量 4K)	数据存储区(容量 8K)	
程序	变量	码本
2.207K	2.968K	2.589K

表 4 MBE-LPC 算法的计算量

Table 4 Calculating Rates of MBE-LPC Algorithm

语音分析和编码	解码和语音合成	总计	利用率
6.40MIPS	9.36MIPS	15.76MIPS	78.8%

从表 3 和表 4 可以看出, 在这种工作模式的条件下, DSP56002 EVM 还剩余 1.793K 的程序存储空间和 2.443K 的数据存储空间以及 21.2% 的运算能力. 一方面, 少的资源需求对算法的 VLSI 实现相当有

利; 另一方面, 可以有更多的资源用于算法的进一步改进, 提高合成语音的质量.

5　结论

提出的 MBE-LPC 算法改善了 LPC 方法语音分析的精度, 兼用了 MBE 方法合成语音的优点. 声码器的计算量和存储量小, 语音质量较 LPC-10e 有一定改善.

参　考　文　献

[1] D. W. Griffin, J. S. Lim, IEEE Trans. on Acoustics, Speech and Signal Processing, 1988, **36**(8) : 1223—1235.

[2] M. S. Brandstein, P. A. Monta, J. C. Hardwick and J. S. Lim, "A Real-Time Implementation of the Improved MBE Speech Coder", IEEE ICASSP90 New York, NY, USA, 1990, 5—8.

[3] Digital Voice System, "Inmarsat-M Voice Coding System Description", Draft Version 1. 2, Inmarsat-M Specification, Inmarsat, London, 1990.

[4] T. E. Tremain, "The Government Standard Linear Predictive Coding Algorithm: LPC-10", Speech Technology, 1982, **1**(2) : 40—49.

[5] F. K. Soong and B. H. Juang, "Line Spectrum Pair and Speech Data Compression", IEEE ICASSP84 New York NY, USA, 1984, 1. 10. 1—1. 10. 4.

[6] Yang Huimin, Chen Hongyi and Sun Yihe, "A 2. 4kbps MBE-LPC Speech Codec Algorithm Suitable for VLSI Implementation", ASICON Shanghai 1996, 143—146.

[7] K. K. Paliwal and B. S. Atal, IEEE Trans. on Speech and Audio Processing, 1993, **1**(1) : 3—14.

[8] F. K. Soong and B. H. Juang, IEEE Trans. on Speech and Audio Processing, 1993, **1**(1) : 15—24.

[9] Motorola Corporation, User's Manual of DSP56002, 1996.

[10] Motorola Corporation, Instruction Set Details of DSP56002, 1996.

[11] Motorola Corporation, User's Manual of DSP56K C Compiler, 1996.