

基于 DSP 的甚低速率语音编码算法及其实现

赵继勇¹, 曹 芳², 梁妙元³, 刘亚峰⁴

(1. 解放军理工大学通信工程学院, 南京 210007; 2. 南京理工大学紫金学院, 南京 210046;

3. 解放军理工大学气象学院, 南京 211101; 4. 解放军 69026 部队, 乌鲁木齐 830002)

摘 要: 在混合激励线性预测(MELP)算法的基础上, 设计一种 1 120 b/s MELP 甚低速率语音编码算法。该算法通过增加帧长、动态比特分配和多帧联合矢量量化、及参数内插等方法降低语音的编码速率, 并已在 TMS320VC5416DSP 芯片上实时实现。采用美国 GL 公司的语音质量评估系统 VQT, 对编解码的实时语音质量进行评估, 语音质量感知评价高于 3 分。实验结果表明, 该算法能够满足实际通信要求。
关键词: 甚低速率语音编码; 动态比特分配; 多帧联合矢量量化; 参数内插

Very Low Bit Rate Speech Coding Algorithm and Its Implementation Based on DSP

ZHAO Ji-yong¹, CAO Fang², LIANG Miao-yuan³, LIU Ya-feng⁴

(1. Institute of Communication Engineering, PLA University of Science and Technology, Nanjing 210007, China;

2. Zijin College, Nanjing University of Science and Technology, Nanjing 210046, China;

3. Institute of Meteorology, PLA University of Science and Technology, Nanjing 211101, China;

4. The 69026 Force of Chinese PLA, Urumqi 830002, China)

【Abstract】 Based on the MELP algorithm, an improved 1 120 b/s MELP very coding algorithm is proposed and implemented. Increase of framesize, dynamic bit allocation, multi-frame joint vector quantization and parameter interpolation are used to reduce the bit rate. The speech coding algorithm is implemented on DSP chip TMS320VC5416 in real time. Tested by GL Voice Quality Tester(VQT), the Perceptual Evaluation of Speech Quality(PESQ) scores of the speech coder is more than 3 points. Experimental result shows that this speech coding algorithm can meet actual communication requirements.

【Key words】 very low bit rate speech coding; dynamic bit allocation; multi-frame joint vector quantization; parameter interpolation

DOI: 10.3969/j.issn.1000-3428.2011.21.089

1 概述

低速语音编码一直是语音研究领域中的一个热点课题, 主要应用于军事保密通信、卫星通信和数字语音存储系统。各种 4 Kb/s 以下的语音编码算法主要归结为混合激励线性预测(Mixed Excitation Linear Prediction, MELP)、多带激励(Multi Band Excitation, MBE)、波形内插(Waveform Interpolation, WI)和正弦变换编码(Sinusoidal Transform Coding, STC)4 种模型。与线性预测模型(Linear Prediction, LPC)相比, 这 4 种模型均充分考虑了语音信号的本质特性, 对模型中的激励源进行了更精细的划分和处理。为了在更低速率上得到高质量的合成语音, 必须在降低码率的同时尽可能保留对激励源参数的编码。为此, 本文提出一种甚低速率语音编码算法, 来实现语音编码系统。

2 甚低速率语音编码算法

2.1 编码基本原理

本文设计并实现一种 1 120 b/s 甚低速语音编码算法, 其基本编码原理为: 原始语音信号经过预处理后, 对每个子帧分别提取线谱对频率(Line Spectrum Frequency, LSF)、基音周期、清浊音模式、能量、非周期标志及傅里叶幅度等参数。

2.2 编码器比特分配

本文算法由 4 个连续语音子帧组成一个超级帧, 子帧帧长 25 ms, 总帧长 100 ms, 编码比特为 112 bit, 总计为 112 bit/100 ms=1 120 b/s。本文算法选择了对合成语音质量影响较大

的线谱对频率、傅里叶幅度、基音周期和清浊音模式、能量以及非周期标志等参数进行量化。具体参数的量化比特分配见表 1。

表 1 1 120 b/s 量化参数比特分配

参数	量化比特/bit
线谱对频率	50
傅里叶幅度	32
基音周期, 清浊音模式	13
能量	12
非周期标志	4
同步	1

2.3 参数量化

2.3.1 线谱对频率矢量量化

本文考虑到线谱对频率的帧间相关性和误差的相对独立性, 同时为了减少编码比特数, 因此采用了帧间内插的方案: 对 2、4 子帧的 LSF 参数采用多级矢量量化, 而 1、3 子帧的 LSF 参数由内插方法得到。2、4 子帧进行四级矢量量化, 量化采用 4 级码本, 各级量化比特数分别为 7 bit、6 bit、6 bit、6 bit, 总共 25 bit。在多级联合优化码本设计^[1]过程中, 考虑各级码本的相互影响, 对各级码本进行了迭代更新, 从而达到多级码本的联合优化。

作者简介: 赵继勇(1978—), 男, 讲师, 主研方向: 中低速语音编码; 曹 芳, 助教; 梁妙元, 讲师; 刘亚峰, 助理工程师

收稿日期: 2011-04-15 **E-mail:** zhaojiyong_ntg@163.com

2.3.2 基音周期和清浊音模式矢量量化

本文采用动态比特分配方案,对浊音子帧多的超级帧,分配较多的比特用于基音周期编码;对浊音子帧少的超级帧,分配较少的比特用于基音周期编码。

本文算法按照 4 个子帧共 16 种清浊音模式,分为 5 类情况对基音周期和清浊音模式进行动态比特分配以及联合矢量量化。编码比特共 13 位,包括模式位和基音位:当子帧全是清音,分配全 0 码字;当仅有 1 个浊音子帧时,基音周期采用 7 bit 标量量化;当有 2、3、4 个浊音子帧时,分别采用 8 bit、10 bit、10 bit 的加权矢量量化算法对基音周期参数进行编码。具体基音周期和清浊音模式编码比特分配方案详见表 2。

表 2 基音周期和清浊音模式编码比特分配

子帧清浊音模式	模式位+基音位	13 比特编码
0000	—	0000000000000
0001	00+7 位	000000,xxxxxxx
0010	01+7 位	000001,xxxxxxx
0100	10+7 位	000010,xxxxxxx
1000	11+7 位	000011,xxxxxxx
0011	10+8 位	00010,xxxxxxxx
0101	11+8 位	00011,xxxxxxxx
0110	100+8 位	00100,xxxxxxxx
1001	101+8 位	00101,xxxxxxxx
1010	110+8 位	00110,xxxxxxxx
1100	111+8 位	00111,xxxxxxxx
0111	10+10 位	010,xxxxxxxxxxx
1011	11+10 位	011,xxxxxxxxxxx
1101	100+10 位	100,xxxxxxxxxxx
1110	101+10 位	101,xxxxxxxxxxx
1111	111+10 位	111,xxxxxxxxxxx

为避免基音量化不准确带来的音调变化,对多帧联合编码的基音周期采用加权矢量量化,量化目标函数采用加权欧氏距离。基音周期量化采用单级码本,利用模拟退火算法^[2]设计码本,量化过程中采用全搜索对码字进行搜索。

2.3.3 能量矢量量化

每个子帧提取一个能量参数,组合成 4 维矢量,在对数域进行量化,并将对数能量取值范围限制在[10,77] dB 内,采用 12 bit 的增益形状矢量量化。在码本设计中使用模拟退火算法;在码本搜索过程中,采用欧氏距离作为目标函数,使用全搜索选择目标函数最小的码本序号进行传输。

2.3.4 傅里叶幅度

在每个子帧中,先对残差信号进行 FFT 得到残差谱,然后残差谱经过峰值提取得到傅里叶谐波幅度,最后对最低 10 阶傅里叶谐波幅度采用 8 位矢量量化,码本查找采用固定加权后的欧氏距离,使低频段比高频段有更大的影响。

3 基低速率语音解码算法

3.1 基本结构解码

解码算法先进行参数译码,再分别合成每个子帧的语音信号。与传统线性预测合成方法不同的是它充分反映了语音信号的本质特征,极大地提高了合成语音的质量。其他后续处理,如混合激励产生、自适应频谱增强、线性预测合成、增益校正、脉冲离散滤波等,与 MELP 算法一致。

3.2 参数解码

线谱对频率解码时先分别对 2、4 子帧的线谱对频率进行解码,再通过内插公式得到 1、3 子帧的线谱对频率。

基音周期解码时先判断清浊音模式,再根据清浊音模式和码本序号对基音周期进行解码。

上述参数保证了合成语音的可懂度,为了提高合成语音的自然度,还需对傅里叶幅度、能量参数以及非周期标志进行恢复,在降低语音编码速率的同时保证了合成语音的质量。

4 系统的实时实现

4.1 芯片选择

本语音编码系统的运算量经过估计小于 100 MIPS,存储空间小于 128k 字,因此采用通用 DSP TMS320VC5416^[3]实现。TMS320VC5416 具有低功耗、运算速度快、存储空间大、外围接口功能全和成本低等优点,可以满足语音编码系统单片实现的需要。

4.2 硬件设计

本系统以 DSP 为主处理和单元,用于完成语音编码系统关键算法和控制功能。模拟接口部分完成语音 A/D、D/A 等功能^[4]。AM29F800 作为存储程序和数据的 Flash,容量为 8 MB。并且,各个接口间的时序可由可编程逻辑器件完成。系统硬件框图如图 1 所示。

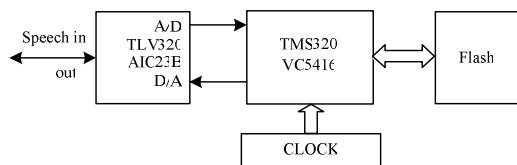


图 1 1120 b/s 语音编码系统硬件框图

本文语音编码系统硬件设计简洁、灵巧,能够充分利用 TMS320VC5416 的硬件资源,在功耗、体积和成本方面具有一定优势。

4.2.1 模拟接口部分

模拟接口部分采用 TI 公司 TLV320AIC23B^[5]完成所需功能:输入放大,抗混叠滤波,A/D,D/A,低通平滑滤波,输出功率推动等。模拟接口支持线路/麦克风输入和线路/耳机输出。TLV320AIC23B 的 PCM 编解码器数字接口为同步串行方式,可直接与 TMS320VC5416 的多通道缓冲串口(McBSP)连接。帧速率为 8 kHz,每次采样收发 16 bit 线性数据。

(1)McBSP0 与 AIC23B 的数据口连接,实现 A/D 数据输入和 D/A 数据输出。TMS320VC5416 将 McBSP0 配置成从模式,与 TLV320AIC23B 的数据口连接;AIC23B 应设置为 DSP 模式与 McBSP0 进行通信。其硬件连接如图 2 所示。

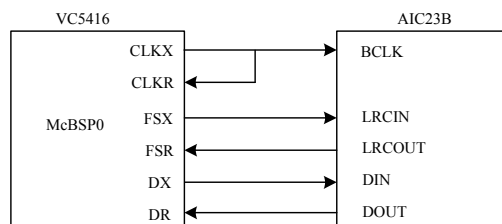


图 2 AIC23B 与 McBSP0 连接框图

(2)McBSP2 与 AIC23B 的控制口连接,对 AIC23B 的寄存器进行设置,其通信协议遵循 SPI 标准,将 McBSP2 设置为主模式。其硬件连接如图 3 所示。

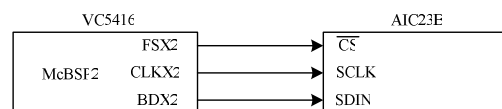


图 3 AIC23B 与 McBSP2 连接框图

4.2.2 存储部分

程序、码本和初始化数据均存储于 Flash 中,本文 Flash 和 DSP 通过并口连接,采用并行加载方式。在加电时,DSP 的自引导程序将代码引导入 DSP 的 RAM 中运行。

4.3 软件系统设计

整个系统软件包括主模块与语音输入/输出模块两部分。

主模块完成系统初始化及语音编解码等功能。语音输入/输出模块完成输入语音采集及合成语音的输出等功能。

4.3.1 软件各模块工作时序

语音 I/O 模块每 125 μ s 运行一次(8 kHz 采样), 从模拟接口读入一个语音样值并向模拟接口输出一个语音样值, 每帧(100 ms)共收发 800 个样点。其余时间均为主模块运行时间。

4.3.2 主模块

开机加电或系统复位后, TMS320VC5416 跳转到系统复位中断服务程序运行自引导程序, 将存于 Flash 中的程序和数据引导至 DSP 内部 RAM, 从程序入口地址开始执行。主模块通过对寄存器设置, 完成 TMS320VC5416 DSP 系统和外围电路初始化; 打开串口中断和其他外部中断源; 根据控制接口的设置, 设置程序运行标志, 控制程序运行分支, 接下去是分帧进行完全相同的处理。图 4 是主模块的程序流程。

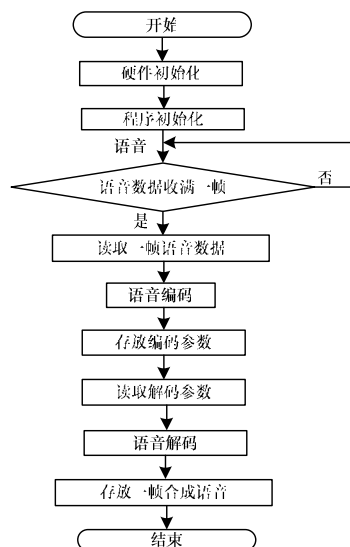


图4 主模块程序流程

4.3.3 语音输入/输出模块

语音输入输出模块由串口中断来触发, 收数据和发数据共用一个串口中断服务程序。在中断服务程序, 从串口接收寄存器读入一个样值到接收缓冲区, 并从发送缓冲区送出一个样值到串口发送寄存器。在接收缓冲收满一帧后, 将接收数据转移交给主模块。在发送数据发送一帧后, 从主模块取一帧新数据用于发送。

5 实验结果

采用 GL 公司的语音质量评估系统 VQT 对编解码的实时语音质量进行评估。该软件可采用 3 种广泛使用的语音质量评估算法 PESQ(Perceptual Evaluation of Speech Quality)、

PAMS(Perceptual Analysis/Masurement System) 和 PSQM(Perceptual Speech Quality Measurement), 对如 VoIP、PSTN 以及移动电话等的通话质量进行评估。本文在对编解码实时语音进行质量评估时, 采用 PESQ 算法, 以 6 段 8 kHz 采样语音作为输入, 其中男声和女声各一半。具体评估结果如图 5 所示。可见, 在 1 120 b/s 速率下, 6 段语音 PESQ 平均得分为 3.01, 其中, 3 段合成语音质量一般, 另 3 段合成语音质量良好。非正式主观试听表明, 本文实现的 1 120 b/s 甚低速率语音编码算法的合成语音具有较好的可懂度、清晰度。

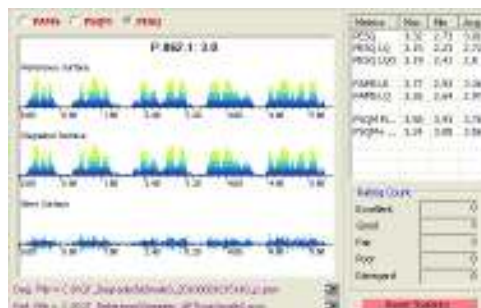


图5 VQT 语音质量评估结果

6 结束语

本文设计并实现一种甚低速率的语音编码算法。在软、硬件设计过程中, 根据算法复杂度等选择 DSP 芯片, 并针对芯片结构完成了算法优化和实现, 围绕硬件接口和功能要求进行外围电路设计。最终通过软、硬件集成和调试, 实时实现了该语音编码系统。VQT 结果表明, 该语音编码算法合成语音具备了一定的自然度, 质量达到了预期的效果, 可应用于短波无线通信中, 而且在其他低速率语音场合具有一定的应用前景。

参考文献

- [1] Chan W Y, Gupta S, Gersho A. Enhanced Multistage Vector Quantization by Joint Codebook Design[J]. IEEE Trans. on Communications, 1992, 40(11): 1693-1697.
- [2] 邹霞, 陈亮, 张雄伟. 甚低速率语音编码中的高效模拟退火算法研究[J]. 系统仿真学报, 2004, 16(10): 2181-2184.
- [3] TMS320VC5416 Fixed-point Digital Signal Processor Data Manual[Z]. (2002-10-12). <http://www.alldatasheet.com>.
- [4] 郭秋平, 项杰. 基于嵌入式系统的 CVSD 语音编解码器的实现[J]. 计算机工程, 2006, 32(16): 213-214.
- [5] Stereo Audio CODEC, 8-to 96-KHz, with Integrated Headphone Amplifier TLV320AIC23B Data Manual[Z]. (2004-12-20). <http://www.ti.com>.

编辑 索书志

(上接第 260 页)

参考文献

- [1] Ogras U Y, Hu Jingcao, Marculescu R. Key Research Problems in NoC Design: A Holistic Perspective[C]//Proc. of the 3rd IEEE/ACM/IFIP International Conference on Hardware Software Code-sign and System Synthesis. Jersey City, USA: IEEE Press, 2005.
- [2] 欧阳一鸣, 董少周, 梁华国. 基于 2D Mesh 的 NoC 路由算法设计与仿真[J]. 计算机工程, 2009, 35(22): 227-229.
- [3] Murali S, Micheli G D. Bandwidth-constrained Mapping of Cores onto NoC Architectures[C]//Proc. of Design Automation and Test in Europe Conference and Exhibition. Washington D. C., USA: IEEE Computer Society, 2004.
- [4] Hu Jingcao, Marculescu R. Energy and Performance Aware Mapping for Regular NoC Architectures[J]. IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems, 2005, 24(4): 551-562.
- [5] Janidarmian M, Khademzadeh A, Tavanpour M. Onyx: A New Heuristic Bandwidth-constrained Mapping of Cores onto Tile-based Network on Chip[J]. IEICE Electronics Express, 2009, 6(1): 1-7.

编辑 刘冰

