Research and Implement on MELP/CELP Hybrid Coding Method Basing on TMS320DM6446 *

MA Zhen[1], LI Jian-lei[1], CHEN Yan-ping[2]

1. Department of Computer Science, Binzhou University, Binzhou, Shandong 256600, China;
2. Shandong Mobile Binzhou branch, Binzhou, Shandong 256600, China

Abstract: This paper introduces the model of Mixed Excitation Linear Prediction (MELP), some vital technologies about MELP and its features basing on discussion of the structure of TMS320DM6446. The MELP/CELP hybrid coding method has been provided here basing on MELP, which combines the two coding method adapting to different speech frame type respectively, and uses CELP in speech frame whose periodicity is bad to make up for MELP, the structure of coder and quantification method are discussed here also. Through implement by TMS320DM6446, the speech quality is close to ADPCM with 32kbit/s coding rate.

Key words: Mixed excitation; Code excited; TMS320DM6446

EEACC: 6120B

基于 TMS320DM6446 的 MELP/CELP 混合编码方法研究与实现 *

马 震[1], 李建磊[1], 陈延萍[2]

1. 滨州学院计算机科学系 山东 滨州 256600;
2. 山东移动滨州分公司 山东 滨州 256600

摘 要: 在讨论 TMS320DM6446 结构的基础上, 讨论了混合激励线性预测 (MELP) 的算法模型, 并详细研究了 MELP 的几种关键技术. 在 MELP 基础上, 提出了 MELP/CELP 两种适用语音帧类型互补的编码方案相结合的语音编码方法, 在周期性不强的语音帧采用 CELP 来弥补 MELP 的不足, 并给出了这种编码器的结构和量化方案. 使用 TMS320DM6446 实现发现可以获得与 32kbit/s 编码速率的 ADPCM 相近的语音质量.

关键词: 混合激励; 码激励; TMS320DM6446

中图分类号: 0157.4 文献标识码: A 文章编号: 1005-9490(2007)06-2174-04

语音是人们交流信息的主要手段之一, 用数字化的方法进行语音的传送、存储、识别、合成等构成了目前数字化通信网中最重要、最基本的组成部分之一. 也正因为信息化社会的高速发展, 频率资源也就显得更加宝贵. 所以, 如何能够利用尽量少的带宽传输语音成为了近年来人们研究的重点.

TMS320DM6446 是 Ti 公司提供的一种基于达芬奇(DaVinci)技术的产品, 主要被应用在可视电话、数字机顶盒、数字监控等领域, 是这些应用中语音信号编解码的核心芯片. 本文在对 TMS320DM6446 的结构特点研究的基础上, 对 MELP/CELP 混合编码方法进行了研究, 并采用 TMS320DM6446 进行实现, 最后对合成语音的质量进行了评估.

1 TMS320DM6446 数字多媒体芯片

TMS320DM6446[1] 是德州仪器 (Ti) 公司提供一种基于达芬奇 (DaVinci) 技术的产品. 包括两个子系统 (DSP 和 ARM) 以及多媒体编解码器、应用编

收稿日期: 2007-09-02
基金项目: 山东省信息产业厅资助 (2005R00012)
作者简介: 马 震 (1980-), 男, 山东滨州市人, 硕士, 研究方向为语音编码、语音合成, 嵌入式系统开发, mazhen_chen@263.net.

程接口(API)等,能大幅降低整体系统开发成本. 下面为 TMS320DM6446 的功能框图.

1.1　TMS320DM6446 功能结构

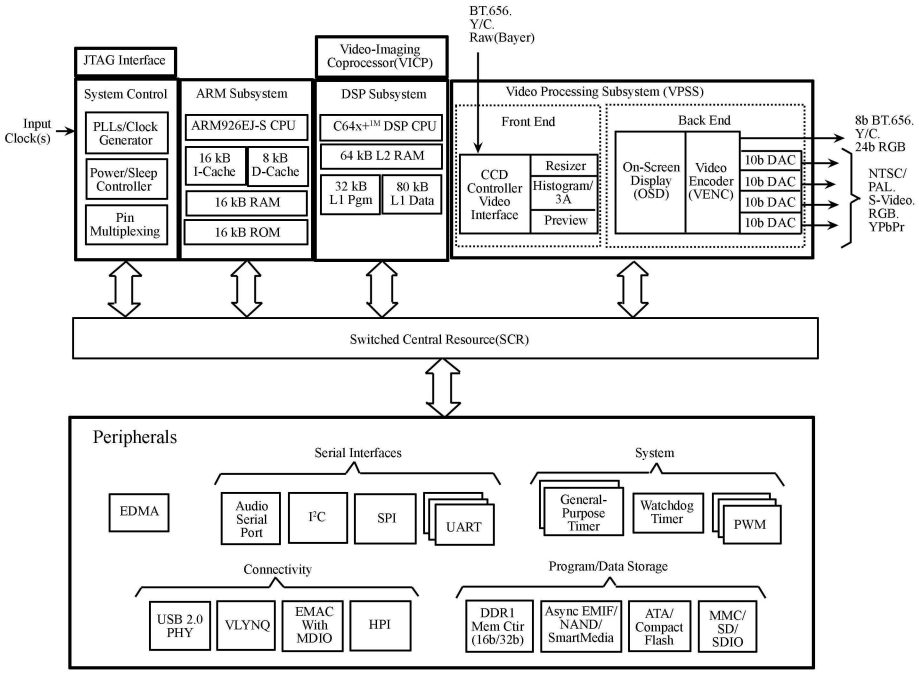

图 1　TMS320DM6446 功能框图

由结构图可以看出,该芯片主要包括两个子系统,分别是低功耗、高性能的 32 位 TMS320C64xDSP 子系统和 ARM926EJ-S 子系统,其工作频率分别达到 594 MHz 和 297 MHz. ARM926EJ-S 子系统具有 16 kB 指令 Cache 和 8 kB 的数据 Cache,TMS320C64xDSP 内核具有 32 kB 程序 RAM/Cache、80 kB 数据 RAM/Cache 及 64 kB 未定义的 RAM/Cache. 由于该芯片采用的是 TMS320C64x DSP 内核,对多媒体处理技术有较好的支持,增强了对语音的编解码能力.

由于语音编解码工作主要围绕 DSP 子系统进行,下面对 TMS320DM6446 的 DSP 子系统进行简单介绍.

1.2　TMS320DM6446 的 DSP 子系统

TMS320C64x DSP 内核是构建在 VelociTI. 2 体系结构的基础上的,是 VelociTI. 2 体系结构的进一步增强,以其 C64x 内核的先进超长指令字(VLIW)结构,获得当前应用设备所需要的音视频处理速度. TMS320DM6446 微处理器有 64 个独立的通道高级 DMA 控制器. DMA 控制器用于可响应内部和外部设备的请求. 在 MPU TI926(ARM926EJ-S)运行条件下,完成外部寄存器/内部寄存器和外部设备之间的数据传输. DMA 的设置取决于 MPU TI926(ARM926EJ-S)内核. 除了上述两款核心功能之外,还带有一些附加外设功能,例如:看门狗(WT)、中断控制器、电源管理控制器、复位控制器,2 个片上振荡器,视频处理子系统(VPSS),异步 EMIFA (NOR Flash,SDRAM),HARD Flash,UART 接口,EMAC 等.

2　MELP 算法模型和关键技术

经典二元激励 LPC 声码器的合成语音可以使人明白谈话的内容,但是它的自然度很差,带有很重的金属声和咔嚓声,难以辨别说话人,主观听觉质量不好. 针对 LPC 模型中存在的一些缺陷,MELP 模型中有针对性的进行了改进. 对于 LPC 中对语音的分类和激励过于简单的问题,MELP 模型采用了周期脉冲和白噪声的混合激励;对于帧内基音周期的时变特性,把语音分为浊音、清音和抖动浊音,对于抖动浊音采用了非周期的脉冲激励;针对二元激励 LPC 忽略了激励信号的形状信息,在 MELP 中编码了激励信号的傅立叶谱的幅度. 而且,MELP 还加入了脉冲散布滤波器和自适应谱增强滤波器来提高语音质量,下面我们就来看一下 MELP 中的几个关键的技术.

2.1　混合脉冲和噪声激励

算法采用多带混合激励模型. 语音被分为 5 个固定的频带 0～500 Hz, 500～1 000 Hz, 1000～2 000 Hz, 2 000～3 000 Hz, 3 000～4 000 Hz, 在每个子带上分别对语音进行浊音强度(Voice

strength)的计算,依据其浊音强度和基音搜索的结果来判断该子带语音是清音、浊音还是抖动浊音.合成时,根据各子带的属性和浊音强度对通过该子带滤波器的噪声、周期脉冲和非周期脉冲进行加权求和,作为激励.采用混合激励,可以减少合成语音中的蜂鸣声.

2.2 非周期脉冲

混合脉冲和噪声激励可以消除 LPC 声码器的嗡嗡声,但是总是无法消除一些孤立的,短暂的音调,会产生一些类似电流声的单音噪声.这是由于在清/浊的过渡段声门脉冲不是很稳定,从而导致激励没有严格的周期性.在过去的 LPC 方法中,这些都是当作浊音来处理,从而会引进一些很奇怪又很刺耳的音调.在 MELP 中,针对这个问题,对基音周期进行抖动,抖动的幅度服从以基音幅度的 ±25 %为上下限的均匀分布.这样就打破了激励信号的严格的周期性,很好的描述了在声门激励不稳定的时候产生的脉冲的特性.

由于这种抖动如果加在强浊音帧的激励中的话,就会破坏语音的周期性,也就是破坏了该帧激励的唯一的参数——基音周期,会使得语音质量急剧变坏,这也是引入抖动浊音这一分类的根本原因.所以要准确的判断,是否可以加抖动,这可以通过计算语音帧的自相关和多峰度来决定.

$$\text{peakness} = \frac{\frac{1}{N}\sum_{n=0}^{N-1} e^2(n)}{\frac{1}{N}\sum_{n=0}^{N-1} e(n)}$$

多峰度的大小描述了峰值是否明显.

2.3 余量信号的傅立叶谱幅度

我们知道各种对语音信号的激励的编码方法,包括多脉冲、码激励、二元激励等都是为了更好的描述余量信号的信息.在浊音中,余量信号所包含的信息量要远远少于原始语音信号,首先在幅度范围上就小于原始语音信号;而且包含较少的几个脉冲,形状上也要比原始语音信号要简单得多.但是,余量信号也是有一定形状的.如果只是采用周期性的脉冲作为激励,虽然可以描述余量信号的周期性信息,却失去了余量信号的形状信息,这就导致合成语音质量的下降.

从时域来描述余量信号的形状信息是比较困难的,因为对于按周期合成的语音来说,周期内的余量信号的长度不定,这就会给量化带来很大的麻烦,所以可以从频域来描述余量信号的形状信息.在 MELP 中,采用对各基音周期谐波处的傅立叶级数幅度进行量化的方法来描述余量信号的形状信息.通过这种方法可以改善 LPC 合成语音中语音嘶哑等弱点,而且提高了抗噪声的性能.

2.4 自适应谱增强

为了使合成语音的频谱能够在共振峰所处的频带与原始的语音信号匹配得更好,我们引入自适应谱增强.这是因为从时域看,合成语音两个主激励之间会衰减的比自然语音要大得多;而从频域看,LPC 滤波器的共振峰要比自然语音的共振峰要缓和.为了解决这个问题,我们引入了自适应谱增强滤波器.它是一个零极点滤波器,由两部分组成.第一部分是一个全极点的滤波器,用来增加每个共振峰的带宽,由 LPC 滤波器在每个 z^{-1} 前面乘以一个因子 A = 0.8 得到;第二部分是为了减小前面全极点滤波器的低通效应而引入,为一个全零点的滤波器,由 LPC 滤波器在每个 z^{-1} 前面乘以一个因子 A = 0.5 后求倒数得到,可以使处于波谷的信号值有所提升.传输函数可以写为:

$$H_{ASE}(Z^{-1}) = \frac{1}{A(0.8q \cdot Z^{-1})} \cdot A(0.5q \cdot Z^{-1})$$

2.5 散布脉冲

为了改善在非共振峰频率处合成语音与原始语音的匹配,MELP 引入了散布脉冲滤波器.在浊音中,每个基音周期并不只是一个主脉冲,声门关闭不完全会产生很多的次要脉冲,引入这个滤波器的目的就是要对合成语音作时域扩展处理,将集中的脉冲激励能量散布开,增加语音的平滑性.MELP 中采用的滤波器的系数可以通过对三角脉冲作 DFT 变换,然后去除掉低频部分,进行归一化处理后再进行 IDFT 变换即可得到.

3 MELP 和 CELP[2] 混合编码

MELP 在稳定的强周期性的语音段,可以得到很高质量的重构语音.但是在爆破音、过渡音、清音段则效果不是很好.这是因为在强周期性的浊音段,有稳定的参数可以提取,而且参数比较简单,最主要的就是基音周期,可以较为精确的描述该语音段的激励信号.而在非周期性或不具备很严格的周期性的语音段则统一的采用白噪声,不能精确的描述语音的激励信号.而码激励线性预测(CELP)编码则因其是以匹配原始语音的波形为目的的,而在非周期或不严格周期语音段有着很好的效果;而这种编码方法却不能象一般的参数编码一样,在强周期性语音段灵活的调节基音周期等,从而精确的匹配原始语音.

从上面的分析,我们可以看到,MELP 和 CELP 这两种编码方法在适用的范围上是互补的,我们下面就讨论一下如何把这两种方法结合在一起.

3.1　MELP/CELP 混合编码简介

在这种混合编码器的编码端,我们把 20 ms 的语音分为一个语音帧. 我们把语音帧分为三类 ——强浊音、弱浊音、清音. 这种分类是根据浊音强度和基音周期来分的. 没有基音周期可言,也就是不具备周期性,而且浊音强度很低的可以分为清音;如果基音周期不是很稳定,变化较快,浊音强度不是很高,这种语音帧我们认为是弱浊音帧;如果基音周期很稳定,浊音强度很高,这种语音帧我们分为强浊音帧. 对于被确定为强浊音的语音帧,我们采用 MELP 方法编码;对于被确定为弱浊音或清音的语音帧,我们采用 CELP 方法编码. 该方法编码器的框图如图 2.

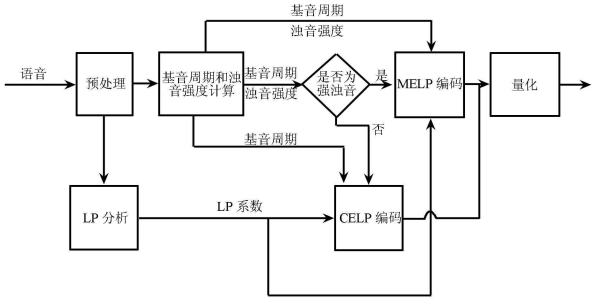

图 2　混合编码器

由于 CELP 是以保持原始语音的波形为目的的,所以它的合成语音会与原始语音保持时间同步. 而作为一种典型的参数编码的 MELP 则主要是提取并编码激励信号的波形参数. 在强浊音帧使用 MELP 编码,主要是提取了原始语音的周期方面的参数,而没有保留相位方面的信息,也就无法与原始语音保持时间的同步.

把参数编码和基于波形的编码相结合在[3-4]中提出过. 但是如果简单的在两者之间切换就会引入噪声,我们在 MELP 中加入了对齐相位,并编码它,从而保证和原始语音的时间同步.

3.2　量化方案

在编码端产生的参数,无论是强浊音帧,还是弱浊音帧、清音帧,都需要编码的是 LP 系数、增益、基音周期,还有一个标志 MELP/CELP 的 bit. 在 MELP 编码的帧中,还要传输傅立叶谱幅度、对齐相位和浊音强度;在 CELP 编码的帧中,要传输固定码本和码本增益. 具体的 bit 分配如下表:

表 1　bit 分配表

参数	bit 数	
	MELP	CELP
LP 系数	24	19
增益	8	5
基音周期	8	5
对齐相位	6	
傅立叶谱幅度	22	
浊音强度	6	
固定码本		44
码本增益		5
保留字	3	
MELP/CELP 标志	1	1
校验位	2	1
总计	80	80

3.3　试验结果

采用上面论述的混合编码器对一段时长 2.5960 s,20768 个样点的语音进行处理. 可以得到与 32 kbit/s 的 ADPCM 相近的语音质量,具有很好的自然度和清晰度. 非正式 MOS 得分可以达到 3.80.

进而,采用混合编码器对一段 3.6590 s,29272 个样点的含白噪声的语音进行编解码处理. 与 32 kbit/s 的 ADPCM 进行 A/B 测试,选择混合编码器语音与选择 ADPCM 语音人数比为 17 ∶13. 非正式 MOS 得分为 3.45. 采用同样的方法,对加入汽车噪声的语音进行处理,A/B 测试结果为 16 ∶14,非正式 MOS 得分为 3.65. 从试验结果来看,对于加入噪声的语音,混合激励编码器也能很好的进行处理.

4　结论

论述了 MELP 声码器的几项关键技术,并进一步把 MELP 与 CELP 相结合,提出了混合声码器的方法. 试验证明,这种技术是有效的,具有和 32 kbit/s 的 ADPCM 相近的语音质量. 从理论上来讲,由于刻画了激励的形状,此种方法应该具备在噪声环境中优于参数编码的性能,试验证明语音质量能与 ADPCM 接近,具有良好的抗噪声能力.

参考文献:

[1]　http://focus.ti.com/general/docs/tidc/tms320c6446.pdf[EB/OL].

[2]　Schroeder M R and Atal B S. Code-Excited Linear Predictive (CELP): High Quality Speech at Very Low Bit Rates,[C]// Proc. IEEE Int. Conf. ASSP,(Tampa),Mar. 1985,937-940.

[3]　Thomson D L and Prezas D P. Selective Modeling of the LPC Residual During Unvoiced Frames; White Noise or Pulse Excitation,[C]// Proc. IEEE Int. Conf. ASSP,(Tokyo) Apr. 1986,3087-3090.

[4]　Kleijn W B. Encoding Speech Using Prototype Waveforms[C]// IEEE Trans. on Speech and Audio Procession,Oct. 1993,1,386-399.