

语音压缩编码技术研究进展

时 磊, 尚秋峰, 李健宇

(华北电力大学, 河北保定 071003)

摘要: 为了满足数字通信及其它商业应用的需求, 语音压缩编码技术得到了迅速发展。介绍了目前语音压缩编码技术的研究进展, 主要包括连续可变斜率增量调制(CVSD)、小波分析、多脉冲激励线性预测编码、散布脉冲码激励线性预测(DP-CELP)、多重脉冲散布非均匀代数码本激励线性预测(MPD-USACELP)、波形内插(WI)、线谱对(频率)(LSP)的量化。对以上算法进行了分析比较, 总结了它们的特点和适用范围, 并介绍了其中一些算法在 DSP 上的实时实现。

关键词: 语音压缩编码; 线性预测(LP); 波形内插; DSP

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-3630(2008)-04-0553-04

The research progress in speech compression coding technology

SHI Lei, SHANG Qiu-feng, LI Jian-yu

(North China Electric Power University, Baoding 071003, Hebei, China)

Abstract: In order to satisfy demands of the digital communication and other commercial applications, the speech compression technology has been developed rapidly. The present research progress in speech compression technology is introduced in this paper including CVSD, wavelet analysis and its application to speech coding, MPLPC, DP-CELP, MPD-USACELP, WI and quantification of LSF. These algorithms are analyzed and compared. Their characteristics and applicable scopes are summarized. Some algorithms applied to DSP are also introduced.

Key words: Speech compression coding; LP; waveform interpolation; DSP

1 引 言

语音是人类最重要、最有效、最常用和最方便的交换信息的形式, 是人们思想疏通和情感交流的最主要途径。在实际的语音通信中, 有些信道难以扩宽且质量很差; 有些信道正被广泛使用, 短期内难以更新; 有些昂贵的信道, 每压缩一个比特都意味着节省开支^[1]。因此, 语音压缩编码无疑在语音通信及人类信息交流中占有举足轻重的地位。

根据语音编码速率的不同, 可以划分为以下的编码算法: 高速率编码 16kb/s~64kb/s、中速率编码 4.8kb/s~16kb/s、低速率编码 2.4kb/s~4.8kb/s、极低

速率编码即小于 2.4kb/s。本文按此分类方法, 将文章中涉及到的语音压缩编码方法分为中高速率编码、低速率编码与极低速率编码三类。

2 主要方法

对于中高速率的语音压缩编码而言, 高速率编码主要以波形编码为主, 其适应能力强、语音质量好, 但编码速率较高。中速率编码以混合编码为主, 基于语音产生模型的假定并采用了分析合成技术, 同时又利用语音的时间波形信息, 增强了重建语音的自然度。低速率及极低速率编码以参数编码为主, 通过建立语音信号的产生模型, 提取代表语音信号特征的参数来编码。

2.1 中高速率编码

2.1.1 连续可变斜率增量调制(CVSD)

收稿日期: 2007-06-23; 修回日期: 2007-09-13

作者简介: 时磊(1983-), 男, 河北保定人, 硕士研究生, 研究方向: 语音信号处理。

通讯作者: 时磊, E-mail: alalar@163.com

CVSD (Continuously Variable Slope Delta Modulation) 针对输入信号振幅的变化率, 调节量化步长的增量值。当输入信号的振幅变化率增大时, 增量值也相应地增大; 当输入信号的振幅变化率减小时, 增量值也相应地减小。文献[2]利用 TI 公司的 TMS320VC5509 定点 DSP 芯片实现多路全双工 16kbps CVSD 语音编解码的方案, 并基于摩托罗拉公司的 MPC800 提出了利用嵌入式系统扩展 DSP 应用的方法。

与其它编码方式相比, CVSD 有如下突出特点: (1) 较高的抗误码能力; (2) 算法简单, 硬件实现容易; (3) 单路应用时无须采取码元与码组同步措施等。因此, 它是一种在卫星通信、移动通信和军事通信等领域应用非常广泛的波形编码方式。

2.1.2 基于小波分析的语音编码方法

小波在时域频域同时具有良好的局部特性, 并且它的基具有多样性, 使得它优于傅立叶分析, 更适合于语音。它不仅可以单独的处理语音, 还可以和其它理论结合用于语音压缩。

小波技术与离散余弦变换(DCT)结合: 离散余弦变换系数能较好地体现语音信号能量在频域中集中分布的特性, 可以重构出高质量的话音, 如果采用矢量量化(VQ)的方法直接对离散余弦变换的系数进行编码, 运算量较大。可以利用小波变换的多分辨分析特性, 对离散余弦变换系数进行离散二进小波变换后, 再对其小波系数进行编码^[3]。其优点为: (1) 编码算法简单; (2) 可方便灵活地控制各频率段的位数分配, 满足不同编码质量的要求; (3) 在 16kbps 数据率时, 可实现与原语音质量相同的重构语音。

2.1.3 多脉冲激励线性预测编码(MPLPC)的改进

原始的 MPLPC 方法中, LP 滤波器的阶数和激励脉冲的个数都是固定的。文献[4]对其改进算法使用了可变阶数滤波器的语音编码方法。主要思想是: 对于不同的语音帧使用不同阶数的线性预测滤波器, 只要低阶滤波器满足要求就不使用高阶滤波器。由于各语音帧滤波器阶数不同, 使用的编码位数也不同。为了使整个语音传输速率基本恒定, 对于滤波器阶数较低的语音帧, 可以相应地增加激励脉冲的个数或增加各脉冲的比特数。

用 MATLAB 对改进 MPLPC 方法进行模拟, 结果表明, 此方法可以在 6kb/s 以下的速率上得到有一定自然度的合成语音, 比同速率下的原始多脉冲激励方法有明显改善。但此方法在有噪声和失真的

信道上传输的情况需要进一步实验, 每帧激励脉冲个数的最佳值需要进一步研究。

2.2 低速率编码

2.2.1 散布脉冲码激励线性预测(DP-CELP)

码激励线性预测(CELP)算法是最近 20 年来语音编码中最为成功的一项技术。但是当速率降至 4 kb/s 以下时, 为了提高量化效率, 必须增加激励矢量的长度, 这时还用很少的符号脉冲来描述激励信号就会导致语音质量的急剧下降, 这就是传统的 CELP 算法在 4kb/s 速率以下质量难以提高的根本原因。

为了解决传统 CELP 算法在 4 kb/s 速率以下质量难以提高的问题, 文献[5]提出了一种散布脉冲 CELP(DP-CELP)语音编码算法。激励矢量由特殊结构的代数码书与固定形式的散布脉冲的卷积获得, 在不增加代数码书搜索复杂度的前提下, 较好地解决了传统代数码书中符号脉冲不足带来的缺陷, 有效地改善了重建语音质量。非正式的主观听力测试表明, 这种 4kb/s DP-CELP 语音编码算法的合成语音质量非常接近 G.723.1 中 6.3kb/s 语音编码器, 可产生高质量的合成语音。

在此算法的基础上, 文献[6]开发了一种基于 TMS320VC5410 定点 DSP 的实时低速率语音压缩系统, 将低比特率语音编码算法应用于实际的语音处理系统。为了降低运算复杂度, 算法程序使用 DSP 汇编语言实现。根据 C5410 定点 DSP 的指令集特性, 以及算法自身的特点, 对整个算法进行了 2 次整体优化和若干次的局部优化, 达到了实时实现的要求。由此, 产生出了性价比极高的语音编解码系统, 在数字存储、多媒体通信等领域有着广泛的应用前景。

2.2.2 多重脉冲散布非均匀代数码本激励线性预测(MPD-USACELP)

由于 CELP 算法在 4kb/s 速率以下质量难以提高, 文献[7]在编码器和解码器中分别引入了不同的脉冲散布技术, 从而达到了利用较少比特提供高质量激励的目的。根据理想代数码本脉冲在子帧中各位置的概率分布规律, 提出了新的非均匀结构代数码本, 提高了代数码本的效率, 并且引入了基音预加重技术优化了自适应码本搜索。

经主观听音和计算机模拟测试, 最后形成的 3.35kb/s 多重脉冲散布非均匀代数码本激励线性预测(MPD-USACELP)语音编码算法的质量达到了北美 IS-54 标准 8 kb/s 算法的水平。

2.3 极低速率编码

2.3.1 波形内插(WI)算法的改进

波形内插语音编码^[8]在 2kb/s~4kb/s 的速率上表现出良好性能,并产生了高质量合成语音。但是这种编码算法的计算复杂度非常高,很难在实际中应用,因此在保持语音质量的前提下,降低 WI 模型的计算复杂度势在必行。

基本的 WI 编码算法需要在编、解码端都要进行复杂的特征波形(CW)对齐,以保证提取的 CW 和合成的 CW 在相位上一致,这使得实时实现整个算法变得非常困难。为了减少这种复杂度,文献[9]提出一种在编码端去除对齐过程,译码端采用三次 B 样条的插值模型。通过对慢渐变波形只量化低频分量,而快渐变波形用正交多项式拟合,实现了变维矢量到定维矢量的转换,进一步降低了算法的复杂度。DRT 测试结果表明,此 2 kb/s 语音编码方法能获得高可懂度的重建语音。

文献[10]在 WI 模型的原理的基础上,针对模型中占主要计算量的 2 个模块(特征波形的表示和对齐)进行了相应的改进。原 WI 模型中的特征波形使用了离散傅里叶级数表示, CW 对齐也是通过频域傅氏级数实现的, CW 的表达和对齐都没有使用快速算法。在改进算法中,先对时域的 CW 尾部补 0,使它的长度为 2 的整次幂,然后作快速傅里叶变换得到 CW 的频域表示。因为时域补 0 起到了对原频域做插值的作用,并不改变原频谱形状,所以再利用三次 B 样条插值技术,就可以得到相应的傅氏级数;对于 CW 对齐则先将它的周期延拓为 2 的整次幂,再利用圆周相关定理通过 FFT 计算出对齐所需要的位移。改进后的算法相对独立没有影响原编码器其它模块,在保持高质量重建语音的前提下使复杂度大幅下降,更易于在单片 DSP 芯片上实现。

2.3.2 奇异值分解的波形内插(SVD-WI)算法

波形内插算法中,针对奇异值分解(Singular Value Decomposition,简称 SVD)理论^[11]用于特征波形的分解时,算法延时长、计算复杂和高比特量化参数的问题,文献[12]提出了一种新颖的基于奇异值分解的特征波形的分解与量化模型。基于降低运算复杂度的考虑,利用语音信号的感知特性,根据所含频率成分的不同,将特征波形的幅度谱分成基本矩阵、过渡矩阵和补充矩阵三个矩阵,并各自采用不同的量化方法;由于特征波形谐波数是时变的,将三个矩阵分为三种组合模式表示特征波形幅度谱,提高了特征波形的表示精度,合成端可以根

据不同的精度要求重建语音。这种方法比传统的 WI 语音编码器减少了一帧延时,主测试表明, SVD-WI 编码器的重建语音质量要好于 2.4kb/s 混合激励线性预测(MELP)编码器。

2.3.3 线谱对(频率)的量化算法

线性预测是许多低速率语音编码方案的核心技术。为降低编码速率,选用高效的线性预测系数编码方法至关重要。线谱对(line spectrum pair,简称 LSP)即线谱频率(line spectrum frequency,简称 LSF)是线性预测系数的等价参数。因此,研究高效率的 LSP 参数量化方法对低速率语音编码算法的研究具有极其重要的意义。

在极低速率的声码器算法研究中,为了实现 LSP 参数的高质量低速率量化,文献[13]提出了一种新的量化算法,称作 P-RS-MSMQ(Pattern-based Residual Split Multi-Stage Matrix Quantization)。此算法在多帧联合矩阵量化的框架下,提出了以下几点新的策略:(1)基于超级帧模式的均值去除和基于超级帧模式转移的帧间预测;(2)将预测余量矩阵分裂成子矩阵,并对各子矩阵分别进行多级矩阵量化;(3)对多帧联合量化引入新的加权策略等。

通过这些措施,有效的提高了量化性能,能够在极低的编码速率下获得高质量的量化效果并且大大降低了量化码本的存储量和运算量,提高了算法的实用性。此算法已经成功地实现了高质量的 1.2kb/s 和 600 b/s 语音编码算法。因此,它可以广泛应用于极低速率语音编码算法中。

为了降低线谱频率(LSF)参数矢量量化器的搜索复杂度和码字存储单元,文献[14]利用 LSF 参数的帧内和帧间相关性,设计了一种 LSF 参数的预测式多级分裂矢量量化器。该量化器对 LSF 参数的预测残差矢量进行两级矢量量化。其中第 2 级的误差矢量分裂成 2 个维数分别为 4 和 6 的子矢量进行矢量量化,采用瞬时联合多级矢量量化器设计算法设计码本,应用 M-L 树搜索算法搜索码字,降低了搜索复杂度和码字存储单元,同时保持了高效的量化性能。

3 结 论

本文介绍了当前语音压缩编码的研究现状,将这些方法按照编码速率进行了分类。其中,小波分析本身是一门交叉学科,小波分析与其它理论的综合运用是今后语音处理技术发展的必然趋势;编码速率在 4kb/s 以下时,为提高码激励线性预测的语音

质量, 对其算法的改进无疑成为当前的研究热点; 作为最有潜力的波形内插语音编码算法, 其今后研究的重点为: 与其它低速率语音编码技术更为合理的相互补充; 特征波形提取和对齐问题; 特征波形分解问题; 特征波形量化问题; 相位恢复问题^[15]。总体来看, 高质量的极低速率编码算法更受到人们的青睐, 不仅是目前的研究热点, 而且是未来的发展方向。另外, 将性能好的算法进行优化和改进, 从而在 DSP 上实时实现则是另一个发展方向。

参 考 文 献

- [1] 郭莉, 郭丽红, 王炳锡. 极低速率语音编码的新发展与应用[J]. 电声技术, 2005, (6): 38-39.
GUO Li, GUO Lihong, WANG Bingxi. New developments and applications of very low bit rate speech coding[J]. Voice technology, 2005, (6): 38-39.
- [2] 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.
GUO Qiuping, XIANG Jie. Implementation of CVSD voice codec on embedded system[J]. Computer Engineering, 2006, 32(16): 213-214.
- [3] 苟大举, 苟平, 周群彪. 语音 DCT 变换的一种小波编码方法[J]. 四川大学学报(自然科学版), 2004, 41(6): 1153-1157.
GOU Daju, GOU Ping, ZHOU Qunbiao. A wavelet transform coding for voice's DCT[J]. Journal of Sichuan University(Natural Science Edition), 2004, 41(6): 1153-1157.
- [4] 蒋保臣, 马震. 改进 MPLPC 语音编码方法[J]. 计算机应用与软件, 2005, 22(8): 77-79.
JIANG Baochen, MA Zhen. Modified MPLPC speech coding method[J]. Computer Applications and Software, 2005, 22(8): 77-79.
- [5] 鲍长春, 高质量的 4kb/s 散布脉冲 CELP 语音编码算法[J]. 电子学报, 2003, (2): 309-313.
BAO Changchun. A high quality dispersed-pulse CELP speech coding algorithm at 4kb/s[J]. Acta Electronica Sinica, 2003, 2: 309-313.
- [6] 鲍长春, 白燕宁, 窦庚欣. 高质量散布脉冲 CELP 算法的定点 DSP 实时实现[J]. 北京工业大学学报, 2006, 32(5): 385-390.
BAO Changchun, BAI Yanning, DOU Gengxin. Real-time implementation of high-quality dispersed-pulse CELP algorithm via fixed-point DSP[J]. Journal of Beijing University of Technology, 2006, 32(5): 385-390.
- [7] 李悦, 唐昆, 崔慧娟, 等. 3.35kb/s 低速率语音编码算法[J]. 清华大学学报(自然科学版), 2004, 44(10): 1410-1413.
LI Yue, TANG Kun, CUI Huijuan, et al. 3.35kb/s low bit-rate speech coding algorithm[J]. Journal of Tsinghua Univ (Sci&Tech), 2004, 44(10): 1410-1413.
- [8] Gottesman O, Gersho A. Enhanced waveform interpolation coding at low bit-rate. Speech and audio processing[J], IEEE Transaction, 2001, 9(8): 786-798.
- [9] 朱娜娜, 鲍长春, 李靓. 一种新的 2kbit/s 波形内插语音编码器[J]. 通信学报, 2004, 25(11): 70-76.
ZHU Nana, BAO Changchun, LI Liang. A new waveform interpolation speech coder at 2kbit/s[J]. Journal on Communications, 2004, 25(11): 70-76.
- [10] 齐峰岩, 鲍长春. 波形内插语音编码中特征波形表达和对齐快速算法[J]. 北京工业大学学报, 2006, 32(6): 514-519.
QI Fengyan, BAO Changchun. Fast algorithms on characteristic waveform representation and alignment in waveform interpolation speech coding[J]. Journal of Beijing University of Technology, 2006, 32(6): 514-519.
- [11] LUKASIAK J, BURNETT I S. Scalable decomposition of speech waveforms[A]. Tsukuba: IEEE Workshop on Speech Coding Proceedings[C]. 2002. 135-137.
- [12] 王贵平, 鲍长春, 张鹏. 基于奇异值分解的低速率波形内插语音编码算法[J]. 电子学报, 2006, (1): 135-140.
WANG Guiping, BAO Changchun, ZHANG Peng. Low bit rates waveform interpolation speech coding base on singular value decomposition[J]. Acta Electronica Sinica, 2006, (1): 135-140.
- [13] 李军林, 崔慧娟, 唐昆. 极低速率语音编码中 LSP 参数的高效量化算法[J]. 清华大学学报(自然科学版), 2004, 44(10): 1422-1425.
LI Junlin, CUI Huijuan, TANG Kun. Efficient LSP quantization algorithm for very low bit rate speech coding[J]. Journal of Tsinghua Univ (Sci&Tech), 2004, 44(10): 1422-1425.
- [14] 李靓, 鲍长春, 白燕宁. 一种高效、低存储的线谱频率参数矢量量化器[J]. 北京工业大学学报, 2005, 31(2): 130-135.
LI Liang, BAO Changchun, BAI Yanning. An efficient LSF parameters quantizer with low storage[J]. Journal of Beijing University of Technology, 2005, 31(2): 130-135.
- [15] 李靓, 鲍长春, 王贵平. 波形内插语音编码技术回顾与展望[J]. 通信学报, 2005, 26(1): 95-103.
LI Liang, BAO Changchun, WANG Guiping. Review of waveform interpolation speech coding[J]. Journal on Communications, 2005, 26(1): 95-103.