

G.729.1 语音编码算法的改进及实现

刘丽群, 黄冰, 宁瑞芳

(桂林电子科技大学 信息与通信学院, 广西 桂林 541004)

摘要: 在 G.729.1 语音编码算法中引入了一种长度为 $N=5 \times 2^m$ ($m \geq 2$) 的 MDCT/IMDCT 实现的快速算法, 编程实现该算法并在 cygwin 中仿真。在本方法中用孤立预乘的方法将 N 点的 MDCT 映射成 $N/2$ 点的 DCT-II 和 DCT-IV, 然后进入加窗阶段。该窗口是一种部分对称的窗, 可以存储使用 $N/2$ 的字, 同时包含了一种只需 4 次乘法在一定的条件下就可以分解 5 点 DCT-II 的方法。与其他算法相比, 该算法可以明显地降低复杂度和减少存储空间。

关键词: G.729.1; 语音编码; 余弦变换; 改进的离散深余弦变换

中图分类号: TP391.42

文献标识码: A

文章编号: 1674-7720(2010)17-0030-03

Improvement and implementation of G.729.1 speech coding algorithm

LIU Li Qun, HUANG Bing, NING Rui Fang

(College of Information and Communication, Guilin Electronics and Science Technology University, Guilin 541004, China)

Abstract: This paper describes an efficient algorithm for implementing MDCT/IMDCT of length $N=5 \times 2^m$ ($m \geq 2$) in G.729.1. The algorithm has implemented in float-point on the platform of cygwin. It utilizes a mapping of MDCT of size N into $N/2$ -point DCT-IV and DCT-II with isolated pre-multiplications, which are subsequently moved in the windowing stage. The window is piece-wise symmetric, and can be store $N/2$ words. The algorithm also uses an efficient factorization of 5 point DCT-II which requires only 4 multiplications. This algorithm offers practically appreciable reduction in complexity and memory usage.

Key words: G.729.1; speech coding; DCT; MDCT

G.729.1 是基于分裂带宽的编码器, 具有灵活的结构, 通过利用正交镜像滤波器组分析和合成滤波器组不仅能够处理采样率为 16 kHz 的输入语音, 还可以处理 8 kHz 采样率的输入语音, 它是一个基于 G.729 的带宽扩展 (50~7 000 Hz) 的、8~32 Kb/s 可分级编码器。由编码器产生的比特流具有可分级性, 包含了 12 个嵌入式层。第 1 层为核心层, 其编码速率为 8 Kb/s, 核心层与原有 G.729 中的比特流形式相同, 这使得 G.729.1 与 G.729 之间具有互操作性。第 2 层为窄带增强层, 每层均在之前一层的基础上增加 2 Kb/s 的编码速率, 编码速率分别为 14 Kb/s、16 Kb/s、18 Kb/s、20 Kb/s、22 Kb/s、24 Kb/s、26 Kb/s、28 Kb/s、30 Kb/s、32 Kb/s。编码器的默认输入为 16 kHz 采样, 16 bit 线性脉冲调制编码语音信号, 也可对 8 kHz 采样的信号进行操作。解码器的输出可以为 16 kHz 或 8 kHz 采样的 16 bit 线性 PCM 语音信号^[1]。

G.729.1 编解码器基于三阶段结构: 嵌入式码本激

励线性预测 (CELP) 编解码、时域带宽扩展 (TDBWE) 以及时域混叠消除 (TDAC) 的估计转换编解码。其中, 嵌入式 CELP 阶段产生第 1 和第 2 层, 生成 8 Kb/s 和 12 Kb/s 窄带合成信号 (50~4 000 Hz)。TDBWE 阶段产生在第 3 层, 生成 14 Kb/s 宽带输出信号 (50~7 000 Hz)。TDAC 阶段工作在改进离散余弦变换域 (MDCT), 生成第 4~12 层, 将信号从 14 Kb/s 提高到 32 Kb/s。TDAC 编解码同时代表 50~4 000 Hz 频带加权 CELP 编解码误码信号和 4 000~7 000 Hz 频带输入信号。算法中处理帧长为 20 ms, 每帧宽带信号样点数为 640^[2]。

本文主要对其 TDAC 部分的 MDCT 变换进行改进并用程序实现。在 G.729.1 算法的时延中, 用于 MDCT 分析 (估计) 的时延为 20 ms, 通过 MDCT 的快速算法来减少该部分的时延, 从而提高算法的实时处理的性能。

改进的离散深余弦 (MDCT) 广泛应用于语音和音频编码、分析和研究滤波器的时域别名取消特性。这是一

种最为繁琐的计算密集型算法,因此需要一种快速算法来实现,到目前为止,已经提出了多种快速算法来实现 MDCT。在这其中许多算法推导变换的长度为 $N=2^m$ 。本文将介绍一种基于长度为 $N=5 \times 2^m$ 的新方法。这种长度的变换已建议在语音和音频编码中使用,典型的采样速率为 8 kHz 或 16 kHz,帧延迟为 10 ms 或 20 ms。ITU-TG.729.1 和 3GPP2EVRC-WB 以及新兴的 ITU-TG..EV-VBR 标准中也提供了这种算法。本文对 G.729.1 中的 MDCT 引入了一种新的算法,该算法更简便和更易于实际应用,非常适合在语音算法中应用。

1 算法的基本原理

设 $\{x(n)\}, n=0, \dots, N-1$ 代表输入采样序列, N 为帧长。在本文中, MDCT 和 IMDCT 的定义如下:

$$X(k) = \sum_{n=0}^{N-1} x(n) \cos\left(\frac{\pi}{2N} (2n+1 + \frac{N}{2})(2k+1)\right), k=0, \dots, \frac{N}{2}-1 \quad (1)$$

$$\tilde{x}(n) = \sum_{k=0}^{\frac{N}{2}-1} X(k) \cos\left(\frac{\pi}{2N} (2n+1 + \frac{N}{2})(2k+1)\right), n=0, \dots, N-1$$

式中, $X(k)$ 是 MDCT 的系数, $x(n)$ 是重建样点。为了简便,本文中忽略了定义中的正常化因子,将 $(N/2 \times N)$ MDCT 定义为矩阵 M :

$$M(i, j) = \cos\left(\frac{\pi}{2N} (2n+1 + \frac{N}{2})(2k+1)\right) \quad (2)$$

$$i=0, \dots, \frac{N}{2}-1$$

$$j=0, \dots, N-1$$

所以 MDCT/IMDCT 可以记为:

$$X = MX \quad (3)$$

$$x = M^T X \quad (4)$$

式中, $X = [X(0), \dots, X(\frac{N}{2}-1)]^T$

$$x = [x(0), \dots, x(N-1)]^T$$

$$\tilde{x} = [\tilde{x}(0), \dots, \tilde{x}(N-1)]^T$$

将 N 点的 MDCT/IMDCT 映射成为 $N/2$ 点的 DCT-IV 如下^[3]:

$$M^T = PSC_{N/2}^{IV} \quad (5)$$

$$M = C_{N/2}^{IV} SP^T \quad (6)$$

式中 P 为:

$$P = \begin{bmatrix} 0 & I_{N/4} \\ 0 & I_{N/4} \\ J_{N/4} & 0 \\ J_{N/4} & 0 \end{bmatrix}$$

其中 $I_{N/4}$ 和 $J_{N/4}$ 为 $N/4 \times N/4$ 的单位矩阵及其相应的逆转矩阵。

$$S = \begin{bmatrix} -I_{N/4} & 0 \\ 0 & I_{N/4} \end{bmatrix}$$

$C_{N/2}^{IV}$ 是一个 $N/2 \times N/2$ 维的 DCT-IV 矩阵,定义如下:

$$C_{N/2}^{IV}(i, j) = \cos\left(\frac{\pi}{2N} (2i+1)(2j+1)\right), i, j=0, \dots, \frac{N}{2}-1$$

根据对称特性可以将 DCT-IV 矩阵映射成为 DCT-II 矩阵,结果如下^[4]:

$$C_{N/2}^{IV} = D(C_{N/2}^{II})^T L^T$$

式中, D 是一个对角线矩阵,其元素为:

$$D(i, i) = 2\cos\left(\frac{\pi}{2N} (2i+1)j\right), i, j=0, \dots, \frac{N}{2}-1 \quad (7)$$

$$L = \begin{bmatrix} 0.5 & 0 & 0 & 0 & \dots & 0 \\ -0.5 & 1 & 0 & 0 & \dots & 0 \\ 0.5 & -1 & 1 & 0 & \dots & 0 \\ -0.5 & 1 & -1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -0.5 & 1 & -1 & 1 & \dots & 1 \end{bmatrix}$$

$C_{N/2}^{II}$ 是一个 $N/2 \times N/2$ 的 DCT-II 的矩阵,其定义如下:

$$C_{N/2}^{II}(i, j) = \cos\left(\frac{\pi}{2N} (2i+1)j\right), i, j=0, \dots, \frac{N}{2}-1 \quad (8)$$

通过以上的变换可以看到,矩阵 D 和窗函数相乘后所得到的结果遵循 IMDCT,这样就可以减少做 $N/2$ 次的乘法,也即可以节省 $N/2$ 的存储空间。MDCT 可以通过逆变换从 IMDCT 中得到。然后进行逆 DCT-II 变换。变换需要一种算法可以适合任意长度的序列。因此本文采用了既对频率状态进行抽取同时又可以执行递归直到 DCT 的长度为 5 点的方法。基于 5 点的 DCT-II 是通过平面旋转的方法来取代系数 $x(3)$ 与 $x(4)$ 之间的 3 次乘法和 3 次加法,用同样的方法对 $x(0)$ 和 $x(2)$ 进行处理。该方法使处理的路径变短、结构的规律性增强、传输过程中数据的动态范围减少。基于 5 点的 DCT-II 变换的方程如下:

$$C^H(0) = x(0) + x(1) + x(2) + x(3) + x(4)$$

$$C^H(1) = \beta\gamma[x(1) - x(3)][x(4) - x(0)]$$

$$C^H(2) = \alpha[x(0) + x(1) + x(3) + x(4)] - x(2) + \delta[x(0) + x(4) - x(1) - x(3)] \quad (9)$$

$$C^H(3) = \beta\gamma[x(1) - x(3)][x(0) - x(4)]$$

$$C^H(4) = x(2) - \alpha[x(0) + x(4) + x(1) + x(3)] - \delta[x(1) + x(3) - x(0) - x(4)]$$

式中, $\alpha = \frac{1}{4}$; $\beta = \cos\left(\frac{3\pi}{10}\right)$; $\gamma = -\cos\left(\frac{\pi}{10}\right)$; $\delta = -\frac{\sqrt{5}}{4}$

逆 DCT 的变换方程如下:

$$a_1 = C^H(2) + C^H(4); a_2 = C^H(4) - C^H(2)$$

$$b_1 = C^H(0) + a_2; b_2 = C^H(0) - \delta a_2 \quad (10)$$

$$b_3 = \beta C^H(1) + \gamma C^H(3); b_4 = -\beta C^H(3) + \gamma C^H(1)$$

$$\begin{aligned} b_5 &= b_2 + \delta a_1; b_6 = b_2 - \delta a_1 \\ x(0) &= b_6 - b_4; x(1) = b_3 + b_5; x(2) = b_1 \\ x(3) &= b_5 - b_3; x(4) = b_4 + b_6 \end{aligned} \quad (11)$$

由于 DCT-IV 中的前项和后项乘法可以在 MDCT 和 IMDCT 分别合并窗函数,因此在许多应用实例中,MDCT 都是使用下面的窗函数:

$$h(n) = \sin\left(\frac{(2n+1)\pi}{2N}\right), n=0, \dots, N-1 \quad (12)$$

用该窗函数和传输因子整合成新的改进的窗函数,其表达式为:

$$w(n) = \begin{cases} 2\cos\left(\frac{\pi}{2N}(2n+\frac{N}{2}+1)\right)\sin\left(\frac{\pi(2n+1)}{2N}\right), n=0, \dots, \frac{n}{4}-1 \\ -2\cos\left(\frac{\pi}{2N}(2n+\frac{N}{2}+1)\right)\sin\left(\frac{\pi(2n+1)}{2N}\right), n=\frac{N}{4}, \dots, N-1 \end{cases} \quad (13)$$

当 $0 \leq n < \frac{N}{4}$ 时:

$$w\left(\frac{n}{4}-1-n\right) = 2\sin\left(\frac{\pi(2n+1)}{2N}\right)\cos\left(\frac{\pi}{2N}\left(2n+\frac{N}{2}+1\right)\right) = w(n) \quad (14)$$

当 $0 \leq n < \frac{3N}{4}$ 时:

$$w\left(n+\frac{N}{4}\right) = 2\sin\left(\frac{\pi(2n+1)}{2N}\right)\sin\left(\frac{\pi}{2N}\left(2n+\frac{N}{2}+1\right)\right) = w(N-1-n) \quad (15)$$

由此说明该窗是准对称的,与对称的正弦窗相比较,该窗也使用相同大小的存储空间。

2 算法的复杂度分析

设: $N=5g2^m$ ($m \geq 2$) 和 $N_f=2^m$, $RM_f(N)$ 和 $RA_f(N)$ 分别代表 N 点 MDCT 的乘法和加法的复杂度, $RM_l(N)$ 和 $RA_l(N)$ 分别代表 N 点 IMDCT 的乘法和加法的复杂度, $RM_d(5)$ 和 $RA_d(5)$ 分别代表快速 5 点 DCT-II 的乘法和加法的复杂度。进一步计算 MDCT 和 IMDCT 的复杂度用下面的计算公式:

$$\begin{aligned} RM_f(N) &= RM_l(N) = N + \frac{N_f}{2} RM_d(5) + \frac{N(m-1)}{4} \\ RA_f(N) &= N + \frac{N_f}{2} RA_d(5) + \frac{3N(m-1)}{4} - \frac{N_f}{2} \\ RA_l(N) &= \frac{N}{2} + \frac{N_f}{2} RA_d(5) + \frac{3N(m-1)}{4} - \frac{N_f}{2} \end{aligned} \quad (16)$$

由式(16)可知,MDCT/IMDCT 算法的复杂度受到快速 5 点 DCT-II 算法复杂度的限制。

3 算法的实现

经过改进后,新的 MDCT 变换算法如图 1 所示。图中,加窗所用的窗函数为一般情况下常使用的窗函数。

从以上分析可知,该算法的复杂度受到 DCTII 算法复杂度的限制,所以在做 DCTII 变换时用的是 HEIDEMAN M T 的基于 5 点的 DCTII 变换的算法。该算法进



图 1 改进后新的 MDCT 变换算法

行 5 点的 DCTII 变换只需 14 次加法和 4 次乘法,而可以提高运算速度和节省运算时间。

本文在 G.729.1 中的 TDAC 模块中引进了 MDCT 的快速算法,在 G.729.1 的默认模式下,该部分的时延为 20 ms。通过使用本文中的方法进行改进,可以使该部分的时延降低约 1/3,其存储空间也可节省约 1/3。从而提高了该算法实现的实时性以及实际应用有了更大的可能性。在原来 G.729.1 的基础上实现了变速率,使该算法可以与 G.729 实现互操作,即该算法包含了 G.729(因在原来 G.729 的基础上引入了 TDBMW 和 TDAC 技术)。

本改进算法的仿真结果(默认模式)如图 2 所示。

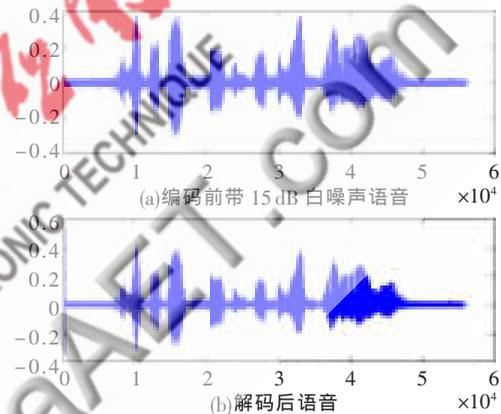


图 2 带 15 dB 白噪声的编解码图

由图 2 可以看到,该算法的降噪性能并不是很理想,加 15 dB 白噪声的语音基本上不可懂,但是解码后的语音虽然包含一定的噪声,而人耳基本上可以识别。该算法在降噪方面还有一定改进的余地,G.729.1 算法的改进还有很大的价值。

参考文献

- [1] ITU-T Rec.G.729.1. An 8-32 Kb/s scalable wideband coder bitstream interoperable with G.729. May. 2006.
- [2] 贾懋坤,鲍长春,李锐.8~64 Kb/s 超宽带嵌入式语音编码方法[J].通信学报,2009(5).
- [3] CHIVUKULA R K, REZNI Y A. Efficient implementation of a class of MDCT/IMDCT filterbanks for speech and audio coding applications[M]. ICASSP, 2008.
- [4] HEIDEMAN M T. Computation of an odd-length DCT from a real-valued DFT of the same length[J]. Signal Processing, IEEE Transactions, 1992,40(1):54-61.

(收稿日期:2010-05-06)

作者简介:

刘丽群,女,1984 年生,硕士研究生,主要研究方向:语音信号处理。