

波形变换域浊音语音的一种 分帧自适应编码算法*

陈云凤 伍康文

(中山大学无线电电子学系)

摘要 波形变换编码算法中,影响语音质量的主要因素是浊音语音编码时产生的量化误差。本文在Mozer算法基础上,提出一种波形变换域分帧自适应编码新算法。采用这种算法用近代的单片数字信号处理器对编译码容易做到实时处理,而短时量化信噪比的平均值可达6dB。

关键词 波形变换,自适应编码,语音信号

波形变换编码算法中,影响语音质量的主要因素是浊音语音编码时产生的量化误差。为了提高重建语音质量,Mozer提出寻找幅度群聚在16级电平附近的新波形进行四比特量化以提高量化信噪比的方法^[1,2]。这种编码算法的缺点是编码时对每一语音帧都要进行多次费时的正、逆傅里叶变换,运算量很大,即使在大型计算机上运行,对编码也无法实时处理。本文在Mozer算法基础上,提出一种波形变换域分帧自适应编码新算法。这种新算法每帧只需进行一次正、逆傅里叶变换,在保持与Mozer算法相同压缩比条件下,其短时量化信噪比的平均值可达26dB。重建语音质量接近于采样数据放出的语音。采用这种算法,用近代的单片数字信号处理器对编译码容易做到实时处理。这种语音数据压缩技术,不但可用于语音合成,而且可推广应用于语音通讯和语音识别这一类需要实时的场合。

1 语音信号的波形变换

人的听觉对声音的感知主要决定于语音信号的短时功率谱,而短时相位谱只起次要作用。可以对语音信号进行波形变换,找出有利于量化和编码的新时域波形代替原语音波形进行数据压缩。图1示出了本文采用的波形变换数据处理过程。语音信号 $S(n)$ 通过基音检测分出浊音段和清音段并检测出浊音段的基音周期,然后按浊音和清音两种类型

本文1992年5月18日收到

* 国家自然科学基金资助项目

划分为不同的分析帧 $S_m(n)$ ， m 表示不同的分析帧。为保持波形变换后浊音语音的准周期性，浊音帧的分析帧长度取等于基音周期。下一步是对各分析帧 $S_m(n)$ 通过离散傅里叶变换分别进行波形变换，本文中波形变换的策略是选取能量集中在基音周期中心附近、周期中心处有正最大样值且中心对称的某一新时域波形，相位的取值是^[3]

$$\left. \begin{aligned} S'_\varphi(k) &= \pi && \text{当 } k = 2l - 1, l = 1, 2, \dots, N/2 \\ S'_\varphi(k) &= 0 && \text{当 } k = 2l, l = 0, 1, \dots, (N/2) - 1 \end{aligned} \right\} \quad (1)$$

式中 N 为分析帧样点数。这种新时域波形具有如下共同特点：①具有中心对称性。因此只要知道了波形的前半周期，就可通过对称性重构后半周期部份；②样值在基音周期中心 $N/2$ 处有最大正值 $X(N/2)$ ，一般说来离基音周期中心越远，信号样值的幅度越来越小。

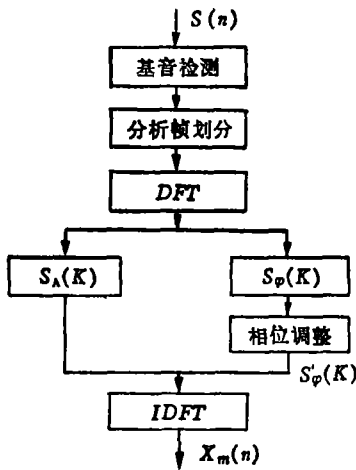


图1 波形变换数据处理过程
Fig.1 Data processing of the waveform transformation

2 波形变换域分帧自适应编码算法

图2示出了波形变换域一个浊音帧进行自适应编码和解码方案。图中表示不同分析帧的下标 m 已经省略。整个过程可分为3个步骤：

第一步，增益因子 G 的提取与正最大值 $X(N/2)$ 的编码和译码。

浊音语音帧与帧之间幅度有很大差异，这种差异可用一个三位二进制数来表示。这个数对应的8个状态表示8种不同幅度差异。增益 G 可由下式计算

$$G = \text{INT}(\log_2 \frac{X_{\max}}{X(N/2)}) \quad (2)$$

式中 X_{\max} 为量化器峰—峰值范围的一半，对12位量化器， $X_{\max} = 7\text{FFH}$ （H表示16进制数）， $X(N/2)$ 为新波形基音周期中心处的正最大值，实验时 G 可由 $X(N/2)$ 作简单的规格化左移求得。 G 值决定后，可先对规格化后的 $x(N/2)$ 进行编码和译码。注意到规格化后的 $x(N/2)$ 其最高2位必为01，编码时可以删去，只记录其随后4位作为代码 $C(N/2)$ ，译码时再在代码前加上2位01组成6位有效值并组成量化值 $\hat{x}(N/2)$ ，则正最大值 $X(N/2)$ 的重建值为

$$\hat{X}(N/2) = \hat{x}(N/2) \cdot 2^{-G} \quad (3)$$

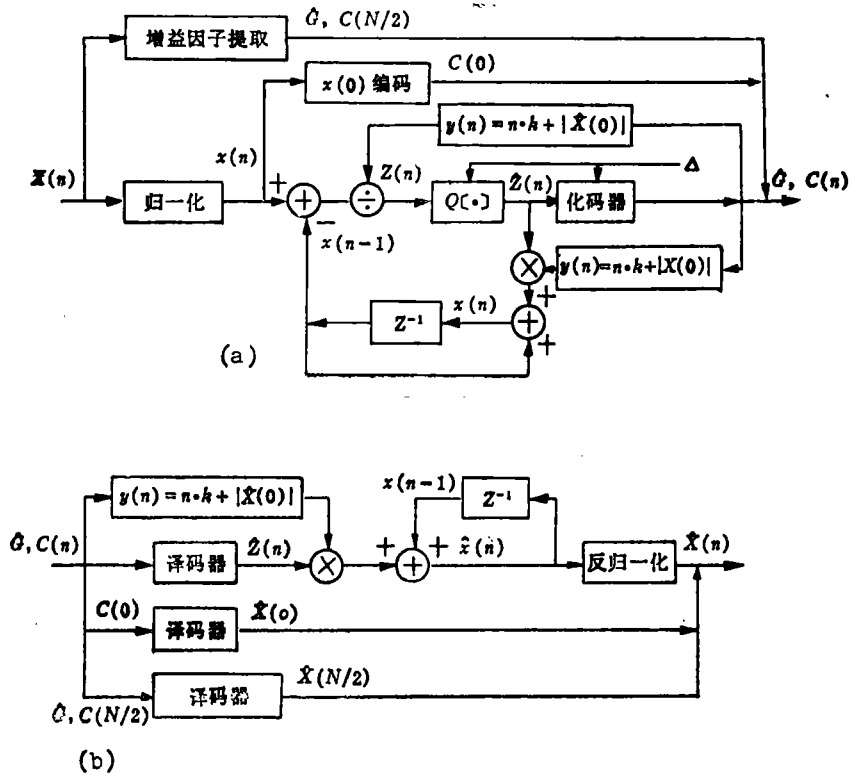


图2 分帧自适应编码(a)和解码(b)方案

Fig.2 Block diagram of short-time adaptive quantizing algorithm in voiced coding systems

第二步，对波形进行归一化处理与 $X(0)$ 值的编码和译码。

设 $x(n)$ 为归一化样值，则

$$x(n) = X(n)/\hat{X}(N/2) \quad n = 0, 1, \dots, (N/2) - 1 \quad (4)$$

以 $x(0)$ 的最高4位作为 $X(0)$ 的代码 $C(0)$ ，译码时由代码 $C(0)$ 求出量化值 $\hat{x}(0)$ 再乘以 $\hat{X}(N/2)$ 即得 $X(0)$ 的重建值 $\hat{X}(0)$ 。

第三步，对其余归一化样值的差值作自适应编码。

因为变换后的新波形，一般说来离基音周期中心越远，信号样值和样值的差值越来越小，这种变化趋势我们可以用常数 K 来近似描述

$$K = (0.5 - |\hat{x}(0)|) / (N/2 - 2) \quad (5)$$

本文采用固定阶距的反馈自适应系统，方法是使差值的幅值尽可能接近于1后才量化，为此先求出半个基音周期内差值的幅值随时间变化近似曲线

$$y(n) = n \cdot k + |\hat{x}(0)| \quad n = 1, 2, \dots, (N/2) - 1 \quad (6)$$

然后用差值除以 $y(n)$ 得 $Z(n)$

$$Z(n) = \frac{x(n) - \hat{x}(n-1)}{y(n)} \quad n = 1, 2, \dots, (N/2) - 1 \quad (7)$$

以 $Z(n)$ 的最高4位作为 $X(n)$ 的代码 $C(n)$ ，译码时由代码 $C(n)$ 求出量化值 $\hat{Z}(n)$ ，再由

下式求出归一化量化值 $\hat{x}(n)$

$$\hat{x}(n) = \hat{Z}(n) \cdot y(n) + \hat{x}(n-1) \quad n = 1, 2, \dots, (N/2) - 1 \quad (8)$$

$X(n)$ 的重建值 $\hat{X}(n)$ 可由下式求出

$$\hat{X}(n) = \hat{x}(n) \cdot \hat{X}(N/2) \quad n = 1, 2, \dots, (N/2) - 1 \quad (9)$$

3 性能测试

在配有TMS32020数字信号快速运算插板ATD-20A的IBM-PC/XT机上,实现了上述编码系统,用分段信噪比法对系统进行了性能测试^[4]。实验时 $X(n)$ 和 $\hat{X}(n)$ 作为系统输入和输出样值,分段长度取32ms,表1列出了汉语普通话6个常见单元音的测试结果。实验材料为彩翎音像有限公司制作,广东省音像出版社发行“六年制小学语文课本第一册”汉语拼音朗读中的男声,由录音机输入,采样频率8kHz,12位A/D采样,波形变换后编码系统的输入 $X(n)$ 亦保持12位。表中只列出每个发音的前10段信噪比及其平均值。

最后,将直接采样的数据与编码后重建语音数据分别送D/A转换器经滤波放大后由扬声器输出进行比较。试听结果表明,重建语音质量接近于采样数据放出的语音。

表1 6个单元音的分段信噪比SNRseg(单位: dB)

Tab.1 Short-time quantization SNR(dB) of the six vowels for mandarin Chinese

	1	2	3	4	5	6	7	8	6	10	平均
a	25.4	26.1	26.7	5.8	25.2	25.9	26.5	26.7	25.9	25.5	26.0
o	25.4	26.8	27.4	26.8	27.6	28.5	27.9	27.0	27.6	27.2	27.2
e	25.2	24.8	26.2	25.1	25.3	25.7	23.9	24.5	25.3	25.8	25.2
i	27.0	28.1	27.3	27.6	27.0	27.7	24.8	22.5	25.0	27.7	26.5
u	25.8	27.5	28.4	29.0	28.1	27.9	27.5	28.6	27.7	27.5	27.8
ü	27.0	27.7	27.1	26.2	24.9	23.6	24.3	22.6	22.8	23.3	24.9

4 讨论

(1) 本文提出的波形变换域分帧自适应编码算法在实际应用时并不需要对每一分析帧都进行波形变换和编码。因为这种算法保留了Mozor方法中“基音周期重复”等数据压缩技术^[1],它具有与Mozor算法同样的压缩比,利用“基音周期重复”我们可以隔几帧计算一次以加快计算速度和进行数据压缩。

(2) 由于实验条件限制,作者无法将两种方法所产生的重建语音进行比较。但与波形变换后直接用四比特量化的结果比较,可以看出,分帧自适应编码比直接用四比特量化的短时信噪比平均高9~10dB。

(3) 本文虽没有对清音的分帧自适应编码进行讨论, 但从原理上完全可推广应用于清音。实际上清音因幅度小, 音长短且幅度变化范围不像浊音那么宽, 因此, 在波形变换后, 清音的编码比浊音容易得多。

参 考 文 献

- 1 Bristow G. *Electronic Speech Synthesis*. Granada Publishing Ltd, 1984. 135~147
- 2 Weinrich D W. *Electronics*, 1980(10):113~118
- 3 陈云凤, 汪立全, 伍康文. 数据采集与处理, 1992, 7(2):9~15
- 4 Kitawaki N, Nagabuchi H. *IEEE Com Magazine*, 1984(10): 36~44

A Short-time Adaptive Quantizing Algorithm for Voice Waveform Based on Waveform Transformation

Chen Yunfeng Wu Kangwen*

Abstract We propose a new short-time adaptive quantizing algorithm on the basis of Mozer algorithm. It is shown that real time coding and decoding procedures can be easily realized in modem digital signal processor with a 26dB short time average quantized signal to noise ratio.

Keywords waveform tranformation, adaptive coding, speech signal

* Department of Radio and Electronics, Zhongshan University