

波形变换域语音波形的帧间差分编码

伍康文 陈云凤

(中山大学无线电电子学系, 广州 510275)

摘要 本文提出一种新的语音波形的帧间差分编码算法. 文中对帧间差分编码所遇到的帧匹配、过渡帧处理和量化器设计等问题进行了讨论. 文末给出了汉语普通话几个常见单元音量化信噪比的测试结果. 实验表明: 采用四比特帧间差分编码, 当信号采样频率为 8kHz 时, 其短时量化信噪比的平均值超过 29dB. 利用归一化新波形的对称性、半周期取零技术和基音重复, 预测误差用 4 比特量化时, 可在数码率 2k~4kbps 情况下得到优质重建语音.

关键词 波形变换, 差分编码, 帧间编码

分类号 TN921.31

语音信号的帧间相关性近年来逐渐受到重视, 利用线谱对(LSP)参量帧间和帧内同时存在的相关性, 对语音进行二维差分编码的研究已见报导^[1]. 本文讨论利用波形参量对语音信号进行帧间差分编码算法, 讨论帧间差分编码中遇到的帧匹配、过渡帧处理和量化器设计等问题, 文末给出了一些编码实验结果.

1 语音帧的波形变换预处理

利用波形参量对语音进行帧间差分编码, 首先应解决如何保证相邻语音帧之间波形有最大的相似性(对应样值有最大的相关性). 由于浊音的非周期性以及相对相位对语音波形的影响, 对浊音即使作基音同步分帧, 相邻帧之间的波形仍有很大差别. 文[2]指出: 根据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这一原理, 改变这些短时相位, 对原始波形进行变换, 可以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 示出了用文[3]方法求出的汉语普通话女声

【O】经波形变换预处理后某邻近 4 帧归一化新波形. 该变换在分析帧划分时, 分析帧数已设定为偶数, 新波形有如下特点: ① 每帧具有中心对称性, 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分; ② 样值在基音周期中心 1/2 处具有最大正值 1, 一般说来离基音周期中心越远, 信号能量越来越小. 从图可见, 这种新波形各基音帧的形状极为相近, 相邻帧对应样值之间有明显的相关性.

收稿日期: 1993-12-03

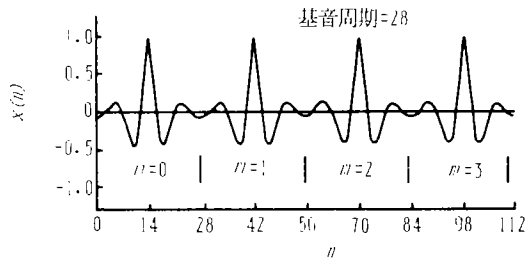


图 1 经波形变换和归一化预处理后的一段浊音语音波形

Fig. 1 A voiced waveform by waveform transformation and normalization

2 帧间差分编码算法

2.1 帧间预测和预测误差

$$\text{设 } X_m(n) = [x_m(n)], \quad n = 0, 1, \dots, N/2 \quad (1)$$

表示第 m 帧语音经波形变换后归一化新波形 \mathcal{D} . 将 $X_m(n)$ 看作帧轴 m 和时轴 n 的二维信号, 根据帧内和帧间各自存在的相关性, 可用线性预测去掉参量的多余度.

假定先前帧量化样值为 $\hat{X}_{m-1}(n)$, 本帧先前量化样值为 $\hat{X}_m(n-1)$, 如用一阶预测, 则预测值:

$$\text{帧内预测 } \tilde{X}_m(n) = a\hat{X}_m(n-1)$$

$$\text{帧间预测 } \tilde{X}_m(n) = b\hat{X}_{m-1}(n)$$

$$\text{二维预测 } \tilde{X}_m(n) = a'\hat{X}_m(n-1) + b'\hat{X}_{m-1}(n) \quad n = 0, 1, 2, \dots, N/2 \quad (2)$$

这里只考虑帧间预测并设预测系数 $b=1$. 帧间预测误差为

$$d_m(n) = X_m(n) - \hat{X}_{m-1}(n) \quad n = 0, 1, 2, \dots, N/2 \quad (3)$$

本文讨论的帧间差分编码, 即讨论预测误差的量化编码.

2.2 帧匹配问题

一般情况下, 浊音的基音周期随声调而变化. 当基音周期变化时, 相邻帧之间找不到对应样值进行差分, 这种情况下必须进行时间规正, 我们叫帧匹配. 根据新波形特点, 可用下述两种方法进行简单的帧匹配运算: ① 补零匹配; ② 帧内差分匹配. 即以每帧中心为基点, 当前帧样值差分对象为先前帧等中心距的量化样值, 如当前帧帧长大于先前帧时, 先前帧两端补零扩宽或将失去差分对象样值与本帧内先前量化样值进行差分.

设先前帧帧长为 $N(m-1)$, 当前帧帧长为 $N(m)$ 且 $N(m) > N(m-1)$, 则:

$$\text{当 } n \leq N(m-1)/2 \text{ 时, } d_m(N(m)/2-n) = X_m(N(m)/2-n) - \hat{X}_{m-1}(N(m-1)/2-n)$$

$$\text{当 } N(m-1)/2 < n \leq N(m)/2 \text{ 时, } d_m(N(m)/2-n) = X_m(N(m)/2-n) - 0$$

$$\text{或 } d_m(N(m)/2-n) = X_m(N(m)/2-n) - \hat{X}_m(N(m)/2-n+1) \quad (4)$$

1) 为讨论方便, 我们先假定 N 为常数. 同时因波形具有中心对称性, 因此可只考虑前半帧

由于元音的基音周期是连续变化的,相邻帧之间基音周期差别很小,又由于归一化新波形能量集中在基音周期中心附近,基音周期两端样值相对较小,可以预料这两种帧匹配对量化信号噪比的影响是很小的.

2.3 过渡帧处理

声母和韵母相连相拼部份存在一个过渡区,在过渡区语音信号特征如短时功率谱等变化迅速,其变换后归一化新波形相邻帧之间差别很大,显然不适合采用帧间编码,此外一个音节的持续段上,由于说话人习惯或呼吸等关系,也会在某个地方出现小的突变,这种突变也会影响帧间编码效果,对这些语音段我们统称为过渡音.

对过渡音我们采用帧内/帧间切换方式处理.由于过渡音没有明显的边界,划分过渡音只能订立一些工程性的切换条件.比较适宜的切换条件是:当帧间编码效果比不上帧内编码效果时,采用帧内差分编码方式.实际操作时在发送端每一帧同时作帧内和帧间差分编码,然后通过量化信噪比来确定两种编码方式的切换,量化信噪比按下式计算

$$\text{SNR}_{\text{seg}} = 10 \lg \left\{ \frac{[x(n) - \overline{x(n)}]^2}{[\hat{x}(n) - x(n)]^2} \right\} \quad (5)$$

在接收端(合成端)可根据帧内/帧间方式开关量直接决定每一帧的编码方式进行解码.

2.4 量化器设计

本文拟用固定阶距的均匀量化器.对均匀量化器,量化阶距 Δ 、电平数目 2^B (B 为比特数)和样值峰值 D_{max} ,必须满足下式

$$2|D_{\text{max}}| = 2^B \Delta \quad (6)$$

为估算 D_{max} 大小,作者对某一男声和女声全部单元音 4 个音调的归一化波形,计算下列差值

$$D_m(n) = X_m(n) - X_{m-1}(n), \quad n = 0, 1, 2, \dots, N/2 \quad (7)$$

并作样值 $D_m(n)$ 的统计分布直方图,统计结果,样值 $D_m(n) < 2^B$ 的概率大于 93.5%,即如果取 $D_{\text{max}} = 2^{-2}$,出现过载的概率不超过 6.5%,考虑到某些帧出现严重过载时,将被划入过渡帧, D_{max} 还可以取小一些.基于上述考虑,本文量化阶距 Δ 取为

$$\Delta = [2(2^B + 1)]^{-1} \quad (8)$$

3 性能测试

在配有 TMS32020 数字信号快速运算插板 ATD-20A 的 IBM-PC/AT 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.实验材料由录音机输入,采样频率为 8kHz,12 位 A/D 采样.实验中采用了补零帧匹配技术,预测误差用 4 比特量化,起始帧和过渡帧采用文献^[4]的帧内 4 比特 ADPCM 算法.表 1 列出了男、女声 6 个常见单元音 4 个音调测试结果.其短时量化信噪比总平均达 29.4.过渡帧占总帧数的 7.5%,即帧间编码帧占 92.5%.

表 1 6 个单元音的分段信噪比平均值(单位: dB)

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

声调	男第一声	男第二声	男第三声	男第四声	女第一声	女第二声	女第三声	女第四声
a	29.6	27.1	28.5	25.2	25.9	28.8	28.4	26.1
o	29.6	28.2	28.0	27.2	31.1	28.8	28.4	29.3
e	28.1	27.0	28.2	25.4	28.5	28.4	28.8	27.1
i	29.4	30.9	28.9	28.1	34.2	33.8	31.1	32.5
u	33.5	30.7	29.6	30.4	34.0	33.3	30.9	31.9
ü	28.0	29.5	27.8	27.1	31.9	33.4	31.1	31.7

总平均 SNR=29.4dB

4 结束语

(1) 本文在讨论帧间差分编码时,没有提及清音的编码.清音因能量小,不存在周期性,清音的编码不用考虑帧匹配技术,但上述算法仍然适用.关于清音的波形变换预处理技术可参考文献^[5].

(2) 本文提及的帧间差分编码基于波形变换预处理,波形变换编码中的数据压缩技术仍可应用.利用归一化新波形的对称性,半周期取零技术和基音重复,当信号采样频率为 8kHz,预测误差用 4 比特量化时,可在数码率 2k~4kbps 情况下得到优质重建语音.

(3) 本文只利用了帧间的相关性进行帧间差分编码,虽然帧间差分信号的帧间相关性已经很小,但帧内相关性仍很大.如果利用帧内和帧间同时存在的相关性,采用二维差分编码,必将进一步提高编码效率.

参 考 文 献

- 1 Kuo Chih chung, Jean Furong, Wang Hisao chuan. Low bit-rate quantization of LSP parameters using two-dimensional differential coding. ICASSP-92(1), S1:1-97~1-100
- 2 Bristow G. Electronic speech synthesis. Granada Publishing Ltd, 1984:135~147
- 3 陈云凤,伍康文.波形变换域浊音语音的一种分帧自适应编码算法.中山大学学报(自然科学版), 1993,32(4):51~55
- 4 陈云凤,伍康文.一种基于波形变换域的分段 ADPCM 语音编码算法.数据采集与处理,1992,7(3):168~175
- 5 陈云凤,汪立全,伍康文.基于波形变换编码的一种语音合成器.数据采集与处理,1992,7(2):86~93

Intra-frame Differential Coding for Speech Waveform in Time Domain

Wu Kangwen* Chen Yunfeng

Abstract We propose a new intra-frame differential coding algorithm for speech waveform in time domain. This coding algorithm treats the waveforms in pre-proceeding as two-dimension signals. X-axis signal is the sequence number of samples in the same frame and Y-axis is the sequence number of frames, which performs the matching differential coding for the two sequent frames. Problems caused by intra-frame differential coding, such as frame matching, transition frame proceeding and quantizer design, are discussed in detail. SNR test results for the some vowels of Mandarin Chinese is presented. The experiment shows that the short-time quantization SNR of 4-bits intra-frame differential coding surpasses 29 dB at 8kHz sampling rate. It is shown that there is a possibility to synthesis excellent speech with bit rate of 2~4kbps.

Keywords waveform transformation, differential coding, intra-frame coding

* Department of Radio and Electronics, Zhongshan University, Guangzhou 510275