

低延迟话音编码算法的增益优化与混合窗设计

张刚 谢克明 张雪英 皇甫丽英
(太原理工大学信息工程学院 太原 030024)

摘要: G.728标准将增益滤波器的偏移不适当地处理成常数32dB, 相当于将20~30ms内短时平稳的语音信号不适当地按整个时间轴上的平稳信号对待。对增益滤波器的改进必然涉及优化混合窗。该文提出一种估计平均分段信噪比的方法用来设计混合窗参数, 并导出激励增益的精确表示。以此为基础, 将Jayant的自适应量化与G.728的增益预测结合起来, 提出一种适合于设计低时延码激励线性预测(LD-CELP)的增益码书的方案。用此方法, 激励的精确值经自适应预测和量化后, 3bit自适应量化和4bit自适应量化分别比G.728的固定量化有0.5dB和6dB的信噪比的改善。当采用4bit自适应量化和64码字的波形码书, 比G.728复杂性降低20%, 同时信噪比提高1dB。

关键词: 语音处理, G.728, LD-CELP, 增益滤波, 混合窗

中图分类号: TN912.3 **文献标识码:** A **文章编号:** 1009-5896(2005)06-0892-04

Optimizing Gain and Designing Hybrid Window of LD-CELP

Zhang Gang Xie Ke-ming Zhang Xue-ying Huangfu Li-ying

(College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China)

Abstract The offset of gain filter is dealt unsuitably with a constant 32dB in G.728 standard. That is, the short-term stationary speech signal during the duration 20~30ms is unsuitably supposed as stationary at whole time axis. It is necessary to design optimizing hybrid window for improving gain filter. In this paper, the exact expression of excitation gain is derived and an estimation method of SNR_{av} used to optimize the parameters of hybrid window is presented. On the basis of above, this paper combines the Jayant adaptive quantization and G.728's gain prediction to propose a scheme fit for designing gain codebook of Low Delay Code Excited Linear Prediction(LD-CELP), in which the exact gain value is adaptively predicted and quantized. In this method, for adaptive quantization gain of 3 bit and 4 bit, the SNR increases by 0.5 and 6 dB separately than that of G.728. If adopting 4 bit adaptive quantization and the shape codebook of 64 codevectors, the G.728' complexity can reduce by 20% while SNR has 1 dB improvement.

Key words Speech processing, G.728, LD-CELP, Gain filter, Hybrid window

1 引言

G.728的增益滤波器利用AR模型估计激励增益 $\sigma_e(n)$ 。按照随机过程理论, 序列 $\{\sigma_e(n)\}$ 应该是平稳的随机序列。也就是说, $\{\sigma_e(n)\}$ 的均值应该是常数。试验表明并非如此。图1给出了一段激励增益的波形, 横轴为时间, 以增益样点数为单位(每个增益样点是对应5个信号采样); 纵轴是增益幅度的对数值(dB)。从图1看出, $\{\sigma_e(n)\}$ 的均值在0~45dB之间振荡。如果去除偏移后增益是平稳的, 则可改善G.728增益滤波器性能。下节介绍本文提出的设计混合窗的一般方法, 将其应用于综合滤波器和感觉加权滤波器也获得了好的结

果。自适应增益预测和量化分别在第3节和第4节介绍, 第5节是试验结果。

2 滤波器的混合窗

2.1 G.728的混合窗

图2显示G.728算法的混合窗^[1,2]。每L个样点进行一次LPC分析, 当前帧信号是 $s_u(m), s_u(m+1), \dots, s_u(m+L-1)$ 。

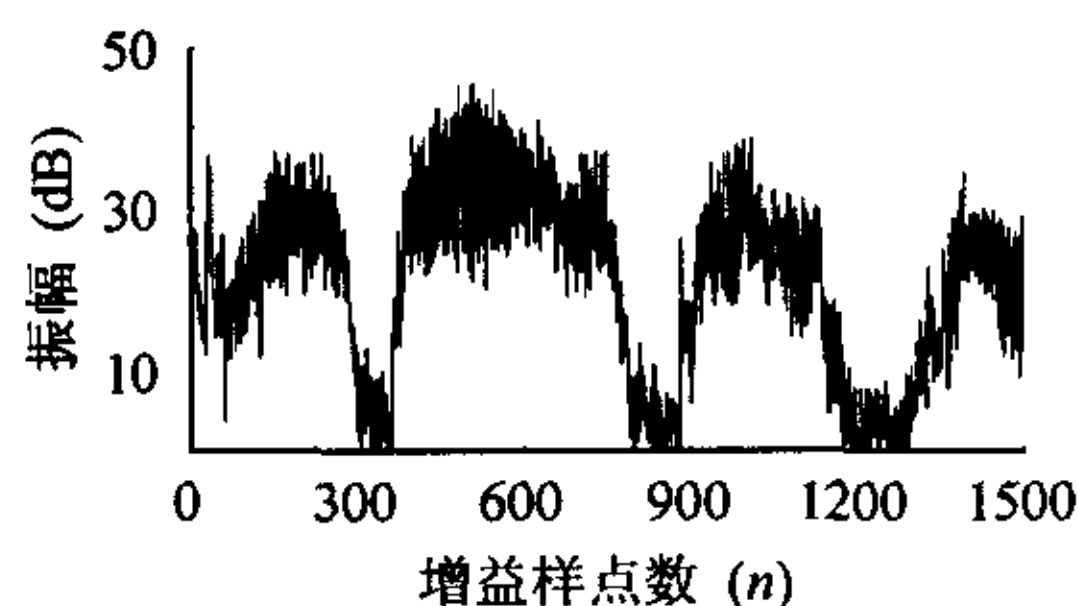


图1 增益波形

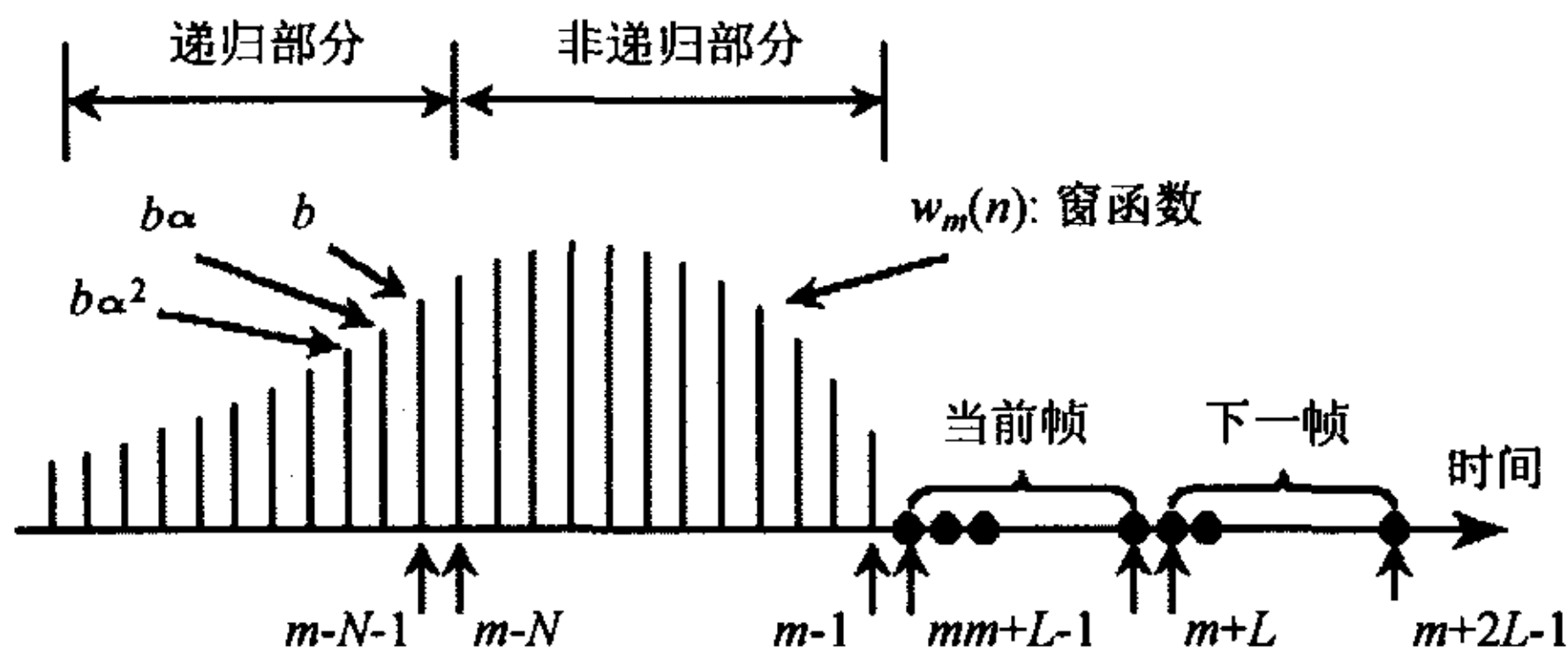


图2 混合窗

混合窗分成两部分:长度为 N 样点的正弦波形是非递归部分,另取指数波形为递归部分。样点 $s_u(m-1), s_u(m-2), \dots, s_u(m-N)$ 由非递归部分加权。从 $s_u(m-N-1)$ 开始,其左边的全部样点由递归部分加权。加权值为 $b, b\alpha, b\alpha^2, \dots$,这里, $0 < b < 1$ 和 $0 < \alpha < 1$ 。在时刻 m ,混合窗函数 $w_m(k)$ 为

$$w_m(k) = \begin{cases} f_m(k) = b\alpha^{-[k-(m-N-1)]}, & k \leq m-N-1 \\ g_m(k) = -\sin[c(k-m)], & m-N \leq k \leq m-1 \\ 0, & k \geq m \end{cases} \quad (1)$$

为有效地估计信号 $\{\sigma_e(n)\}$,加窗后可表示为

$$s_m(k) = s_u(k)w_m(k) = \begin{cases} s_u(k)f_m(k) = s_u(k)b\alpha^{-[k-(m-N-1)]}, & k \leq m-N-1 \\ s_u(k)g_m(k) = -s_u(k)\sin[c(k-m)], & m-N \leq k \leq m-1 \\ 0, & k \geq m \end{cases} \quad (2)$$

M 阶LPC分析用到参数 M, L, N, b, c 和 α 。这些参数的不同赋值便可得到不同的混合窗。

2.2 选择窗函数

由式(1)和图2,混合窗非递归部分形状由其第1个窗函数数值确定。这里

$$w_0 = -\sin[c(m-1-m)] = \sin(c), \text{ 或者 } c = \arcsin(w_0) \quad (3)$$

为了平滑联接混合窗的递归与非递归两个部分。递归部分的第1个值应为

$$b = w_m(m-N-1) = \sin[c(N+1)] \quad (4)$$

因为混合窗的递归部分单调递减, w_0 一定与某个 $w_m(k)$ 最接近,不妨假定 $w_m(k) = w_0$ 。于是,

$$b\alpha^{-[k-(m-N-1)]} = w_0 \quad (5)$$

$$\alpha = \sqrt[k-(m-N-1)]{w_0 / \sin[c(N+1)]} \quad (6)$$

由上可知,只有 w_0 与 k 是决定性参数而 c, b 和 α 可分别由其导出。这里 L 和 N 与G.728相同,而平均分段信噪比用作选择参数 w_0 与 k 的判据。实验表明,优化后的窗可以使滤波器的系数更好地适应语音信号的时变性。

3 自适应增益预测

3.1 偏移自适应

令 $\beta_e(n) = \log[\sigma_e(n)]$ 表示第 n 个被增益调节的激励矢量

的均方根的常用对数,而 $g(n)$ 是 $\beta_e(n)$ 的均值(偏移),即有

$$\beta_e(n) = g(n) + \sigma'_e(n) \quad (7)$$

因序列 $\{\beta_e(n)\}$ 是短时平稳的。可以期望去除偏移后,序列 $\{\sigma'_e(n)\}$ 是平稳的。这里

$$\sigma'_e(n) = \beta_e(n) - g(n) = \log[\sigma_e(n)] - g(n) \quad (8)$$

从而 $\sigma'_e(n)$ 可以用线性预测滤波器估计,而 $g(n)$ (序列 $\{\beta_e(n)\}$ 的均值)由下式进行估计:

$$g(n) = r \cdot \beta_e(n-1) + (1-r) \cdot g(n-1) \quad (9)$$

这里, $r \in (0,1)$ 是一个自适应参数。该模型实际上是一个有低通特性的一阶递归滤波器。这个滤波器可以捕捉到激励增益振荡成分中缓慢变化的部分,从而使激励增益滤波器的偏移随输入信号的变化缓慢地变化,而不是一个常数。新的自适应激励增益预测器成为

$$\log \sigma(n) = \sum_{i=1}^p p_i [\log[\sigma_e(n-i)] - g(n-i)] + g(n) \quad (10)$$

3.2 激励增益的精确表示

G.728根据公式(11)进行码书搜索:

$$D_{\min} = \sigma^2(n) \|\hat{x}(n) - g_i H(n) y_j\|^2 \quad (11)$$

这里 (n) 是激励增益的估计值。 $\hat{x}(n)$ 是被 (n) 调节的目标矢量, $H(n)$ 是预测器的单位冲激响应, g_i 是增益码字, y_j 是波形码矢。由于式(11)的最小化等价于下式的最大化

$$\hat{D}_{\max} = 2g_i P^T(n) y_j - g_i^2 E_j \quad (12)$$

这里 $P(n) = H^T \hat{x}(n)$, $E_j = \|H y_j\|^2$ 。在式(12)中,令 $\partial \hat{D}_{\max} / \partial g_j = 0$ 有

$$g_j = [P^T(n) y_j] / E_j \quad (13)$$

这就是增益码字的精确值。为表达简单,令 y_j 是归一化波形码书,则式(11)可以写成

$$D_{\min} = \|\mathbf{x}(n) - G_j(n) H(n) y_j(n)\|^2 \quad (14)$$

这里 $\mathbf{x}(n) = \sigma(n) \hat{x}(n)$ 和 $G_j(n) = \sigma(n) g_j(n)$ 分别是未经增益调节的目标矢量和未经增益调节的激励增益的精确值。令 $\hat{G}_j(n)$ 表示 $G_j(n)$ 的量化值^[3],由G.728协议, $\hat{G}_j(n)$ 是式(10)中的 $\sigma_e(n)$,令 $g_j(n) = G_j(n) / \sigma(n)$,则对数域中的增益预测残差是

$$\log_2 g_j(n) = \log_2 G_j(n) - \log_2 \sigma(n) \quad (15)$$

用 $Q\{\cdot\}$ 表示信号 $\{\cdot\}$ 的量化值。则

$$\log_2 \hat{G}_j(n) = Q\{\log_2 g_j(n)\} + \log_2 \sigma(n) \quad (16)$$

3.3 窗的选择和增益滤波器定阶

当用式(10)选择增益滤波器的窗和阶时,量化信号 $\hat{G}_j(n)$ 还不存在。必须用真值 $G_j(n)$ 代替,这样平均分段信噪

比SNR_{av}一直处于最佳状态。用SNR_{av}无法判定窗和阶选择得是否适合于增益预测器,但是可以对SNR_{av}进行估计。令 $\varepsilon_j(n) = G_j(n) - \sigma(n)$ 是预测残差,则有

$$g_j(n) = G_j(n) / \sigma(n) = G_j(n) / [G_j(n) - \varepsilon_j(n)] = \text{snr}_j(n) / [\text{snr}_j(n) - 1] \quad (17)$$

这里 $\text{snr}_j(n) = G_j(n) / \varepsilon_j(n)$ 是n时刻信号 $G_j(n)$ 与残差 $\varepsilon_j(n)$ 之比,从而有

$$\text{snr}_j(n) = g_j(n) / [g_j(n) - 1] \quad (18)$$

由式(18)可知, $g_j(n)$ 越接近于1, $\text{snr}_j(n)$ 对信噪比的贡献越大。取 $\text{snr}_j(n)$ 的一个观测值 snr , 可以对所有的j计算满足 $\text{snr}_j(n) \geq \text{snr}$ 的个数。假定 Δ 是SNR_{av} 的一个预先设定的水平, 即有

$$\text{SNR} = 10 \lg N^{-1} \sum_{k=1}^N \text{snr}_j^2(k) \geq 20 \lg |\text{snr}| \geq \Delta \quad (19)$$

或

$$10^{\Delta/20} \leq |\text{snr}| = (1 - g_j(n)^{-1})^{-1} \quad (20)$$

由此得到优化方法: 对于两种不同的解决方案, 分别用式(10)计算各自的激励增益预测值 $\sigma(n)$ 。此时 $\sigma_e(n)$ (或 $\hat{G}_j(n)$) 用 $G_j(n)$ 代替并且有 $g_j(n) = G_j(n) / \sigma(n)$ 。选择SNR_{av} 的水平 Δ 并考查相应的范围:

$$(1 + 10^{-\Delta/20})^{-1} \leq g_j(n) \leq (1 - 10^{-\Delta/20})^{-1} \quad (21)$$

计算两个方案中满足式(21)的 $g_j(n)$ 比例, 比例较大的方案更好。

4 增益自适应量化(GAQ)

4.1 原理

图3是GAQ块图^[5,6]。量化 $\log_2 g_j(n)$ 获得4bit索引 $I(n)$ 送往解码端, $I(n)$ 同时在本地解码得到量化差值信号 $\log_2 \hat{g}_j(n)$, 与增益预测值 $\log_2 \sigma(n)$ 相加后, 得到本地重建信号 $\log_2 \hat{G}_j(n)$ 。预测器是12阶的AR模型, 量化阶距 $\Delta(n)$ 由输入信号自适应控制。通过选择一组最佳量化器参数可以得到 $\log_2 g_j(n)$ 的最佳量化效果。图4是 $\log_2 g_j(n)$ 的分布 $p(n)$ 。不计符号位, $\log_2 \hat{g}_j(n)$ 是8个量化电平 $\eta_i (i=0, \dots, 7)$ 之一, 量化方差为

$$\sigma_e^2 = \sum_{i=0}^7 E[(\eta_i - \log_2 g_j(n))^2] = \sum_{i=0}^7 \int_{\xi_i}^{\xi_{i+1}} (\eta_i - x)^2 p(x) dx \quad (22)$$

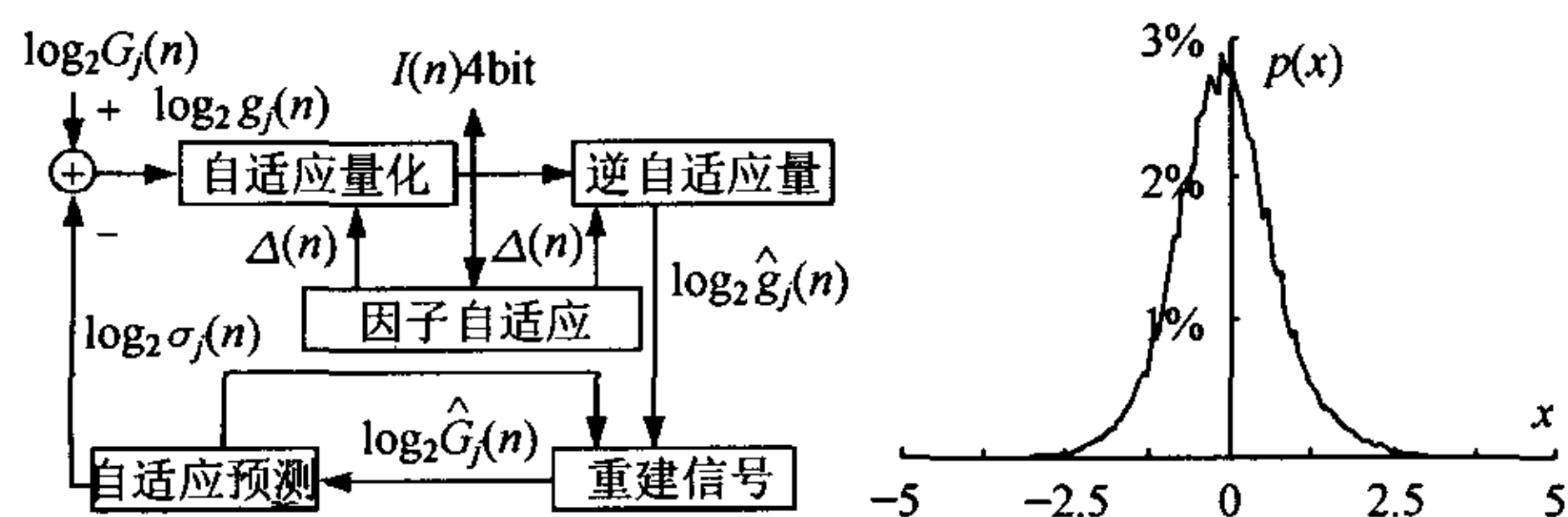


图3 编码器块图

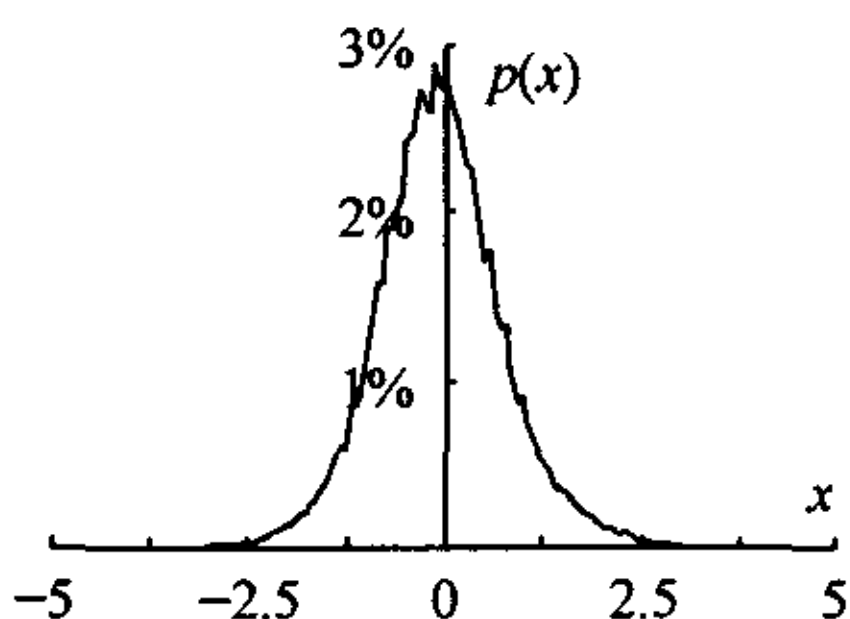


图4 增益的概率分布

选择参数 ξ_i 和 η_i 使 σ_e^2 最小化, 这里 $\xi_0 = -\infty, \xi_8 = +\infty$ 。令 $\partial \sigma_e^2 / \partial \xi_i = \partial \sigma_e^2 / \partial \eta_i = 0$ 可得

$$\eta_i \left[\int_{\xi_i}^{\xi_{i+1}} x p(x) dx \right] / \left[\int_{\xi_i}^{\xi_{i+1}} p(x) dx \right] \quad (23)$$

$$\xi_i = (\eta_{i-1} + \eta_i) / 2 \quad (24)$$

式(23)说明 η_i 的最优位置处于 ξ_i 和 ξ_{i+1} 的距心, 而式(24)说明 ξ_i 的最佳位置是 η_i 和 η_{i+1} 的算术平均。参数 ξ_i 和 $\eta_i (i=0, \dots, 7)$ ^[4] 迭代计算的方法如下: 取步长 Δ 划分区间 $[\eta_i, \eta_{i+1}]$, 对每个 Δ , 计算其中 $\log_2 g_j(n)$ 的样点个数 f_i , 除以总样点数 F 得到 Δ 内的样点频数 f_i / F , 用它近似代替式(23)的分布函数 $p(n)$ 。取 Δ 的中值为 x , 当 Δ 足够小时, $\log_2 g_j(n)$ 可以看作均匀分布。给出 ξ_i, η_i 的初值后可以得到分布函数 $p(n)$, 然后可由式(23), (24)计算新的 ξ_i, η_i , 如此迭代直到其值稳定。

4.2 最佳修正因子

量化阶距的自适应鲁棒乘子算法为: $\Delta(n) = M[I(n-1)] \Delta^\beta(n-1)$, 这里 $0 < \beta < 1$ 是提高量化器鲁棒性的衰减因子。步长修正因子 M 是前一个索引 $I(n-1)$ 的函数。信号的概率密度越小, 则 M 越大。为使信号能量不变时系统具有尽可能高的平稳的SNR, 修正因子应满足 $\prod_{i=0}^7 M_i^{p_i} = 1$, 这里 $p_i (i=0, \dots, 7)$ 是帧 i 的占有概率。 M_i 根据量化增益的平均分段SNR来选择, 而离散参数 $W[I(k)]$ 与 $M[I(k)]$ 的关系是 $W[I(k-1)] = 2^5 \log_2 W[I(k-1)]$ 。

5 试验结果与讨论

首先用40句话(20句男声和20句女声, 时长共计78s)估计式(9)的参数 γ , 然后用100句话(50句男声和50句女声, 时长共计160s)测试自适应偏移算法。图5是用式(9)估计的均值序列 $\{g(n)\}$ 的波形, 它与图1的波形类似。图6是去除偏移后的增益波形, 其特征正是我们期望的。

5.1 增益自适应量化

选择5种增益量化方案进行比较, 图7是其量化信噪比结果。横轴是测试语句的序列号, 纵轴是以分贝表示的量化信噪比。可以看出对同样的量化比特, 自适应量化比固定

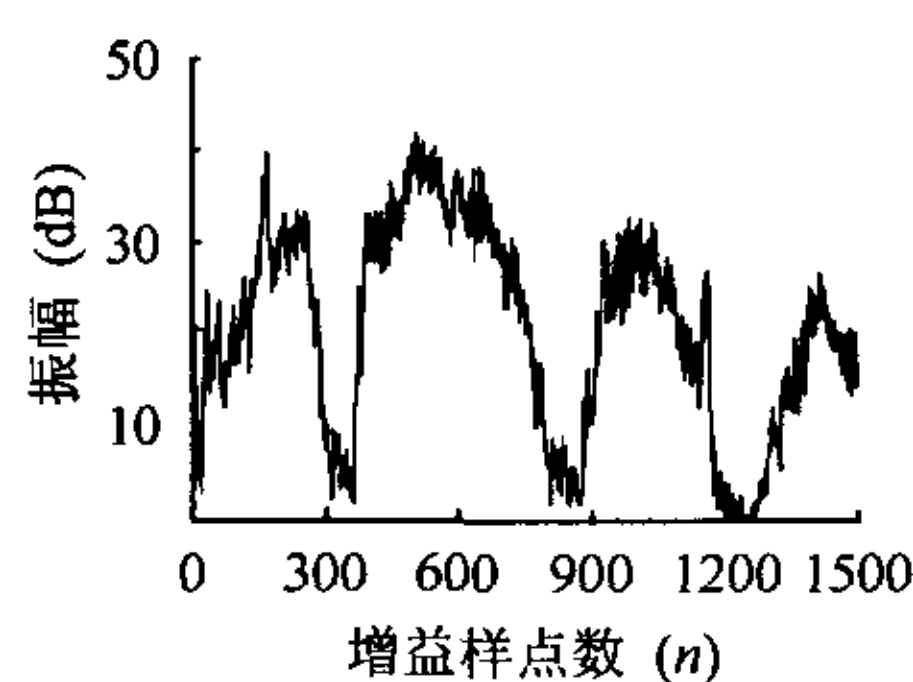


图5 均值序列 $g(n)$ 的波形曲线

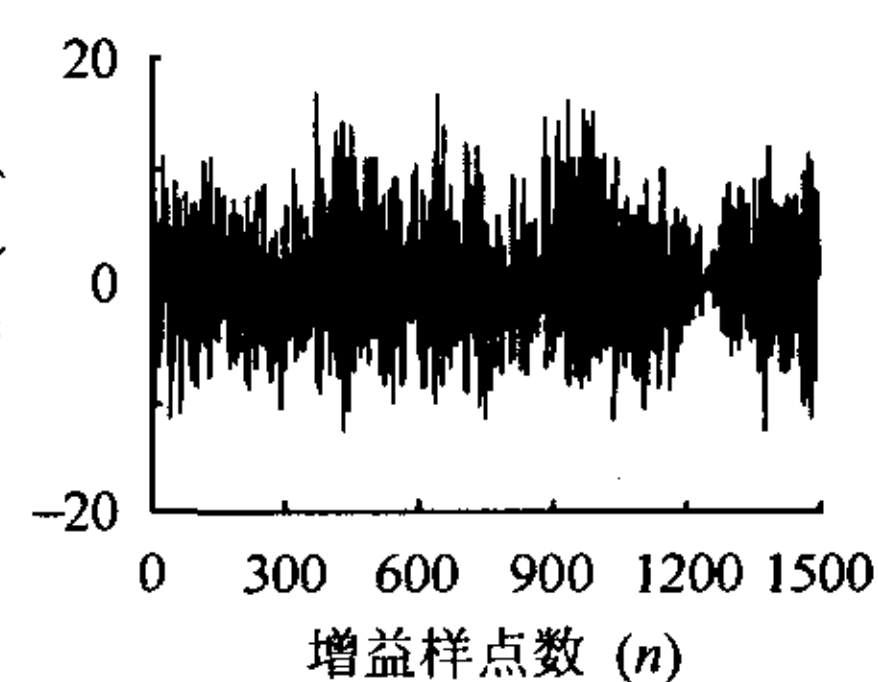


图6 去偏后增益样点n的增益波形曲线

量化高出2dB左右。而3bit和4bit自适应量化分别比G.728固定量化高出0.5dB和6dB。对 $g_i(n)$ 4bit量化, 离线时信噪比为17.7dB。在线时达19.45dB。

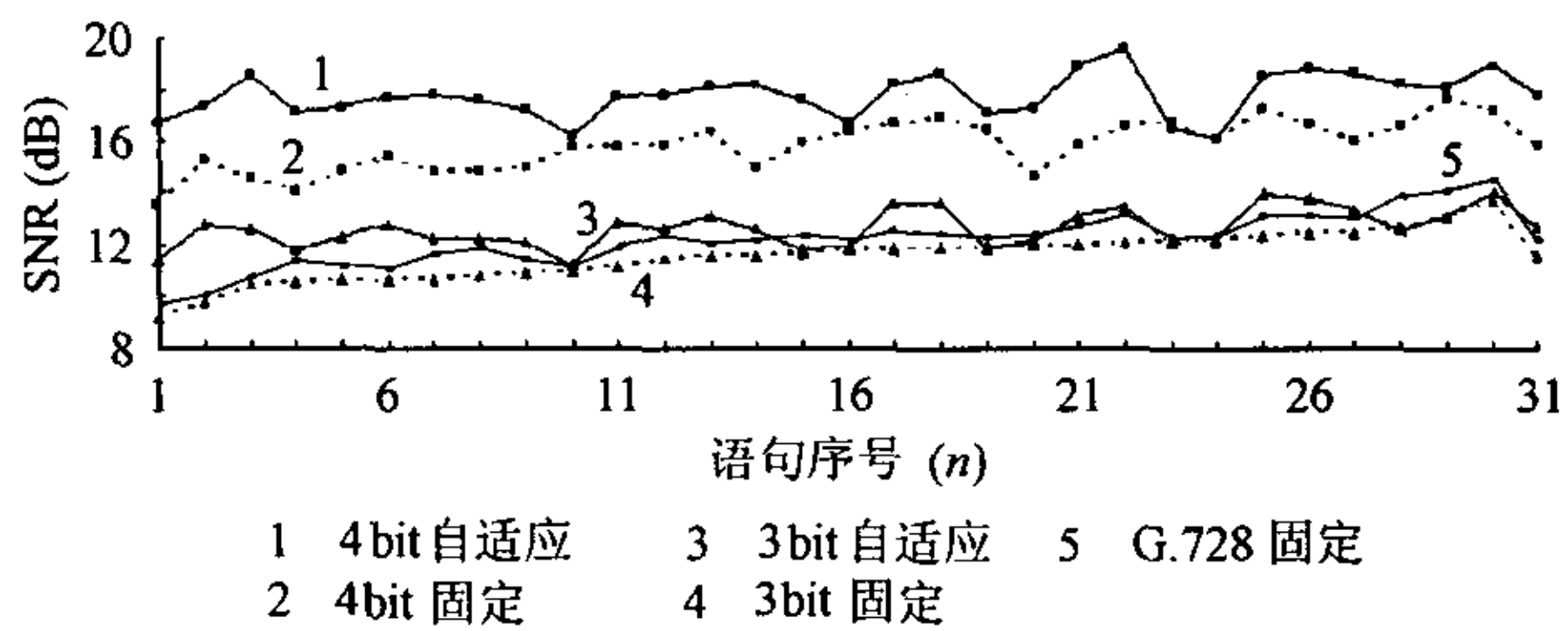


图7 5种量化方案的比较

5.2 语音编码

通过训练30句话61000个矢量, 从G.728的128个归一化波形码字中, 选出使用频率最高的64个作新算法的波形码书。码书搜索期间, 对每个输入语音矢量, 首先计算目标矢量, 然后用式(4)对64个归一化波形码字计算64个增益 $g_j(n)$ 的精确值, 并分别对其量化。相对于 \hat{D}_{\max} 的一对 $\hat{g}_j(n)$ 和 y_j 是最佳增益量化值和最佳波形码矢。经30句话测试, 其SNR是22.16dB, 比G.728高出1dB, 另外选用54句话进行测试, 也得到相同结果。

参考文献

[1] CCITT. Recommendation G.728, Coding of speech at 16kbit/s

using low-delay code excited linear prediction[S]. Geneva, 1992-9.

[2] CHEN J H. A robust low-delay CELP speech coder at 16 kb/s [D/B]. Advances in Speech Coding, 1991: 30 – 32.

[3] Zhang G, Xie K M, Zhang X Y, Huangfu L Y. Optimizing gain codebook of LD-CELP[C]. Proceeding of ICASSP2003, Hong Kong, 2: 149 – 152.

[4] 皇甫丽英, 张刚, 等. 改进G.728语音编码算法的增益滤波[J]. 太原理工大学学报, 2002, 33(3): 232 – 233.

[5] Jayant N S. Adaptive quantization with a one-word memory[J]. Bell Syst. Tech. J., 1973, 52: 1119 – 1144.

[6] Cumiskey, Jayant, et al.. Adaptive quantization in differential PCM coding of speech[J]. Bell Syst. Tech. J., 1973, 52: 1105 – 1118.

张刚: 男, 1953年生, 教授, 研究方向为信号处理与嵌入式系统。

谢克明: 男, 1944年生, 教授, 研究方向为智能信息处理与智能控制。

张雪英: 女, 1964年生, 教授, 研究方向为信号处理与语音识别

皇甫丽英: 女, 1977年生, 讲师, 研究方向为语音编码。