低延迟码激励语音编码算法的最佳增益滤波器

张起贵　　张 刚
(太原理工大学信息工程学院　太原　030024)

摘　要　该文采用加权 L-S 算法、有限记忆算法以及 BP 神经网络算法分别与 G.728 标准使用的 Levinson_Durbin(L-D)方法进行 5 样点激励增益滤波方案比较测试，发现编码效果均好于 G.728。其中加权 L-S 方法语音编码效果最好, 其平均分段 SNR 高出 G.728 算法 0.76dB。用该方法评价了 16 样点激励矢量增益滤波器和 20 样点激励矢量增益滤波器，加权 L-S 方法同样效果最佳。

关键词　增益滤波器，加权 L-S，有限记忆，G.728，信噪比估计
中图分类号：TN912.3　　**文献标识码**：A　　**文章编号**：1009-5896(2006)08-1533-04

The Optimal Gain Filter of LD-CELP

Zhang Qi-gui　　Zhang Gang
(*College of Information Engineering*, *Taiyuan University of Technology*, *Taiyuan* 030024, *China*)

Abstract　The recommendation G.728 depends on the Levinson-Durbin algorithm to update gain filter coefficients. In this topic, it is replaced by three different methods which are the weighted L-S recursive filter, the finite memory recursive filter and the BP neural network, respectively. Using these three gain filter the speech coding effect is all better than the G.728. The weighted L-S algorithm has the best result. Its average segment SNR is higher than the G.728 about 0.76dB. It is also used to evaluate the case that excitation vector is 16 and 20 samples respectively; the weighted L-S algorithm has similarly the best result.

Key words　Gain Filter, Weighted L-S, Finite memory, G.728, SNR estimation

1　引言

码激励语音编码算法目前几乎都采用增益-波形乘积码书结构，增益量化器的良好性能对 LD-CELP 至关重要。G.728 在对比了固定系数的 Jayant[1]增益滤波器后选择 10 阶 LPC 模型并采用 Levinson-Durbin(L-D)方法更新系数[2,3]。但是文献[2]没有给出这两个滤波器的性能比较，其原因可能是由于考察增益预测器时量化器还不存在，因此无法用量化信噪比评价滤波器。本文提出一种信噪比估计方法使优化增益预测器可以与量化器优化问题分开处理。然后用 3 种新的滤波器替代 G.728 的 L-D 方法，它们分别是加权 L-S 算法、有限记忆算法以及 BP 神经网络。分别重新迭代计算各自的量化区间和量化电平后，3 种新滤波器增益量化方案的语音编码效果均好于 G.728，其中加权 L-S 方法 SNR 高出 G.728 算法 0.76dB。

2　信噪比估计

G.728 算法根据式(1)进行码书搜索:

$$D_{\min}=\sigma^2(n)\left\|\hat{\boldsymbol{x}}(n)-g_i\boldsymbol{H}(n)\boldsymbol{y}_j\right\|^2 \tag{1}$$

这里 $\sigma(n)$ 是激励增益的预测值。$\hat{\boldsymbol{x}}(n)$ 是被 $\sigma(n)$ 调节的目标矢量，$\boldsymbol{H}(n)$是短时预测器的单位冲激响应，g_i 和 $\boldsymbol{y}_j$ 分别是增益和波形码矢。式(1)的最小化与式(2)最大化等价

$$\hat{D}_{\max}=2g_i\boldsymbol{P}^{\mathrm{T}}(n)\boldsymbol{y}_j-g_i^2E_j \tag{2}$$

式(2)中 $\boldsymbol{P}(n)=\boldsymbol{H}^{\mathrm{T}}\hat{\boldsymbol{x}}(n)$，$E_j=\left\|\boldsymbol{H}\boldsymbol{y}_j\right\|^2$。若令 $\partial\hat{D}_{\max}/\partial g_j=0$，可得增益码字的精确值：

$$g_j=[\boldsymbol{P}^{\mathrm{T}}(n)y_i]/E_j \tag{3}$$

为表达简单，令 $\boldsymbol{y}_j$ 是归一化波形码书，式(1)可写成

$$D_{\min}=\left\|\boldsymbol{x}(n)-G_j(n)\boldsymbol{H}(n)\boldsymbol{y}_j(n)\right\|^2 \tag{4}$$

这里 $\boldsymbol{x}(n)=\sigma(n)\hat{\boldsymbol{x}}(n)$ 而 $G_j(n)=g_j(n)\sigma(n)$分别是未经增益调节的目标矢量和激励增益的精确值。由于 $g_j(n)=G_j(n)/\sigma(n)$，可知对数域预测残差是

$$\log_2 g_j(n)=\log_2 G_j(n)-\log_2\sigma(n) \tag{5}$$

令 $\hat{G}(n)$ 表示 $G_j(n)$的量化值, 用 $Q\{\bullet\}$表示对信号$\{\bullet\}$的量化, 则

$$\log_2\hat{G}_j(n)=Q\{\log_2 g_j(n)\}+\log_2\sigma(n) \tag{6}$$

选择预测器时信号尚未量化，所以只能用其真值代替，这导致系统处于最佳状态。因此，无法用量化信噪比评价预测器性能。但是可以对信噪比进行估计[3]。由预测残差 $\varepsilon_j(n)=G_j(n)-\sigma(n)$知

$$g_i=\frac{G_j(n)}{\sigma(n)}=\frac{G_j(n)}{G_j(n)-\varepsilon_j(n)}=\frac{\mathrm{snr}_j(n)}{\mathrm{snr}_j(n)-1} \tag{7}$$

这里 $\mathrm{snr}_j(n)=G_j(n)/\varepsilon_j(n)$ 是 n 时刻信号 $G_j(n)$与预测残差 $\varepsilon_j(n)$之比，从而有

$$\mathrm{snr}_j(n)=g_j(n)/[g_j(n)-1] \tag{8}$$

由式(8)知 g_j 越接近于 1，$\mathrm{snr}_j(n)$对信噪比的贡献越大。

2006-01-14 收到，2006-06-13 改回
国家自然科学基金(60372058)和山西省自然科学基金(20041046)资助课题

取 $\mathrm{snr}_j(n)$的一个观测值 snr，对所有的 j 可以计算满足 $\mathrm{snr}_j(n)$ >snr 的个数。假定 Δ 是 SNR 的一个预先设定的水平。即有

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

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

当 $g_j(n)>1$ 时有 $g_j(n) \leq (1-10^{-\Delta/20})^{-1}$，当 $g_j(n)<1$ 时有 $g_j(n) \geq (1+10^{-\Delta/20})^{-1}$。由此得到信噪比估计方法：对于不同的预测器方案分别计算各自的增益预测值 $\sigma(n)$，此时量化信号 $\hat{G}_j(n)$ 用 $G_j(n)$ 代替。选择 SNR 的水平Δ并考查相应的范围

$$(1+10^{-\Delta/20})^{-1} \leq g_j(n) \leq (1-10^{-\Delta/20})^{-1} \tag{9}$$

计算每个方案中 $g_j(n)$满足不等式(9)的比例，比例较大的方案更好。

3 增益量化

3.1 原理

图 1 是增益量化原理块图[4,5]。对 $\log_2 g_j(n)$量化获得 3bit 索引 $I(n)$被送到解码端，同时 $I(n)$在本地解码后得到量化差值信号 $\log_2 \hat{g}_j(n)$；加上增益预测值 $\log_2\sigma(n)$便得到本地重建信号 $\log_2 \hat{G}_j(n)$；输入到 10 阶 LPC 预测器可产生增益预测值 $\log_2\sigma(n)$；与增益采样 $\log_2 G_j(n)$相减可得新的预测残差 $\log_2 g_j(n)$。

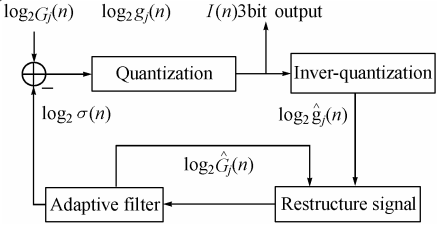

图 1 激励增益量化器块图

Fig.1 Excitation gain quantizer

3.2 最优量化参数

$\log_2 g_j(n)$的最佳量化效果可以通过选择一组最佳量化器参数获得。图 2 是 $\log_2 g_j(n)$的分布 $p(n)$。不计符号位，$\log_2 \hat{g}_j(n)$是 4 个量化电平 η_i (i=0 … 3)之一，量化误差的方差为

$$\sigma_e^2 = \sum_{i=0}^{3} E\left[(\eta_i - \log_2 g_j(n))^2\right] = \sum_{i=0}^{3}\int_{\xi_i}^{\xi_{i+1}} (\eta_i - x)^2 p(x)\mathrm{d}x \tag{10}$$

选择参数 ξ_i 和 η_i 使 σ_e^2 最小化，这里 $\xi_0 = -\infty$，$\xi_4 = +\infty$。令 $\partial\sigma_e^2/\partial\xi_i = \partial\sigma_e^2/\partial\eta_i = 0$，可得

$$\eta_i = \int_{\xi_i}^{\xi_{i+1}} xp(x)\mathrm{d}x \Big/ \int_{\xi_i}^{\xi_{i+1}} p(x)\mathrm{d}x \tag{11}$$

$$\xi_i = (\eta_{i-1} + \eta_i)/2 \tag{12}$$

式(11)说明 η_i 的最优位置处于 ξ_i 和 ξ_{i+1} 的距心，而式(12)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 η_i (i=0, … ,3)迭代计算的方法如下[5]：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(10)中的分布函数 $p(n)$。取 Δ 的中值为 x，当 Δ 足够小时，$\log_2 g_j(n)$可以

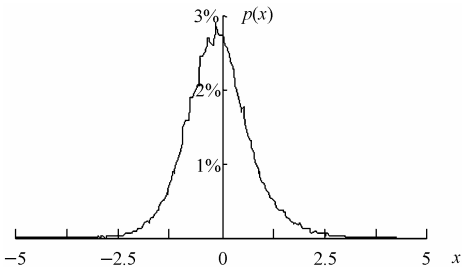

图 2 增益的分布

Fig.2 Gain distribute

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(11)，式(12)计算新的 ξ_i，η_i，如此迭代直到其值稳定。

4 增益滤波器评价

4.1 加权 L-S 算法

对于 p 阶 AR 模型

$$x_t = \boldsymbol{\Phi}^{\mathrm{T}}(t)\alpha(d,n) + e_t(d,n) \tag{13}$$

其中 $\alpha(d,n) = (\alpha_1(d,n), \cdots, \alpha_p(d,n))^{\mathrm{T}}$ 是基于采样信号 $x_{d,}\ x_{d+1,}\ \cdots\ x_n$ 对模型参数的估计。角标 T 表示向量或矩阵的转置。矢量 $\boldsymbol{\Phi}(t) = (x_{t-1}, \cdots, x_{t-p})^{\mathrm{T}}$ 是最新的 p 个采样，而 $e_t(d,n)$是方差为 σ^2 的零均值白噪声。令

$$\boldsymbol{H}(d,n) = (\boldsymbol{\Phi}(d), \boldsymbol{\Phi}(d+1), \cdots, \boldsymbol{\Phi}(n))^{\mathrm{T}}$$
$$\boldsymbol{Z}(d,n) = (x_d, x_{d+1}, \cdots, x_n)^{\mathrm{T}}$$
$$\boldsymbol{V}(d,n) = (e_d(d,n), e_{d+1}(d,n), \cdots, e_n(d,n))^{\mathrm{T}}$$
$$\boldsymbol{P}(d,n) = [\boldsymbol{H}^{\mathrm{T}}(d,n)\boldsymbol{H}(d,n)]^{-1}$$

式(13)可写成矢量形式

$$\boldsymbol{Z}(d,n) = \boldsymbol{H}(d,n)\boldsymbol{\alpha}(d,n) + \boldsymbol{V}(d,n) \tag{14}$$

相应的增长记忆加权 L-S 滤波器为递归公式

$$P(d,n+1) = \frac{1}{\lambda}[I - K_+(d,n+1)\Phi^{\mathrm{T}}(n+1)]P(d,n) \tag{15}$$

$$K_+(d,n+1) = P(d,n)\Phi(n+1)/\gamma_+(d,n+1) \tag{16}$$

$$\gamma_+(d,n+1) = \lambda + \Phi^{\mathrm{T}}(n+1)P(d,n)\Phi(n+1) \tag{17}$$

$$\hat{e}_{n+1}(d,n) = x_{n+1} - \Phi^{\mathrm{T}}(n+1)\hat{\alpha}(d,n) \tag{18}$$

$$\hat{\alpha}(d,n+1) = \hat{\alpha}(d,n) + K_+(d,n+1)\hat{e}_{n+1}(d,n) \tag{19}$$

初始化时，取 $\Phi^{\mathrm{T}}(0)$为零矩阵，$P(0,n)$是 $\delta^{-1}I$，其中 δ 是一个很小的值，典型取值为 0.01 或更小。采用不等式(9)判定滤波器性能，G.728 的 L-D 方法在 Δ=10dB 时结果为 0.467372。表 1 是 10 阶加权 L-S 滤波器在不同权值下的试验结果。可以看出最好结果为λ=0.982。

表 1 10 阶加权 L-S 算法

Tab.1 10-th order weighted L-S filer

权值	Δ=10dB	Δ=20dB	Δ=40dB
0.7	0.376214	0.121026	0.0120986
0.8	0.428924	0.139822	0.0139927
0.9	0.492718	0.16346	0.0164229
0.95	0.52484	0.175417	0.0175205
0.985	0.537051	0.179451	0.017964
0.982	0.537257	0.179531	0.0179327

4.2 有限记忆算法

利用式(20)–(24) 可以去除最远的历史数据对增长记忆

公式(15) – 式(19)的影响，此时令加权因子λ=1。所以只要知道如何从 $\hat{\alpha}(d,n+1)$ 推出 $\hat{\alpha}(d+1,n+1)$ 即可。

$$P(d+1,n+1)=[I+K\_(d+1,n+1)\Phi^{\mathrm{T}}(d)]P(d,n+1) \quad (20)$$

$$K\_(d+1,n+1)=P(d,n+1)\Phi(d)/\gamma\_(d+1,n+1) \quad (21)$$

$$\gamma\_(d+1,n+1)=1-\Phi^{\mathrm{T}}(d)P(d,n+1)\Phi(d) \quad (22)$$

$$\hat{e}_d(d,n+1)=x_d-\Phi^{\mathrm{T}}(d)\hat{\alpha}(d,n+1) \quad (23)$$

$$\hat{\alpha}(d+1,n+1)=\hat{\alpha}(d,n+1)-K\_(d+1,n+1)\hat{e}_d(d,n+1) \quad (24)$$

令记忆长度τ = n+1 – d。选择τ 的过程从 70 样点逐次加 10 样点进行记忆长度实验，直到最后选出最佳结果。表 2 是不同记忆长度的试验结果。从表 2 中可以看出记忆长度为 210 时，算法效果最好。

表 2 有限记忆算法

Tab.2　Finite memory filter

记忆长度	Δ=10dB	Δ=20dB	Δ=40dB
70	0.494479	0.164397	0.0164962
80	0.505794	0.168841	0.016966
140	0.524942	0.176437	0.0176754
150	0.525548	0.176635	0.0178414
210	0.526329	0.17662	0.01773
220	0.526084	0.176544	0.0177224

4.3　BP 神经网络

BP 神经网络的输入数据采用线性归一化方案预处理。用 *I-H-O* 表示其结构框架，其中，*I* 是输入节点数，*H* 和 *O* 分别是隐层和输出节点数。例如 BP 网络结构 5-4-7-1 表示 5 个输入节点, 4 个第 1 隐层和 7 个第 2 隐层节点，1 个输出节点。确定神经网络结构时先选用一个隐层并使用较少的隐层神经元，不断地增加隐层神经元，直到效果满意或再增加一个隐层。表 3 是较好的神经网络结构。

表 3 单隐层固定系数神经网络

Tab.3 BP neural network a hidden layer

BP 结构	Δ=10dB	Δ=20dB	Δ=40dB
10-5-1	0.461361	0.147223	0.0147601
9-3-1	0.488967	0.161131	0.0161372
4-10-1	0.497861	0.167168	0.0169031

5　语音编码测试

表 4 至表 6 分别是上述 3 种滤波器重新迭代的量化区间和量化电平。

表 4 加权 L-S 量化器

Tab.4 Weighted L-S quantizer

量化器输入范围 $\log\|g_l(n)\|$	$\|I(n)\|$	量化器输出
[1.20897, +∞)	3	1.77599
[0.27048, 1.20897)	2	0.63936
[– 0.53666, 0.27048]	1	– 0.10131
[– ∞, – 0.53666]	0	– 0.97355

表 5 有限记忆量化器

Tab.5 Finite memory quantizer

量化器输入范围 $\log\|g_l(n)\|$	$\|I(n)\|$	量化器输出
[1.24175, +∞)	3	1.81916
[0.283695, 1.24175)	2	0.66433
[– 0.540965, 0.283695]	1	– 0.09694
[– ∞, – 0.540965]	0	– 0.98499

表 6 神经网络 4-10-1 量化器

Tab.6 4-10-1 BP quantizer

量化器输入范围 $\log\|g_l(n)\|$	$\|I(n)\|$	量化器输出
[2.19897, +∞)	3	2.86077
[1.0263, 2.19897)	2	1.53716
[– 0.050015, 1.0263)	1	0.51543
[– ∞, – 0.050015]	0	– 0.4154

用 15 分钟语音信号约 700 万样点进行编解码测试结果如表 7。可以看出 3 种新滤波器方案的语音编码效果均好于 G.728 的 L-D 方法。加权 L-S 方法语音编码效果最好, 其平均分段 SNR 高出 G.728 算法 0. 76dB。这与采用不等式(9)判断滤波器优劣的结果是一致的。就 10 阶预测器计算复杂性而言，G.728 中 L-D 的计算量为 10 次除法和 1810 次乘加法；由于不必进行加窗计算, 加权 L-S 算法仅为 2 次除法和 650 次乘加；有限记忆算法是加权 L-S 算法的两倍，为 4 次除法和 1310 次乘加；计算量最低的是结构为 4-10-1 的固定权值神经网络，仅为 122 次乘加，其平均分段 SNR 也高出 G.728 约 0.156dB。

表 7 语音编码效果

Tab.7 Speech coding effect

增益滤波器算法	平均分段信噪比(dB)
G.728 的 L-D 方法	18.4506
有限记忆	18.9696
加权 L-S	19.2119
4-10-1 神经网络	18.6065

6　16 维和 20 维矢量

进行 16 样点激励矢量或 20 样点激励矢量增益的滤波器实验对提高 LD-CELP 算法的编码速率有重要意义。表 8 和表 9 列出了 3 种 10 阶 AR 滤波器在 16 和 20 维激励矢量的试验结果，其中加权 L-S 算法的权值选择为 0.982，有限记忆算法的记忆长度选作 210 个样点。表 10 和表 11 分别是 16 和 20 维激励矢量的 BP 神经网络非线性滤波的实验结果。从中可以看出对 16 样点和 20 样点激励矢量增益，加权 L-S 算法仍然具有最好的性能。根据第 5 节的结果，有理由相信上述 3 种算法应用于语音编码后，效果也会优于 G.728 的 L-D 方法。

表 8 16 维样点的试验结果

Tab.8 Result of L-D for 16 Dim. vectors

算法	Δ=10dB	Δ=20dB	Δ=40dB
加权 L-S	0.646814	0.237109	0.023973
有限记忆	0.592455	0.212723	0.021686
L-D 方法	0.58858	0.210464	0.0213981

表 9 20 维样点的试验结果

Tab.9 Result of L-D for 20 Dim. vectors

算法	Δ=10dB	Δ=20dB	Δ=40dB
加权 L-S	0.665101	0.24872	0.025646
有限记忆	0.62788	0.231733	0.023809
L-D 方法	0.612049	0.225571	0.0230044

表 10 16 维样点的试验结果(BP 神经网络)

Tab.10 Result of BP for 16 Dim. vectors

BP 结构	Δ=10dB	Δ=20dB	Δ=40dB
3-4-1	0.626763	0.224912	0.0227424
3-6-1	0.6229	0.223508	0.0225901
3-8-1	0.62617	0.225001	0.016774
5-2-4-1	0.61828	0.221953	0.0225449
3-7-5-1	0.631193	0.226982	0.0231033
4-5-3-1	0.629496	0.226839	0.0229274

表 11 20 维样点的试验结果(BP 神经网络)

Tab.11 Result of BP for 20 Dim. vectors

BP 结构	Δ=10dB	Δ=20dB	Δ=40dB
2-8-1	0.647425	0.241631	0.0246328
2-9-1	0.647417	0.241656	0.02467167
2-10-1	0.643073	0.239902	0.0245125
2-5-7-1	0.643597	0.240527	0.0245996
2-5-9-1	0.643297	0.240343	0.0245384
2-5-10-1	0.644981	0.241927	0.0247044

7 结束语

按照本文提出的判断增益滤波器性能的方法，所选择的 3 种激励增益滤波器算法效果均优于 G.728 的 L-D 方法。对于 5 样点、16 样点和 20 样点激励矢量，加权 L-S 算法效果最好。用于 5 样点激励矢量语音编码结论依然成立：加权 L-S 算法具有最好的性能；其平均分段 SNR 比 G.728 高 0.76dB，运算量仅为 G.728L-D 算法的 30%。固定权值结构为 4-10-1 的 BP 神经网络运算量最低，仅为 G.728 L-D 算法的 6.7%，但其平均分段 SNR 高出 G.728 约 0.156dB。

参 考 文 献

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

[2] Chen J H. A robust low-delay CELP speech coder at 16 kb/s. Advances in Speech Coding, Kluwer Academic Publishers, 25-35, 1991.

[3] CCITT, Recommendation G.728. Coding of speech at 16kbit/s using low-delay code excited linear prediction. Geneva, 1992.

[4] 薛春雨. 低延迟码激励话音编码算法中激励增益滤波器的优化与选择. ［硕士论文］，太原理工大学，2005.

[5] Zhang G, Xie K M, Zhang X Y, Huangfu L Y. Optimizing gain codebook of LD-CELP. Proceeding of ICASSP, Hong Kong, 2003, Vol.2: 149–152.

张起贵： 男，1963 年生，副教授，主要研究方向为语音编码.

张 刚： 男，1953 年生，教授，主要研究方向为语音编码与嵌入式系统.