

语音压缩编码中的两种基音预测器¹

鲍长春 戴逸松* 樊昌信

(西安电子科技大学信息科学研究所 西安 710071)

*(吉林工业大学 长春 130025)

摘 要 本文研究了语音压缩编码中基音预测器的两种情形,即开环和闭环结构。提出了一些简化基音预测器求解方法,比较了几种条件下的基音预测器性能。给出了计算机仿真结果。

关键词 语音处理, 线性预测, 基音预测

中图分类号 TN912.3

1 引 言

当今在中、低速率语音压缩编码方案中,最普通的语音合成模型为 B.S. Atal 提出的随机激励线性预测模型或码激励线性预测模型^[1],如图 1(a)所示。这种模型的基本思想就是利用几乎是白的信号激励两个线性滤波器,一个考虑了浊音语音的周期性,另一个则考虑了语音的短时谱包络。这种模型来源于它的逆过程——预测过程,如图 1(b)所示。 $F(z)$ 为短时预测滤波器,或共振峰滤波器,它排除了近样点冗余度; $P(z)$ 为长时预测滤波器,或基音滤波器,它排除了远样点冗余度,经过两级预测得到的归一化残差信号近似服从于标准正态分布。

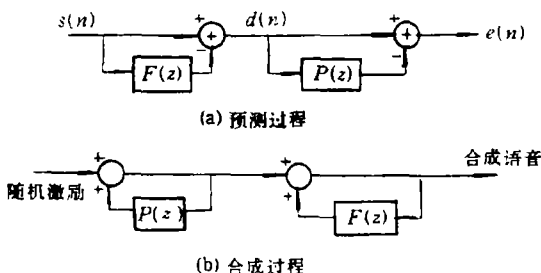


图 1

目前的一些中、低速率语音压缩编码方案中,诸如 16kb/s 低延时码激励线性预测 (LD-CELP) 编码器^[2],北美数字蜂窝无线标准 8kb/s 矢量和激励码激励线性预测 (VSELP) 编码器^[3],泛欧数字蜂窝无线标准 13kb/s 带长时预测器的、规则脉冲激励 (RPE-LTP) 编码器^[4],美国联邦标准 (FS1016)4.8kb/s 码激励线性预测 (CELP) 编码器^[5],均采用了长时预测器

(基音预测器),保证了重建语音的质量。在高比特速率,可以分配足够多的比特给激励信号,以保证它们重建谐波结构,这时长时预测器接近一个准确模型。然而,在中、低比特速率,激励没有如此多的易变性,语音质量主要取决于长时预测器的性能。

本文的目的就是研究组合短时和长时预测器在开环、闭环形式下的各种性能和简化基音预测器算法的一些方法。预测器的级联形式为 S-L(即短时预测器在前),实现方式为横向结构,给出了实验结果。

¹ 1994-11-07 收到, 1995-06-24 定稿

2 短时和长时预测器

一般地, 短时预测器传递函数为

$$F(z) = \sum_{k=1}^{N_f} \alpha_k z^{-k}, \quad (1)$$

典型的阶数 N_f 介于 8 和 16 之间, 噪声反馈滤波器 (或称感知加权滤波器) 的系统函数通常为 $W(z) = F(z/\alpha)$, 这里 $0 < \alpha < 1$ 。通过改进谱水平低的那些区域的信噪比 SNR, 可以减少输出语音的感性失真, 然而这个改进是以降低共振峰区域的 SNR 为代价的。在接收端, 短时合成滤波器的传递函数为 $H_f(z) = 1/[1 - F(z)]$ 。

基音预测滤波器有较小数量的抽头 N_f , 与这些抽头相联系的延迟集结在基音延迟值的周围, 对于 1, 2, 3 个抽头的基音预测器的系统函数为

$$P(z) = \begin{cases} \beta_1 z^{-M}, \\ \beta_1 z^{-M} + \beta_2 z^{-(M+1)}, \\ \beta_1 z^{-M} + \beta_2 z^{-(M+1)} + \beta_3 z^{-(M+2)}. \end{cases} \quad (2)$$

基音延迟 M 通常随系数一起修正, 基音合成滤波器的系统函数为 $H_p(z) = 1/[1 - P(z)]$ 。

在语音编码方案中, 基音预测器在浊音语音中是最有用的, 这是因为浊音语音被表征为由基音周期分隔的样点间大量相关的准周期信号, 在 S-L 级联形式下, 短时预测器在很大程度上排除了近样点间的相关, 得到的短时预测残差信号是一主要由基音尖峰信号构成的低密度准周期信号, 基音预测器作用于这个残差信号上。如果基音延迟是整数个样点, 则一抽头基音预测器能去除基音周期相关, 对于非整数基音周期, 多抽头基音预测器和分数延迟基音预测器能排除这些远抽样点相关。

3 开环和闭环基音预测器及其简化算法

3.1 开环结构下的基音预测器及其简化算法

本文采用协方差法进行开环基音预测, 对于三抽头基音预测, 其系统方程为

$$\begin{bmatrix} \phi(M, M) & \phi(M, M+1) & \phi(M, M+2) \\ \phi(M+1, M) & \phi(M+1, M+1) & \phi(M+1, M+2) \\ \phi(M+2, M) & \phi(M+2, M+1) & \phi(M+2, M+2) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \phi(0, M) \\ \phi(0, M+1) \\ \phi(0, M+2) \end{bmatrix}, \quad (3)$$

这里

$$\phi(i, j) = \sum_{n=0}^{N-1} d(n-i)d(n-j), \quad (4)$$

$d(n)$ 为基音预测器的输入信号, 即短时预测残差。一般矩阵 Φ 不是 Toeplitz 阵, 可用 Cholesky 分解法解系统方程。对于合理的帧体积和较小数量的抽头, Φ 可以修正为具有很小预测增益损失的 Toeplitz 阵, 那么 Levinson 递推算法可用于确定预测器系数, 而 Φ 的 Toeplitz 阵性质不一定保证 $H_p(z)$ 的稳定性, 但确实是系统方程的一种有效解法。文献 [6] 讨论了基音滤波器的稳定性问题, 由不稳定性引起的预测增益下降并不很大。本文提出了一种开环结构下的基音预

测滤波简化算法,即忽略协方差阵中主对角线上下部分的元素,这些元素是具有较小延迟的相关项。这些近似基于这样的观察,即短时预测滤波器趋于排除掉短时相关,运用这种改变,矩阵 Φ 变成了对角阵。对于三抽头基音预测器,可得如下系数阵

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \phi(0, M)/\phi(M, M) \\ \phi(0, M+1)/\phi(M+1, M+1) \\ \phi(0, M+2)/\phi(M+2, M+2) \end{bmatrix}, \quad (5)$$

这里 $\phi(i, j)$ 为短时残差信号 $d(n)$ 的相关函数。这种简化方法得到的预测增益与全复杂解相比较仅下降 0.1dB(40 样点帧长)。得到了预测系数后,由最小均方误差 $\epsilon^2 = \phi(0, 0) - c^T \alpha$ 可得到基音估计公式为

$$c^T \alpha \approx \sum_{t=M}^{M+N_p-1} \frac{\phi^2(0, t)}{\phi^2(t, t)}. \quad (6)$$

那么使得 (6) 式最大的 M 即为所求。对于一抽头基音预测器 (6) 式不存在近似关系,即

$$c^T \alpha = \frac{\phi^2(0, M)}{\phi^2(M, M)} = \frac{\sum_{n=0}^{N-1} d(n)d(n-M)}{\sum_{n=0}^{N-1} d^2(n-M)}. \quad (7)$$

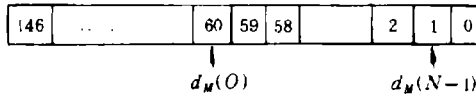


图 2

由于 (6) 式计算量很大,故本文采取了利用一抽头基音预测器延迟作为估算值,运算量下降了 2/3,而预测增益下降很少。为便于计算,本文依据基音延迟 M 的统计范围 20~147,设计了一个 147 位的短时残差 $d(n)$ 的移位寄存器。移位寄存器结构如图 2 所示,图中的数字 0~146 表示移位寄存器位置。

将每帧得到的短时残差推入移位寄存器,以备计算 $d(n-M), 0 \leq n \leq N-1$ 项。令 N 为基音预测器帧长,当 $M \geq N$ 时,从移位寄存器可直接得到所要求的 $d(n-M), 0 \leq n \leq N-1$ 。例如,当 $M = N+1$ 时,对应于 $d(n-M)$ 的值为 $d_M(n), 0 \leq n \leq N-1$,如图 2 所示。当 $M < N$ 时,前 M 个 $d_M(n), 0 \leq n \leq N-1$,在移位寄存器为 0,1, ..., $M-1$ 位,将这 M 个样点重复即可形成长度为 N 的序列 $d_M(n), 0 \leq n \leq N-1$ 。例如,当 $M=20$ 时,形成的序列占据移位寄存器位置为 19,18, ..., 0,19,18, ..., 0,19,18, ..., 0;当 $M=58$ 时,形成的序列占据移位寄存器位置为 57,56, ..., 57,56。

3.2 闭环结构下的基音预测器及其简化算法

本文提出一种新的自激励一抽头闭环基音预测器,它不同于 B.S. Atal 提出的多脉冲激励情况下的闭环法^[7]。新方法定义的闭环基音预测器,更易于理解。闭环基音预测器的初衷就是使得预测到的基音延迟 M 更好地跟踪实际语音的周期性变化。但它的预测增益并不比开环法高,这是因为它引入了感性加权滤波器,使得浊音区的预测增益下降,导致平均预测增益下降,但它确实使得平均分信噪比升高。关于这个结果将在后面的仿真实验中看到。下面给出自激励一抽头闭环基音预测器框图,如图 3 所示。

原语音与合成语音间的感性加权
平方误差为

$$E = \sum_{n=0}^{N-1} [s''(n) - \beta s'(n)]^2, \quad (8)$$

其中 $s''(n)$ 为加权原语音减掉零输入 $s_0(n)$ 后得到的语音, $s'(n)$ 为过去激励 $d(n-M)$ 通过感知加权合成滤波器的输出语音, 表示为

$$s'(n) = \sum_{k=0}^n h(k) d(n-M-k). \quad (9)$$

通过使 E 最小, 即令 $\partial E / \partial \beta = 0$ 得

$$\beta = \sum_{n=0}^{N-1} s''(n) s'(n) / \sum_{n=0}^{N-1} s'^2(n) \quad (10)$$

$$E_{\min}(M) = \sum_{n=0}^{N-1} s''^2(n) - E'_{\max}(M), \quad (11)$$

$$E'_{\max}(M) = \left[\sum_{n=0}^{N-1} s''(n) s'(n) \right]^2 / \sum_{n=0}^{N-1} s'^2(n). \quad (12)$$

选择使 (12) 式最大的 M 即为所求。这里对基音延迟 M 的搜索仍然使用了 3.1 节中介绍的 147 移位寄存器。计算 (12) 式的计算量很大, 尤其是涉及到了卷积运算 $s'(n)$ 。这时我们可以从 147 位移位寄存器所固有的性质中得到简化算法。

当 $M \geq N$ 时, 过去激励 $d(n-M)$, $0 \leq n \leq N-1$, 存在着叠接 1 的属性, 即

$$d_M(n) = SR(M-n-1), \quad 0 \leq n \leq N-1, \quad N \leq M \leq 147 \quad (13)$$

式中 $SR(n)$ 为移位寄存器中内容值。这时我们可以得到一递推算算法如下:

$$\left. \begin{aligned} s'_M(n) &= \sum_{k=0}^n d_M(k) h(n-k) = \sum_{k=0}^n SR(M-k-1) h(n-k) \\ &= SR(M-1) h(n) + \sum_{k=0}^{n-1} SR[(M-1)-k-1] h(n-1-k) \\ &= SR(M-1) h(n) + s'_{M-1}(n-1), \quad 1 \leq n \leq N-1, \\ s'_M(0) &= SR(M-1) h(0). \end{aligned} \right\} \quad (14)$$

这时计算卷积项 $s'(n)$ 仅需 N 次乘法, 而纯卷积则需 $N(N+1)/2$ 次乘法。当 $M < N$ 时,

$$d_M(n) = SR(M-1-n \% M), \quad 0 \leq n \leq N-1, \quad 20 < M < N, \quad (15)$$

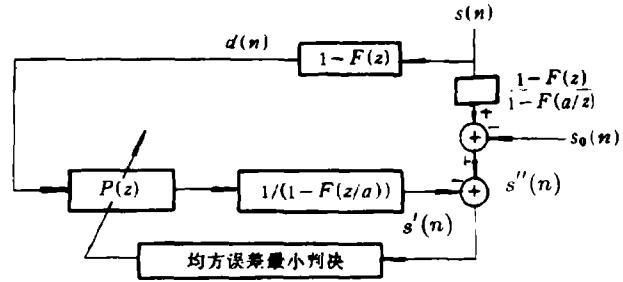


图 3

这里符号 % 表示求余, 此时递推公式为

$$\begin{aligned} s'_M(0) &= SR(M-1), \quad n=0; \\ s'_M(n) &= SR(M-1)h(n) + s'_{M-1}(n-1), \quad 0 < n < M; \\ s'_M(n) &= SR(M-1)h(n) + s'_{M-1}(n-1) + s'_M(n-M) - s'_{M-1}(n-M), \quad M \leq n \leq N-1. \end{aligned} \quad (16)$$

3.3 开环结构下的分数延迟基音预测器

由于语音的基音周期平滑变化, 故基音延迟不一定为整数值, 可能为实数, 这将有利于高基音语音的重建, 保持其恢复的谐波结构不致于破坏很大。本文在开环一抽头基音预测器基础上, 研究了分数延迟基音预测器的结构及其简化算法。依据多相滤波原理^[8], 我们可以得到非整数延迟 $M + l/D$ 的一抽头基音预测误差滤波器传递函数为

$$P(z) = 1 - \beta \sum_{k=0}^{q-1} p_l(k) z^{-(M-I+k)} \quad (17)$$

式中 I 为低通滤波器固有延迟, q 为多相滤波器系数, $q = [N/D]$, $N = 2ID + 1$, D 为分数延迟的分割数, $p_l(k), l = 0, 1, \dots, D-1$, 为汉明窗加权的 $\text{sinc}(x)$ 函数。选择不同的 D, I 值得到不同的预测性能。文献 [9] 给出了如何求解分数延迟的方法, 本文在开环情况下给出了它的算法。由 (17) 式得预测误差 $e(n)$ 为

$$e(n) = d(n) - \beta \sum_{k=0}^{q-1} p_l(k) d(n - M + I - k). \quad (18)$$

使 $\varepsilon^2 = \sum_{n=0}^{N-1} e^2(n)$ 最小, 可得

$$\beta = \frac{\sum_{n=0}^{N-1} [d(n) \sum_{k=0}^{q-1} p_l(k) d(n - M + I - k)]}{\sum_{n=0}^{N-1} [\sum_{k=0}^{q-1} p_l(k) d(n - M + I - k)]^2}, \quad (19)$$

误差函数

$$E(M) = \sum_{n=0}^{N-1} d^2(n) - E'(M), \quad (20)$$

$$E'(M) = \frac{\{\sum_{n=0}^{N-1} [d(n) \sum_{k=0}^{q-1} p_l(k) d(n - M + I - k)]\}^2}{\sum_{n=0}^{N-1} [\sum_{k=0}^{q-1} p_l(k) d(n - M + I - k)]^2}. \quad (21)$$

求得使 $E'(M)$ 最大的 M 即为所求。此处也应用了 3.1 节中给出的移位寄存器。大量的计算来源于卷积项 $\sum_{k=0}^{q-1} p_l(k) d(n - M + I - k)$, 因为对于每个分数延迟都要计算这一项。本文建议一种近似解法, 即在确定完整数延迟后, 再在此整延迟值周围进行小范围搜索求分数延迟。这大大降低了计算量, 而预测增益与全搜索相比下降较小, 语音质量几乎没有降低。由于人为引入了多相滤波, 使得预测增益下降, 但编码时确提高了平均分信噪比, 这是由于分数延迟更能准确地跟踪实际语音基音周期性变化的结果。

4 预测器性能分析和比较

本节将给出预测器在开、闭环情况下的性能分析, 包括预测器增益、平均分信噪比与帧长、加权因子和增量搜索范围的关系。短时预测分析采用了自相关法, 帧间搭接半帧; 长时预

测在奇子帧采用了增量搜索, 增量搜索范围分别为 $SUB = 64, 32$, 在求分数延迟时增量搜索范围分别 $SUB1=16, 2$ 。语音数据为四男四女讲话。表 1 给出了主帧长 240 点, 子帧长 60 点时开、闭环情况下的预测增益。表 2 给出了 $SUB=64$ 时预测增益与帧长的关系。表 3 给出了在 4.8kb/s CELP 编码方案中开、闭环情况下平均分信噪比。图 4 给出了某段语音几种情况下的预测增益曲线; 图 5 给出了一抽头, $SUB=64$ 时几种情况下的平均分信噪比曲线; 图 6 给出了 4.8kb/s CELP 语音编码方案中不同信号的波形。

表 1 主帧 240 点, 子帧 60 点时开、闭环预测增益 (dB)

		短时预测增益 $N_F=10$	开环多抽头基音预测增益			开环一抽头分延迟基音预测器增益 ($D=12, I=4$)		闭环一抽头基音预测器增益		
			$N_p=1$	$N_p=2$	$N_p=3$	$SUB1=16$	$SUB1=2$	$\gamma=0.8$	$\gamma=0.4$	$\gamma=0.2$
SUB=128	女性	7.461	3.258	3.418	3.410	3.157	3.109	2.714	3.170	3.234
	男性	7.223	2.834	3.040	2.958	2.720	2.631	2.296	2.675	2.787
SUB64	女性	7.461	3.029	3.256	3.260	3.038	2.933	2.584	2.979	3.009
	男性	7.223	2.598	2.906	2.859	2.535	2.452	2.147	2.459	2.558
SUB32	女性	7.461	2.833	3.114	3.139	2.945	2.818	2.326	2.733	2.786
	男性	7.223	2.536	2.836	2.791	2.493	2.392	2.117	2.410	2.498

表 2 SUB=64 时, 预测增益 (dB) 与帧长的关系

		主帧 =120		主帧 =160		主帧 =240		
		子帧 =30	子帧 =60	子帧 =40	子帧 =80	子帧 =48	子帧 =60	子帧 =120
开环 $N_p=1$	女性	7.903		7.803		7.461		
	男性	4.434	3.026	3.972	2.612	3.394	3.029	1.862
	女性	7.552		7.460		7.223		
闭环 $N_p=1, \gamma=0.8$	男性	3.290	2.552	3.062	2.001	2.808	2.598	1.400
	女性	7.903		7.803		7.461		
	男性	3.729	2.595	3.443	2.269	2.790	2.719	1.544
开环 $D=12, I=4$ $SUB1=2, N_p=1$	女性	7.552		7.460		7.223		
	男性	2.694	2.131	2.550	1.556	2.338	2.147	0.997
	女性	7.903		7.803		7.461		
闭环 $D=12, I=4$ $SUB1=2, N_p=1$	男性	4.243	2.931	3.748	2.548	3.231	2.933	1.860
	女性	7.552		7.460		7.223		
	男性	3.087	2.428	2.859	1.859	2.578	2.452	1.382

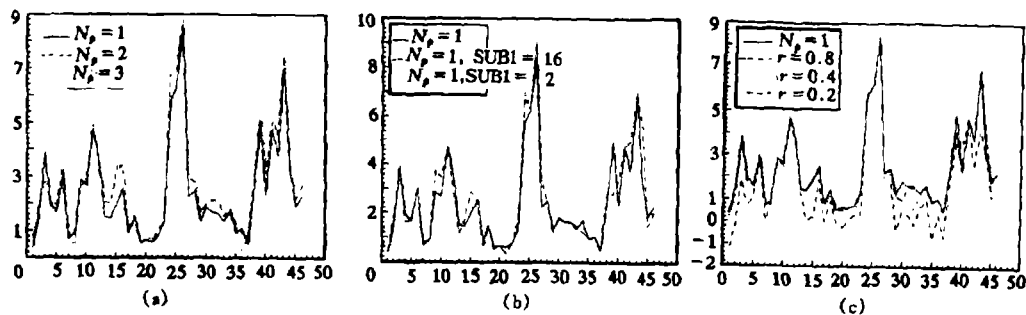


图 4

从表 1 和图 4 可得结论: 多抽头基音预测器增益高于单抽头基音预测器增益; 非整数基音延迟预测器增益低于不带分数延迟预测器增益, 闭环基音预测器增益与加权因子 γ 有关。从表 2 和图 5 可得结论: 预测增益与帧长成反比。从表 3 得出结论: 使用闭环基音预测器得到的平均分

段信噪比远远高于使用开环基音预测得到的平均分段信噪比,使用分数延迟得到的平均分段信噪比最高。分数延迟搜索范围 SUB1=2 时,尽管平均分段信噪比比 SUB1=64 时下降,但主观非正试听要比未加分数延迟时好得多,运算量大大下降。

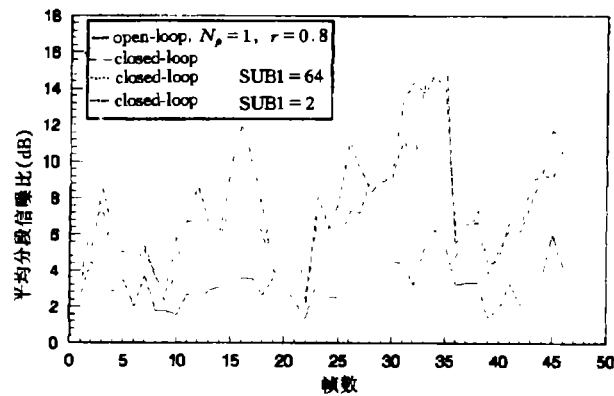


图 5

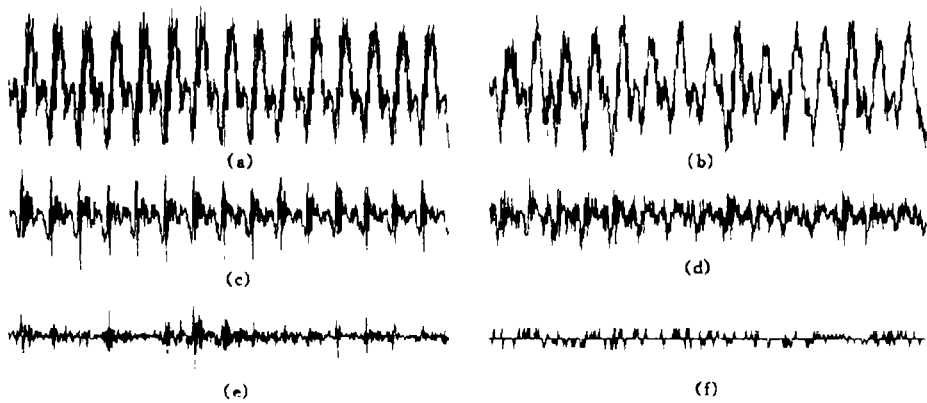


图 6

5 结 束 语

本文详细研究了两种结构的基音预测器,提出了一些简化算法,给出了性能比较,为同行在这方面的研究提供了一些借鉴。由于实验是在有限的语音数据上进行的,所以,更详细的研究还待继续。

表 3 SUB=64 时, 开、闭环情况下的分段信噪比 (dB)						
	开环 $N_p=1$			闭环 $N_p=1(D=12,I=4)$		
	$\gamma=0.8$	$\gamma=0.4$	$\gamma=0.2$	无分数延迟	有分数延迟	有分数延迟
				SUB1=64 SUB1=2		
女性	3.107	1.722	1.388	6.800	7.464	6.997
男性	3.453	1.943	1.337	6.779	7.260	7.010

参 考 文 献

[1] Schroeder M R, Atal B S. Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates. in Proc. IEEE Int. Conf. ASSP, Tampa, FL: Mar., 1985, 937-940.

[2] Cuperman V, Gersho A , Pettigrew R. IEEE Trans. on Com., 1992, 40(1): 129-139.

- [3] Gerson I, Jasiuk M A. Vector Sum Excited Linear Prediction(VSELP) Speech Coding at 8kbps. in Proc. IEEE Int. Conf. ASSP, New Mexico, USA: 1990, 461-464.
- [4] Vary P, *et al.* Speech Codec for the European Mobile Radio System. in Proc. IEEE Global Comm. Conf., Dallas, Texas: 1989, 1065-1069.
- [5] Campbell J P, Welch V C, Tremain T E. An Expandable Error-Protected 4800bps CELP Coder(U. S. Federal Standard 4800bps Voice Coder), in Proc. IEEE Int. Conf. ASSP, Glasgow, Scotland: 1989, 735-738.
- [6] Ramachandran R P, Kabal P. IEEE Trans. on ASSP, 1987, ASSP-35(3): 937-946.
- [7] Singhal Sand Atal B S. Improving Performance of Mutipulse LPC Coders at Low Bit Rates. Proc. ICASSP, San Diego, CA: Mar., 1984, 1.3.1-1.3.4.
- [8] Crochiere R, Rabiner L. Multirate digital signal processing. Engle Wood Cliffs, NJ: Prentice, 1983, Ch.6, 271-274.
- [9] 鲍长春, 等. 通信学报, 1994, 15(4): 53-61.

TWO KINDS OF PITCH PREDICTOR IN SPEECH COMPRESSING CODING

Bao Changchun Fan Changxin

(*Xidian University, Xi'an 710071*)

Dai Yisong

(*Jilin Univerdity of Technology, Changchun 130025*)

Abstract This paper studies two kinds of methods for pitch predictor in speech compressing coding, i. e., open-loop and closed-loop structure. Some of simplified approachs for solving pitch predictor equation are suggested, and the performances are compared under several conditions. The computer simulation results are shown.

Key words Speech processing, Linear Predictive, Pitch predictive

鲍长春: 男, 1965 年生, 副教授, 博士后, 中国电子学会高级会员, 研究方向为语音压缩编码, 多媒体通信等.

戴逸松: 男, 1936 年生, 教授, 博士生导师, 研究领域为弱信号检测, 随机数字信号处理, 语音数字信号处理, 神经网络及其应用, 谱估计等.

樊昌信: 男, 1931 年生, 教授, 博士生导师, 中国电子学会会士, 研究领域为语音和图象信号处理、B-ISDN 等.