

改进低比特率的 LPC 声码器

王 俊 生

(中国科学院声学研究所)

1983年12月29日收到

摘要 线性预测技术已广泛用于语言信号处理,特别是用于设计低比特率的声码器。但是传统的定帧长分析方法不能很好适应语言的非平稳过程,同时由此得到的语言参数(例如预测误差)用于提取音调也容易出错,因而影响了合成语言的质量。为此我们进行了三方面的改进:(1)用自适应梯型算法代替现有定帧分析的线性预测方法,以便得到更准确的声道参数;(2)有限的变帧抽样语言参数,改善了合成语音过渡区的性能;(3)改进 Gold-Rabiner 的基音提取技术,使音调提取方法更简单可靠。传送数据率为 2400 和 1200 比特每秒。在计算机上模拟结果表明,2.4kb/s 的方案所合成的语言较为自然易懂,且不难分辨熟人口音,1.2kb/s 方案的合成语言也未严重降级。采用缓存器后,两种方案均可在固定数据率信道上传输。

IMPROVEMENT OF LOW BIT RATE LPC VOCODER

WANG JUN-SHENG

(The Institute of Acoustics, Academia Sinica)

Received December 29, 1983

Abstract Linear predictive technique has been widely used for speech signal processing, especially for low bit rate vocoder in recent years. However the conventional fixed frame processing does not suit well with speech signal for its nonstationary behavior. Also it is not reliable to extract pitch from the parameters (e.g. residual) obtained by this method. Due to these reasons, speech quality can be adversely affected. In order to improve the quality and intelligibility, in this paper it presents improvement to the following three aspects: (1) Using adaptive ladder form algorithm instead of conventional LPC to obtain more accurate estimation of the vocal-tract transfer function (2) Adapting non-uniform sampling of speech parameters with constrained interval to improve performance of synthesized speech during fast transients. (3) Improving Gold-Rabiner pitch extraction technique for saving time and reliable. The computer simulated results show that the 2.4 kb/s vocoder gives the synthesized speech with good intelligibility and quality, and the quality of the 1.2 kb/s one is not degraded seriously. Transmission over fixed rate channels can be accomplished, using transmit and receive buffers.

一、引 言

到目前为止,在窄带语言通信领域,声码器仍处于独占的地位,因为其他编码方法很难达

到数据率 4.8kb/s 以下^[1]。人类语言直接数字化传输大约需要 64kb/s (8000Hz 采样, 8 比特量化) 典型的 ADPCM 可压缩比特率到 16kb/s^[2], ATC 技术可到 8kb/s^[3], 而低比特率的声码器可把语言信号的带宽压缩到 4.8, 2.4kb/s, 甚至几百比特秒^[4], 当然这要损失一定的语言质量。它生成的语言自然度和辨认人的特性均没有前述方法好, 但它依然有相当好的可懂度, 可利用现有的电话线路和卫星信道传输, 同时它是参数型的合成方法, 保密性能好, 是国防和政府部门保密通信的重要手段, 所以很多国家和政府资助这方面研究, 为提高其语言质量和进一步减少数据率。

随着电子技术的发展, 特别是 LSI 和 VLSI 技术的迅速发展, 为数字化实现比较复杂的算法提供了条件。早期的模拟式通道、共振峰型声码器已全可数字化实现, 从七十年代开始, 人们在这方面的研究兴趣已广泛地转向线性预测技术 (LPC)^[5] 和同态 (homomorphic) 处理方法^[6], 由于线性预测技术优于通道声码器^[7], 且它的运算复杂性又少于同态处理方法, 所以近十年来, LPC 声码器的研制最为广泛, 最近用新的微处理器 (NEC 的 UPD7720) 已制成体积小 (18 平方英寸一块电路板) 功耗低 (5.5W) 的 LPC 声码器^[8]。

硬件实现技术的迅速发展, 势必推动各种新算法的研究, 因为现有的 LPC 声码器的语言质量并不令人满意, 除了数据率的限制条件外, 普遍采用的定帧分析方法, 均匀抽样参数并不能很好表达语言的实际过程, 因为语言是一种非平稳的随机信号, 有些音素 (如 p, t, k) 持续期短至几个毫秒, 语音间的过渡变化很快, 通常定帧 (20—30ms) 分析显然不能很好表达这些快速变化。为克服这个问题, 我们采用自适应梯型算法^[9,10], 于 1981 年美国 102 次 ASA 会议上提出一个变帧率设计方案^[11], 在计算机上模拟实现了变帧的 LPC 声码器, 虽然数据率降到平均 800bits/s, 仍然保持相当满意的语言质量。但这种算法的数据率完全随着语言性质的自然变化而变, 很难在固定码率的信道上传输。本文现提出一种改进方案, 主要改进如下三方面: (1) 实现参数时间维选择, 消除声门脉冲干扰, 提高声道参数分析精度; (2) 改进基音提取算法; (3) 实行过渡区的有限变帧, 从而实现了 2400, 1200bits/s 的改进型的 LPC 声码器, 详细叙述在下面几节。

二、提高声道参数分析精度

本节我们讨论如何利用自适应的前窗 (PW) 型算法 (以下简称为 PW 算法), 提高声道参数的分析精度, 并和传统的分帧 LPC 方法进行比较。

语言的线性预测编码基于把语言信号模拟为一个时变线性系统的输出, 在短时分析中, 这个系统可近似为常参数的线性滤波, 并写作为

$$Y(z) = H(z)E(z) \quad (1)$$

其中 $E(z)$ 和 $Y(z)$ 是 z 变换域的输入和输出, $H(z)$ 是系统函数。考虑鼻音的表达, $H(z)$ 应该具有极点和零点 (ARMA 模型), 但为了实现的简单性, 只要有足够的阶次, 也可由全极点型 (AR 模型) 近似^[12], 此时

$$H(z) = 1 / \left(1 + \sum_{i=1}^P a_P^{(i)} z^{-i} \right), \quad (2)$$

在时域里, 这一模型等效为

$$e_T = y_T + \sum_{i=1}^P a_P^{(i)} y_{T-i} \quad (3)$$

其中 y_T 是 T 时刻的语言样点值, $a_P^{(i)}$ 是 P 阶线性预测的第 i 个预测系数, e_T 是 T 时刻的预测误差. P 个预测系数是声道系统的模型参数, 表征人的声道特征, 现有多种线性估算方法, 其目的都是为了准确求解这组预测系数或与之相对应的一组梯型参数(反射系数). 通常的线性预测分析(自相关或协方差方法)是分帧处理方法, 是应用最小均方误差(LMS)准则得到预测参数的正则方程组. 然而这个 LMS 准则对语言信号不是完全适合的, 在音调脉冲附近的语言波受到声门激励的干扰, (3) 式中 e_T 不可能是最小, 因此分帧处理技术将导致估算参数的误差, 这种误差往往产生声道共振峰增宽, 引起合成语言的含糊不清(fuzzy)^[13], 而且分帧处理方法中难以消除这种误差.

K. Steiglitz 等曾做过一个理论上的模拟计算^[14], 分析帧里若包含一个输入激励脉冲, 较窄的共振峰带宽能产生 45—51% 的估算误差.

由于声门脉冲的干扰, 即使在稳态的浊音区, 估算参数也因分析帧的位置不同而有差异. 为了克服这个问题, 提高参数的分析精度, 人们曾用协方差方法做音调同步分析^[15], 虽然有效, 但音调同步分析需要事先知道音调脉冲的准确位置, 而音调脉冲位置与语言波的实际峰位置又不一致, 所以在实际的分析系统中很难做到音调同步.

本文采用快速递归的 PW 算法, 用以得到预测方程的精确解, 同时应用似然变量 γ 的幅度阈值控制声道参数的时间维选择, 很简便地去掉音调脉冲的影响.

关于 PW 算法的推导可见文献 [16], 在此只列出归一化的计算公式:

$$\begin{aligned} \tilde{k}_{P+1,T} &= \tilde{k}_{P+1,T-1}(1 - \tilde{\varepsilon}_{P,T}^2)^{1/2}(1 - \tilde{r}_{P,T-1}^2)^{1/2} + \tilde{\varepsilon}_{P,T}\tilde{r}_{P,T-1} \\ \tilde{\varepsilon}_{P+1,T} &= (\tilde{\varepsilon}_{P,T} - \tilde{k}_{P+1,T}\tilde{r}_{P,T-1})/[(1 - \tilde{k}_{P+1,T}^2)(1 - \tilde{r}_{P,T-1}^2)]^{1/2} \\ \tilde{r}_{P+1,T} &= (\tilde{r}_{P,T-1} - \tilde{k}_{P+1,T}\tilde{\varepsilon}_{P,T})/[(1 - \tilde{k}_{P+1,T}^2)(1 - \tilde{\varepsilon}_{P,T}^2)]^{1/2} \end{aligned} \quad (4)$$

初始值

$$\begin{aligned} R_T &= \lambda R_{T-1} + Y_T^2 \\ \tilde{\varepsilon}_{0,T} &= \tilde{r}_{0,T} = Y_T/R_T^{1/2} \end{aligned}$$

其中 $\tilde{k}$, $\tilde{\varepsilon}$ 和 $\tilde{r}$ 分别是归一化反射系数, 向前和向后误差, λ 是所谓忘却系数. 通常选值为 0.95—0.995.

PW 算法公式的推导是基于最小二乘方 (LS) 准则, 同样在音调脉冲附近的误差不符合这个假定条件, 因为企图用声门激励前的样点去预测声门激励后的波形显然是不合情理的. 但这种方法是逐点更新 (update) 参数, 即每点都有一组预测参数的解, 可以由时间维的选择来去除这些不合理的参数组, 其方法有两种: (1) 用预测误差的均方根值作为阈值, 预测误差大于这个阈值 2—2.5 倍的参数不选; (2) 用似然变量 γ 的幅度为阈值, 因为 γ 不仅反映声门脉冲的位置, 而且还能反映声门脉冲影响的区域^[16], 取一个绝对门限(本文用 0.4), γ 大于这个阈值的参数组不要, 用这种参数选择方法, 还可以消除偶然的脉动干扰. 前者用预测误差是公式 (4) 中已有的递归参数, 后者 γ 在归一化的标准公式中没有, 而且是更有效的方法, 尤其是在下节的音调提取中很有用, 因此在归一化的 PW 算法中, 有必要导出 γ 的递归关系式, 这可由非归一化的 γ 递归表达式^[16]

$$\gamma_{P,T} = \gamma_{P-1,T-1} + \varepsilon_{P,T}^2/R_{P,T}^6 \quad (5)$$

其中 ε , R^e 是非归一化的向前预测误差及其方差. 和归一化其他参数一样, 用两步归一化方法, 实际上 (5) 式后一项可写成一步归一(用误差的平方根)的形式, 所以

$$\gamma_{P,T} = \gamma_{P-1,T-1} + \bar{\varepsilon}_{P,T}^2 \quad (6)$$

再用 $(1 - \gamma_{P-1,T-1})^{1/2}$ 归一第一步得到的 $\bar{\varepsilon}_{P,T}$, 经过整理即得

$$\gamma_{P,T} = \gamma_{P-1,T-1} + (1 - \gamma_{P-1,T-1})\bar{\varepsilon}_{P,T}^2 \quad (7)$$

程序设计时要将 (7) 式放到 (4) 式的循环中去, 没有增加多少计算的负担.

为了说明音调脉冲对估算声道参数的影响, 在此举一个合成元音分析的例子, 合成的声道由 10 阶梯型参数表达, 激励用实际分析中的逐点误差, 这样使合成元音很近似实际元音. 然后用 PW 和 LPC 两种算法分析这一合成元音, 所得三个共振峰的中心频率 (CF) 和带宽 (BW) 的实际值和估算值列于表 1, PW 估算参数包括有无激励脉冲的影响. 比较表 1 中的三种估算参数可见, 消除激励脉冲影响的估算参数最为准确, 声门激励对共振峰带宽的影响比频率影响大.

PW 算法不仅可以改进稳态元音的声道参数分析精度, 而且还可以用来提高语音过渡区参数估算的准确性, 因为 PW 的参数跟踪快, 现让我们看图 1 和图 2, 两图上面的波形是用下面的六条实线表示的反射系数 (k_1-k_6) 合成的. 即六条实线值是上面语波的实际声道参数, 图 1 上的六条虚线是用 PW 算法估算得到的; 图 2 上的六条虚线是用 Levinson 算法估算得到的, 帧长 20ms, 每次移动一点, 前面加 Hamming 窗. 两图对比说明, PW 算法参数收敛快, 在

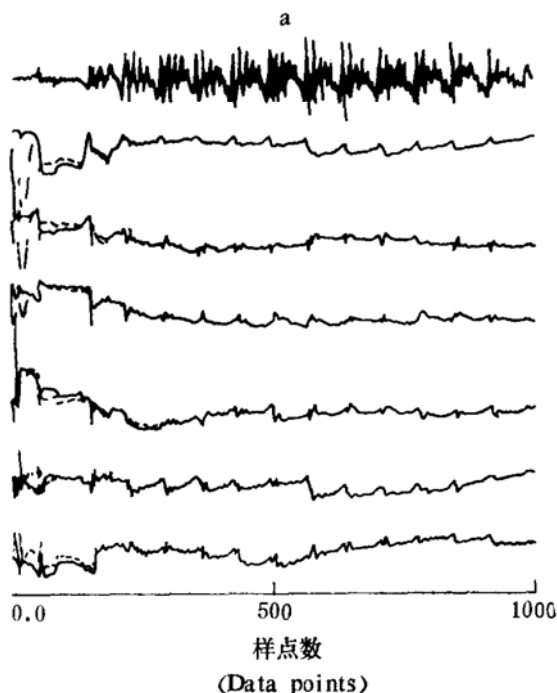


图 1 语波的实际和 PW 估算的反射系数的比较
Comparison between modeling and estimated reflection coefficients using PW
a 合成的波形 (synthesized)
——实际合成参数 (modeling)
----估算参数 (estimated)

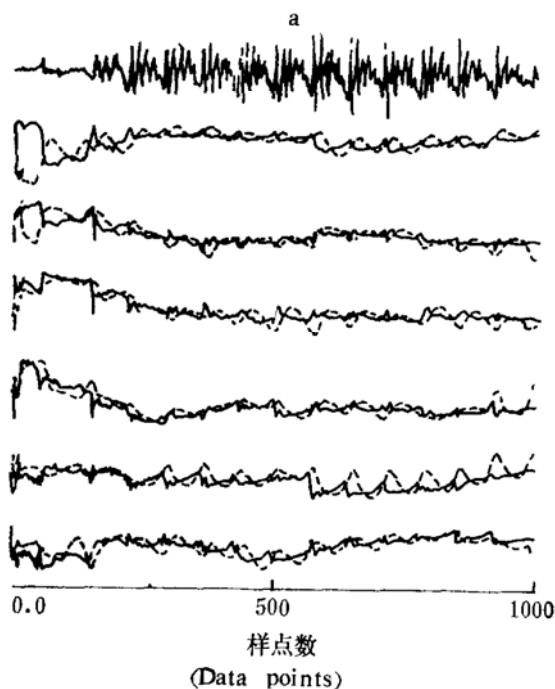


图 2 语波的实际和 Levinson 递归估算的反射系数比较
Comparison between modeling and estimated reflection coefficients using levinson
a 合成的波形 (synthesized)
——实际合成参数 (modeling)
----估算参数 (estimated)

表 1 合成元音的共振峰频率和带宽的估算

Estimated formant frequencies and bandwidths for synthesized vowel /a/ of "nan"

实 际 (Actual)	CF(Hz) 估 算 (Estimated)			实 际 (Actual)	BW(Hz) 估 算 (Estimated)		
	没有调脉冲 PW NO Pitch Pulse	有调脉冲 PW Includes Pitch Pulse	LPC		没有调脉冲 PW NO Pitch Pulse	有调脉冲 PW Includes Pitch Pulse	LPC
693.7	692.7	692.5	689.2	62.6	62.8	64.8	60.4
1483.6	1473.0	1473.2	1465.6	70.2	79.4	84.6	92.4
2976.3	2981.3	2979.1	3018.2	131.1	132.6	140.0	179.3

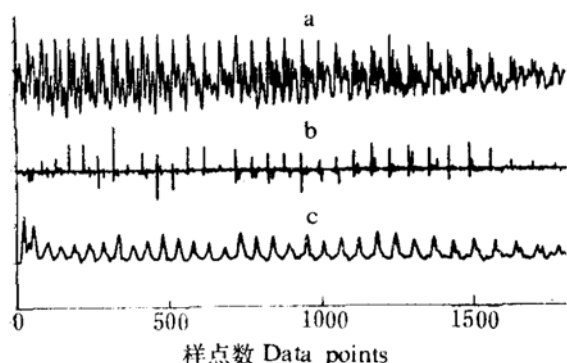
逼近真值后几乎完全和原模型参数符合,即使模型参数在声门脉冲位置有跳变,估算参数也能跟踪很好,可见这种算法有快速自适应性能. 开始部分偏离程度,各阶 k_i 不一样,渐近快的几个毫秒,慢的多至 20ms. 起始部分的偏离主要是零初值引起的,如果用某些经验值启动将会改善,本文对此不细讨论了. 对比之下,Levison 算法的参数收敛慢,不仅在起始部分,而且整个过程是波动的,其波动周期和音调周期是符合的,因为它跟踪不上音调脉冲的变化. 由此也说明用分帧方法,帧长至少要大于一个音调周期,才能得到较好的平均参数,可见,这种方法是无法跟踪语音快速过渡的,会影响语言的自然度.

三、音调提取的改进

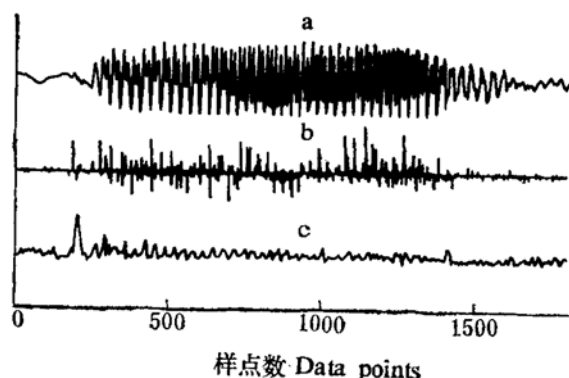
在低比特率声码器设计中,音调的准确提取和浊清音的正确判别是提高合成语言质量的重要因素,所以在这类声码器中,音调提取部分的计算量可能相当,乃至大于声道参数分析部分的计算量. 现在已有很多关于音调提取的方法,但实际上应用最广的主要有两种:一种是我们在文献[17]中介绍的平均幅差函数方法(AMDF),它可用于预测误差或直接用于低通滤波的语言信号;另一种是 B. Gold 和 L. B. Rabiner 提出的并行处理技术^[18](下称 G-R 方法),因为这两种技术运算简单,易于硬件实现,由其后者运算和存贮要求均少于前者,并且可产生音调跟踪,但用于谐波较丰富的元音和浊擦音时会有严重偏差.

本文采用 PW 梯型算法中的似然变量和 G-R 方法中的部分参量相结合,进行时间维并行处理,从而简化原并行处理技术,并提高了两种方法单独工作的性能.

我们在文献[16]中,详细地讨论了用 PW 算法中 γ 等参数去测定音调脉冲的位置,虽然在很多情况下,它对检测音调周期是很有效的,但还存在一些不足,如用于鼻音和第一共振峰低的语音(i, u)等就遇到困难. 前者因鼻道的反共振影响,使 γ 可能出现多脉冲;后者因第一共振峰落到基频范围,使音调脉冲跳变不明显,这些现象见图 3 和图 4 两例. 两图上面的波形分别是“这样”(zhè Yàng)和“多”(duō)的发音,中间是微分的 γ 脉冲,由此可见图 3 中间的脉冲除反映音调脉冲外,在/ng/的发音部分有多余的脉冲,且有些音调脉冲和相邻的比之太小;图 4 中的音调脉冲有的丢失,有的二次谐波跳变太大,显得不规则. 为了改进似然变量用于音调提取的性能,对 γ 进行低通滤波,就是两图下面的波形,可见图 3 的问题得到改善,而图

图3 γ 波形 (Waveforms)

- a. 原始语波 original (zhè yàng) b. $\Delta\gamma_{s,T} \cdot \varepsilon_{s,T}$
c. 低通滤波的 $\gamma_{s,T}$ Lowpass filtered $\gamma_{s,T}$

图4 γ 波形 (Waveforms)

- a. 原始语波 original (duo) b. $\Delta\gamma_{s,T} \cdot \varepsilon_{s,T}$
c. 低通滤波的 $\gamma_{s,T}$ Lowpass filtered $\gamma_{s,T}$

4 上的有些部位峰值平坦而弱, 音调周期难辨. 但这类语音如 (i, u, m, n) 等波形简单, 用 G-R 方法测定这些语音基频很有效.

G-R 方法是在低通滤波后的语言信号上测量六个参量如图 5, 每个参量都单独进行极值检测, 将检测到的六个参数峰值周期组成一个矩阵, 然后按帧统计峰值周期的重复频率, 挑其重复频率最高者为该帧的音调周期, 详见文献 [18]. 这种技术是模拟人观察语言波形寻找周期性的方法, 当波形简单如图 4 的“多”很容易, 但对复杂波形就较困难(如图 6). 图 6 的上面是原始语波 (fǎn) “反”, 中间是低通滤波后的波形, 采用六极点切比雪夫低通, 截止频率为 800Hz, 从这个波形上测量 G-R 方法中的参量, 定准基频周期不太容易. 图 6 下面的波形是我们用同样滤波器滤似然变量 γ , 其音调峰值周期明显. 我们结合这两种方法, 取三个参数: γ , M_1 (正峰值) 和 M_4 (负峰值), 用类似我们在文献 [16] 中介绍的脉冲周期检测办法, 自适

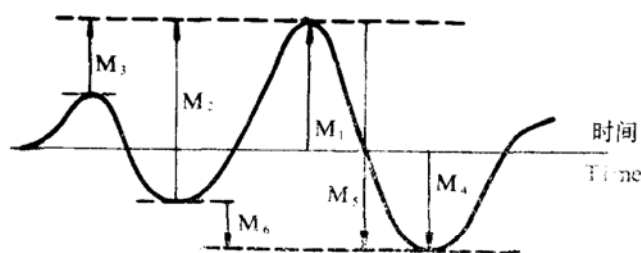


图5 G-R方法参量

Basic measurements made on the filtered speech

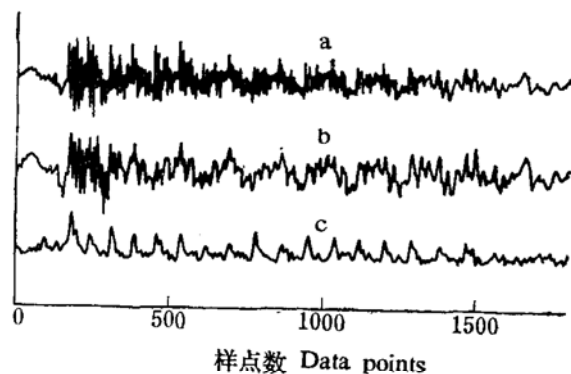


图6

- a. 原始语波 (fǎn) original
b. 低通滤波后的波形 Lowpass filtered
c. 低通滤波的 $\gamma_{s,T}$ Lowpass filtered $\gamma_{s,T}$

应变换峰值门限和搜索窗长, 除浊音起始一帧外, 每个参数保留三个周期, 与分析帧长无关, 这样避免了短帧长音调周期 AMDF 方法中的困难. 最后进行如 G-R 方法的重复频率的比较. 这样一个改进音调提取方案, 计算机上模拟表明, 效果很好, 看图 7 一个合成句子“很多苍蝇飞去吃”的音调周期估算结果, 图上句点“○”表示实际周期, “+”点是上述方法估算的音调周期,

只有在一个起点有稍大一些偏差外其余符合很好。

浊清音的判别是和音调的估算密切相关的,我们根据三个因素做浊清音判断,它们是

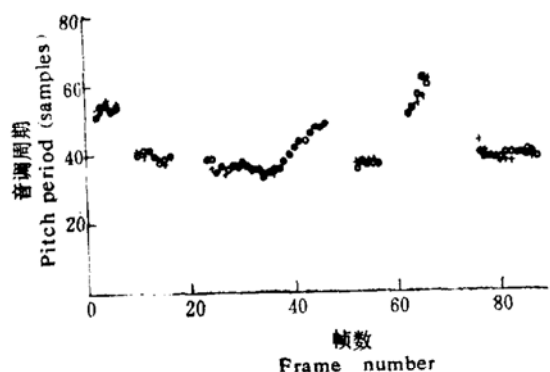


图7 合成语音的实际和估算音调周期的比较
Comparison between actual and estimated pitch period for a synthesis utterance
○ 实际值 (Actual) + 估算值 (Estimated)

上述三个参数周期数的和 N_T , 均方根能量 E 和第一个反射系数 k_1 。按照我们模拟的情况, 字长 12bit 的 A/D 转换器, 8000Hz 采样, 清音条件是 $N_T < 6$ 和 $E < 50$; 或 $N_T < 6$, $E < 150$ 及 $k_1 < 0$, 如果不合这些条件就判为浊音, 如果浊音帧三个参数的峰值周期均无重复, 则仍判回清音, 最后再进行前后帧的比较, 纠正不合理的偏差。

四、参数变帧抽样和量化

固定帧率不变帧长的声码器, 参数传递简单,

例如 2400bits/s 的声码器, 只要每秒传送 50 帧, 每

帧参数量化为 48bit, 但这种均匀抽样的参数不能很好表达语言信号的非均匀特性。PW 算法很容易实现变帧率的参数抽样, 从而适应语言的非平稳过程。但如分析帧长的变化完全随机, 这可能导致语言过渡区很高频率, 带来参数传输的困难。本文采用折衷办法, 在保持一个基本帧长 BFL 的参数抽样情况下, 变化语音过渡段的帧长, 并限制最小帧长为 LFL, 随变化快慢, 传送的参数按最小帧长的整数倍抽样。过渡区由相邻两个最小帧的反射系数的平方距离

$$D_F = \sum_{i=1}^P |\bar{k}_{i,F}^2 - \bar{k}_{i,F-1}^2| \quad (8)$$

来确定, 其中 $\bar{k}_{i,F}$ 为 F 帧的第 i 个平均反射系数, 取平均时除去音调脉冲的影响。经过很多数据试验, 我们得到一个经验阈值 $D_F = 0.7$, 只有在语音的过渡区才出现高于这个阈值的距离, 需要传送短帧参数, 其余则按基本帧长实行参数抽样。如果把最小帧长限制得很小, 或在其内再行比较, 可把浊清音的边界判得更细致一些。通过实行过渡区的变帧参数抽样, 较好地表达了语音过渡区的特性, 达到减少以致消除了合成语言局部模糊不清的感觉, 提高了语言的自然度。

为了降低传送的比特率, 除了合理的分帧抽样参数外, 量化编码是另一个重要因素。近两年来发展了一些新的参数量化方法, 例如矢量量化^[19], 分段矢量量化^[20]和矩阵量化技术^[21], 利用这些技术可把数据率压缩到 800—150bits/s。这些方法展现了语言数据压缩, 实现极低比特率的通信的美好前景, 但目前的语言质量还比较差, 使用受到限制, 还有得到它的 codebook 很费事, 所以这些方法尚待进一步研究改进。本文只讨论实用的标量量化参数技术。

已经证明, 梯型参数(反射系数)具有较好的量化特性^[22], 在它的数值范围 $(-1, 1)$ 直接线性量化, 方法最简单, 但低效能。因为反射系数的变化所引起的频谱偏差是非线性的, 所以通常采用对数面积比或反正弦的非线性变换,

$$A_i = \log \frac{1 + k_i}{1 - k_i} \quad (9)$$

$$\theta_i = \sin^{-1} k_i \quad (10)$$

然后对 A_i 或 θ_i 进行线性量化,但这种方法在 k 系数达不到或很少出现的地方,量化级有些浪费。因此我们采用分段线性量化方法^[23],它是基于 k 系数的统计分布来确定比特的分配,同时兼顾 k 变化的非线性因素。我们对汉语反射系数的统计结果如图 8,虽然这个统计的数据还不够广泛,但其分布规律和英语基本相同^[23], k_1 很接近 1, k_2 趋向 -1,其余的阶次越高越向零点集中,鉴于这种几率分布,可以把每个 k_i 分成三段:

$$\theta_1(R_0^i \leq k_i \leq R_1^i), \theta_2(R_1^i < k_i \leq R_2^i),$$

和

$$\theta_3(R_2^i < k_i \leq R_3^i),$$

每段给于不同的量化级,具体的数值分配见表 2,各系数的编码由下面公式得到,

$$C_i = \{[(R_3^{(i)} - k_i)\theta_3^{(i)} / (R_3^{(i)} - R_2^{(i)}) + 1]\},$$

$$R_2^{(i)} < k_i \leq R_3^{(i)}$$

$$C_i = \{[(R_2^{(i)} - k_i)\theta_2^{(i)} / (R_2^{(i)} - R_1^{(i)}) + \theta_3^{(i)} + 1]\}$$

$$R_1^{(i)} < k_i \leq R_2^{(i)}$$

$$C_i = \{[(R_1^{(i)} - k_i)\theta_1^{(i)} / (R_1^{(i)} - R_0^{(i)}) + \theta_3^{(i)} + \theta_2^{(i)} + 1]\} \quad (11)$$

$$R_0^{(i)} \leq k_i \leq R_1^{(i)}$$

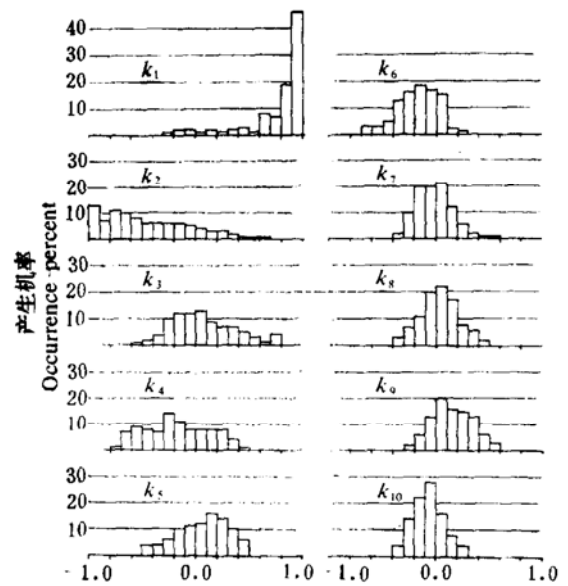


图 8 反射系数的分布

The percentage of occurrence of the reflection coefficients

表 2 分段线性量化反射系数的比特分配

The bit allocation of piecewise linear quantization for coding the reflection coefficients

Coefficients 反射系数	$R_0(k_i, \min)$	R_1	R_2	$R_3(k_i, \max)$	量化级数 Number of Quantization Levels			比特数 Number of bits
					Q_1	Q_2	Q_3	
K_1	-0.8	0.6	0.8	1.0	14	6	12	5
K_2	-1.0	-0.4	0.2	0.8	14	10	8	5
K_3	-0.6	-0.3	0.5	0.8	3	10	3	4
K_4	-0.8	-0.5	0.3	0.6	3	10	3	4
K_5	-0.6	-0.3	0.2	0.6	3	9	4	4
K_6	-0.8	-0.5	0.2	0.4	3	11	2	4
K_7	-0.5	-0.2	0.2	0.6	2	4	2	3
K_8	-0.5	-0.2	0.2	0.6	2	4	2	3
K_9	-0.4	-0.1	0.3	0.6	2	4	2	3
K_{10}	-0.4	-0.2	0.1	0.4	2	4	2	3

其中 C_i 是第 i 个反射系数 k_i 的编码, $[[\]]$ 双括号表示取整数, $\theta_i^{(i)}$ 是分配给第 i 个系数 i 段的量化级, 如果分配给 k_i 的比特数为 B_i 则满足:

$$\theta_1^{(i)} + \theta_2^{(i)} + \theta_3^{(i)} = 2^{B_i}$$

表 1 中的数据是要记忆的, 但比存贮对数表要的单元少得多, 最主要的是这种量化方法充分利用比特数的分配。

根据上述分帧抽样参数和量化方法,可以实现 2400, 1200bits/s 的低数据率的声码器。由采用频率 F_s , 所要传送的比特率 R_d 和每帧参数组的比特数 N_b , 可以确定基本帧长,

$$\text{BFL} = \frac{F_s}{R_d} \cdot N_b \quad (12)$$

最小帧长 LFL 取 BFL/2 或 BFL/4, 例如我们用 $F_s = 8000\text{Hz}$, 当 $R_d = 2400\text{bits/s}$, $N_b = 54\text{bit}$ (分配如下: 音调周期 $T = 6$, 能量 $E = 5$, 帧长指示 $L = 3$, 同步 $\text{syn} = 1$ 和 10 个反射系数用 38bit), 则 $\text{BFL} = 180$ 样点, 取 $\text{LFL} = 45$; 当 $R_d = 1200\text{bits/s}$, $N_b = 45\text{bit}$ ($T = 5$, $E = 5$, $L = 2$, $\text{syn} = 1$ 和八个反射系数用 32bit), 则 $\text{BFL} = 300$ 样点, 取 $\text{LFL} = 150$ 。两种速率的浊清音按 T 是否为零来判断, 为了在固定数据率的信道上传送, 收发两端均要有两帧比特数的缓冲器。

五、结 论

本文提出用自适应前窗型算法, 改进现有 LPC 声码器的定帧分析方法, 提高了声道参数的分析精度, 改进了应用多年的 Gold-Rabiner 音调检测技术, 用有限变帧抽样参数, 既可较好表达语音的过渡, 也利于在固定数据率信道上传输。

我们用这些改进技术, 在计算机上模拟实现了 2400, 1200bits/s 的 APC 声码器, 这里用 APC 名称是为表达用自适应梯型算法实现对 LPC 声码器的改进。虽然合成的语言质量尚未做正式评价, 但计算机模拟的很多男女语音的试听表明, 2400bits/s 的方案, 语言较为自然易懂, 能辨认熟人口音, 1200bits/s 的方案, 语言并未严重降级, 语言质量仍可为人接受, 众所周知, 现在有很多 2400bits/s 和 4800bits/s 相兼容的 LPC 声码器产品, 而兼容 1200bits/s 的产品尚未出现。我们这两种码率的方案正开始硬件试制, 由于这种算法比一般的 LPC 复杂, 计算量较大, 但归一化后的算法参数小于 1, 便于定点运算。关于如何用高速微处理器的实现, 及语言质量的正式评价会出现在未来的硬件报告中。

参 考 文 献

- [1] Atal, B. S. and Remde, J. R., "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rate", *Proc. IEEE ICASSP*, (1982), 814—817.
- [2] Noll, P., "A Comparative Study of Various Schemes for Speech Encoding", *Bell System Tech. J.*, **54** (1975), 1597—1614.
- [3] Flanagan, J. L., "Opportunities and Issues in Digitized Voice", *EASCON*, (1978), 709—712.
- [4] Sambur, M. R., "An Efficient Linear Prediction Vocoder", *Bell System Tech. J.*, **54**(1975), 1693—1723.
- [5] Atal, B. S. and Hanauer, S. L., "Speech Analysis and Synthesis by Linear Prediction of Speech Wave", *J. Acoust. Soc. Am.*, **50**(1971), 637—655.
- [6] Oppenheim, A. V. and Schaffer, R. W., *Digital Signal Processing*, (Prentice-Hall, Englewood Cliffs, NJ, 1975.)
- [7] Birch, J. N., "Low Bit Rate Voice Processors", *EASCON*, 1974, 59—63.
- [8] Feldman, J. A., Hofstetter, E. M., Malpass, M. L., "A Compact Flexible LPC Vocoder Based on a Commercial Signal Processing Microcomputer", *IEEE Trans. ASSP-31*(1983), 252—257.
- [9] Lee, D. T. L., Morf, M., Friedlander, B., "Recursive Least Squares Ladder Estimation Algorithms", *IEEE Trans. on Circuits and Systems*, **CAS-28**(1981), 467—481.
- [10] Friedlander, B., "Lattice Filter for Adaptive Processing", *Proc. of IEEE*, **70**(1982), 829—865.
- [11] Wang, J. S., Lee, D. T. L., Morf, M. and Peterson, A. M., "A Variable Bit Rate LPC Vocoder Using

- Normalized Ladder Forms", 102nd ASA Meeting, (Miami Beach, Floride, 1981), Paper R17.
- [12] Rabiner, L. R. and Schater, R. W., *Digital Processing of Speech*, (Prentice-Hall, Englewood Cliffs NJ., 1978.)
- [13] Kang, G. S. and Everett, S. S., "Improvement of the LPC Analysis", *Proc. IEEE ICASSP*, (1983), 89—92.
- [14] Steiglitz, K. and Dickinson, B., "The Use of Time-Domain Selection for Improved Linear Prediction", *IEEE Trans.*, **ASSP-25**(1977), 34—39.
- [15] Chander, S. and Lin, W. C., "Experiment Comparison Between Stationary and Nonstationary Formulations of Linear Prediction Applied to Voiced Speech Analysis", *IEEE Trans.*, **ASSP-22**(1974), 403—415.
- [16] 王俊生等, 音调脉冲准确定位的新方法, 声学学报, **8**(1983), No. 4, 236—249.
- [17] 王俊生等, 预分浊音型的 LPC 基音提取法, 声学学报, **5**(1980), No. 5, 286—290.
- [18] Gold, B. and Rabinear, L. B., "Parallel Processing Techniques for Estimating Pitch Periods of Speech in The Time Domain", *J. Acoust. Soc. Am.*, **64**(1969), 442—448.
- [19] Wong, D. Y., Juang, B. H., Gray, A. H., Jr., "An 800 Bits/sec. Vector Quantization LPC Vocoder", *IEEE Trans.*, **ASSP-30**(1982), 770—780.
- [20] Roucos, S., Schwartz, R. M., Markhoul, J., "A segment Vocoder at 150 B/s", *Proc. IEEE ICASSP*, (1983), 61—64.
- [21] Wong, D. Y., Juang, B. H., Cheng, D. Y., "Very Low Data Rate Speech Compresion With LPC Vector and Matrix Quantization", *Proc. IEEE ICASSP*, (1983), 65—68.
- [22] Viswanathan, R., Makhoul, J., "Quantization Properties of Transmission Parameters in Linear Prediction Systems", *IEEE Trans.*, **ASSP-23**(1975), 309—321.
- [23] Un, C. K. and Yang, S. C., "Piecewise Linear Quantization of LPC Reflection Coefficients", *Proc. IEEE ICASSP*, 1977, 417—420.