

线谱频率参数的分裂矢量量化¹

鲍长春 樊昌信* 王都生*

(北京工业大学电子工程系 北京 100022)

*(西安电子科技大学信息科学研究所 西安 710071)

摘要 本文简要介绍了线谱频率 (LSF) 参数的定义和性质, 给出了一种求解 LSF 参数的有效方法。在 LPC 语音编码中设计了一种新的 LSF 参数分裂矢量量化器。主观和客观评价结果表明, 以每帧 25bit 对 LSF 参数编码, 该量化器提供了透明的量化质量。

关键词 语音编码, 线性预测, 矢量量化, 线谱频率

中图分类号 TN912.3

1 引言

为了表达语音的短时谱包络信息, 线性预测编码 (Linear Predictive Coding) 参数已广泛应用于各种语音编码。它对保证语音质量和压缩速率起着直接作用。目前表达 LPC 参数的最有效方式为线谱频率 (LSF)。LSF 参数在数学角度上完全等价于其它的线性预测编码系数, 诸如, LPC 系数和 PARCOR 系数等。另外 LSF 参数有一些特别的性质, 使得它比其它的系数更富有吸引力^[1,2], 例如, 一个线谱频率参数的误差仅仅影响全极点模型中邻近这个参数对应频率处的语音谱, 而不影响其它地方。因此, LSF 参数可以根据人的听觉特性来分配量化的比特数, 对敏感频率段对应的 LSF 参数分配较多的比特数。对不敏感频率段对应的 LSF 参数分配较少的比特数。实验结果表明, 在相对低的编码速率, 使用 LSF 参数能获得高质量的合成语音, 主观性能也证明它能产生高质量的合成语音。例如, 以每帧 32 40bit 对 LSF 参数进行标量量化能合理、准确地表达语音短时谱特征。

对于低比特率语音编码, 用尽可能少的比特量化 LSF 参数是非常重要的。为了用最小数量的比特表达语音的谱包络信息, 人们在开发 LSF 参数的标量量化和矢量量化方案过程中已经进行了大量的工作。尤其是 LSF 参数的矢量量化, 是目前人们热衷研究的课题。尽管矢量量化在相同的编码比特数要比标量量化给出低的量化失真, 但矢量量化存在两方面的问题: (1) 大的码本需要极其大数量的训练数据, 训练过程化费太多的计算时间; (2) 对于矢量量化编码, 存贮和计算要求过高。基于这些问题, 本文提出了一种减少计算复杂性和存贮需求的次优 LSF 矢量量化方法, 次优的矢量量化器将 LSF 参数分裂成三组, 每组单独进行矢量量化, 每帧用 25bit 对 LSF 参数编码获得了透明的 LPC 信息量化。

2 LSF 参数的定义和性质

2.1 LSF 参数的定义

LSF 参数是线性预测参数的另一种表现形式。如果把声道等效为 $p+1$ 段声管级联而成, 则线谱参数表示声门完全开启或完全闭合状态下声管的谐振频率^[3]。

¹ 1997-01-17 收到, 1997-11-11 定稿

中国博士后科学基金和国家自然科学基金 (69602001) 资助和 ISN 国家重点实验室部分支持

在语音的线性预测分析中, 一短段语音被看成是一全极点滤波器 $H(z) = 1/A(z)$ 的输出, 这里 $A(z)$ 为预测误差滤波器, 给出如下:

$$A(z) = 1 + \alpha_1 z^{-1} + \cdots + \alpha_p z^{-p}. \quad (1)$$

(1) 式中, p 为线性预测阶数, $\{\alpha_i\}_{i=1,2,\dots,p}$ 为线性预测系数. $A(z)$ 满足下面递归关系:

$$A_n(z) = A_{n-1}(z) - k_n z^{-n} A_{n-1}(z^{-1}), \quad n = 1, 2, \dots, p, \quad (2)$$

这里 $A_0(z) = 1$, $\{k_i\}_{i=1,2,\dots,p}$ 被称为 PARCOR 系数. PARCOR 系数在数学上完全等价于 LPC 系数, 换句话说, PARCOR 系数和 LPC 系数表达了相同的谱信息. 因为 LPC 系数有较宽的动态范围以及在合成滤波器中的不稳定性, 所以它不适合于量化. 然而对于 PARCOR 系数, 如果满足 $|k_i| < 1$, 则合成滤波器是稳定的. PARCOR 系数也称为声道声管模型的反射系数.

在 (2) 式里, 当 $n = p + 1$ 时有

$$A_{p+1}(z) = A_p(z) - k_{p+1} z^{-(p+1)} A_p(z^{-1}), \quad (3)$$

$k_{p+1} = 1$ 或 $k_{p+1} = -1$. $k_{p+1} = 1$ 时相当于声道完全闭合, $k_{p+1} = -1$ 时相当于声道完全开启. 此时 $A_{p+1}(z)$ 可分解为

$$A_{p+1}(z) = \frac{1}{2} [P(z) + Q(z)], \quad (4)$$

$$P(z) = A_p(z) - z^{-(p+1)} A_p(z^{-1}), \quad k_{p+1} = 1; \quad (5)$$

$$Q(z) = A_p(z) + z^{-(p+1)} A_p(z^{-1}), \quad k_{p+1} = -1. \quad (6)$$

经过简单变换可得到

$$P(z) = (1 - z^{-1}) \prod_{i=2,4,\dots,p} (1 - 2z^{-1} \cos \omega_i + z^{-2}), \quad (7a)$$

$$Q(z) = (1 + z^{-1}) \prod_{i=1,3,\dots,p-1} (1 - 2z^{-1} \cos \omega_i + z^{-2}). \quad (7b)$$

从 (7) 式很容易看到 $e^{j\omega_i}$, $i = 1, 2, \dots, p$ 是多相式 $P(z)$ 和 $Q(z)$ 的根. 参数 $\{\omega_i\}_{i=1,2,\dots,p}$ 定义为 LSF 参数. 这里要求 p 为偶数. 着重注意的是 $\omega_0 = 0$ 和 $\omega_{p+1} = \pi$ 分别为 $P(z)$ 和 $Q(z)$ 的固定根, LSF 参数不包含这两个根. 由此可见 LSF 参数的物理意义可以解释为在声门完全闭合和开启两个特定条件下的声管谐振频率集.

2.2 LSF 参数的性质

LSF 参数有一些很有用的性质, 有利于高效量化编码和保证系统稳定. 多相式 $P(z)$ 和 $Q(z)$ 有如下两个性质: (1) $P(z)$ 和 $Q(z)$ 的所有根位于单位圆上. (2) $P(z)$ 和 $Q(z)$ 的根在单位圆上相互交错, 尤其一直满足如下有序关系:

$$0 = \omega_0 < \omega_1 < \omega_2 < \cdots < \omega_{p-1} < \omega_p < \omega_{p+1} = \pi. \quad (8)$$

如果 LSF 参数满足上面两个性质, 则 $A(z)$ 具有最小相位性质. 这样在 LSF 域内量化 LPC 信息时, 能很容易保证 LPC 合成滤波器的稳定性.

3 LSF 参数的确定

我们根据求多项式实根的二分法, 由 LPC 系数求 LSF 参数, 该方法简便实用, 介绍如下.

由线性预测器的传递函数 $A(z) = \sum_{i=0}^{10} \alpha_i z^{-i}$, $\alpha_0 = 1$, 根据 (5) 式和 (6) 式可构造下面两个多项式 ($m = 10$):

$$\begin{aligned} P(z) &= A(z) - z^{-11} A(z^{-1}) \\ &= 1 + p'_1 z^{-1} + p'_2 z^{-2} + \cdots + 0 \cdot z^{-m/2} \\ &\quad - p'_{m/2-1} z^{-(m/2+1)} - \cdots - p'_1 z^{-m} - z^{-(m+1)}, \end{aligned} \quad (9)$$

$$\begin{aligned} Q(z) &= A(z) + z^{-11} A(z^{-1}) \\ &= 1 + q'_1 z^{-1} + q'_2 z^{-2} + \cdots + q'_{m/2} z^{-m/2} \\ &\quad + q'_{m/2-1} z^{-(m/2+1)} + \cdots + q'_1 z^{-m} + z^{-(m+1)}. \end{aligned} \quad (10)$$

$Q(z)$ 是对称多项式, $P(z)$ 是非对称多项式, 通过如下变换可得两个对称多项式:

$$P'(z) = \frac{P(z)}{1-z^{-1}} = 1 + p_1 z^{-1} + \cdots + p_{m/2} z^{-m/2} + \cdots + p_1 z^{-(m-1)} + z^{-m}, \quad (11)$$

其中 $p_1 = p'_1 + 1$, $p_i = p'_{i-1}$, $i = 2, 3, \cdots, m/2$.

$$Q'(z) = \frac{Q(z)}{1+z^{-1}} = 1 + q_1 z^{-1} + \cdots + q_{m/2} z^{-m/2+1} + \cdots + q_1 z^{-(m-1)} + z^{-m}, \quad (12)$$

其中 $q_1 = q'_1 - 1$, $q_i = q'_{i-1}$, $i = 2, 3, \cdots, m/2$.

由于 $P'(z)$, $Q'(z)$ 在单位圆上有零值, 所以, 求 $P'(z)$, $Q'(z)$ 的根 $\{e^{j\omega}\}$ 的问题, 就变成了求如下两个多项式根的问题.

$$\begin{aligned} P_a(x) &= 32x^5 + 16p_1 x^4 + 8(p_2 - 5)x^3 + 4(p_3 - 4p_1)x^2 \\ &\quad + 2(p_4 - 3p_2 + 5)x + p_5 - 2p_3 + 2p_1, \end{aligned} \quad (13)$$

$$\begin{aligned} Q_a(x) &= 32x^5 + 16q_1 x^4 + 8(q_2 - 5)x^3 + 4(q_3 - 4q_1)x^2 \\ &\quad + 2(q_4 - 3q_2 + 5)x + q_5 - 2q_3 + 2q_1, \end{aligned} \quad (14)$$

这里 $x_i = \cos(\omega_i)$, $\omega_i = 2\pi f_i/f_s$, $-1 \leq x_i \leq 1$, $0 < \omega_i < \pi$, f_s 为采样率.

本文将 $[0, \pi]$ 区间化分为 314 份 (即给定 Δx), 采用二分法求方程 $P_a(x)$, $Q_a(x)$ 的根. 首先, 求 $P_a(x)$ 的 5 个根, 计算 $P_a(i\Delta x)$ 与 $P_a[(i+1)\Delta x]$, $i = 0, 1, 2, \cdots, 314$. 如果 $P_a(i\Delta x)$ 与 $P_a[(i+1)\Delta x]$ 同号, 则在 $(i\Delta x, (i+1)\Delta x)$ 区间 $P_a(x)$ 无根, 继续查找直到 $P_a(i\Delta x)$ 与 $P_a[(i+1)\Delta x]$ 异号为止. 然后再将 $(i\Delta x, (i+1)\Delta x)$ 区间两等份, 依上法查找直到有一根达到精度, 再令 $i = i + 1$, 继续找到其余的根. 求出 $P_a(x)$ 的 5 个根后, 其次根据 LSF 的交错性 $\omega_1 < \omega_2 < \cdots < \omega_{10}$, 可求得 $Q_a(x)$ 的另外 5 个根, 从而得到 10 个 LSF 参数 $lsf(0), lsf(1), \cdots, lsf(10)$ (频率值). 由于事先可将 $\cos(i\Delta x)$, $i = 0 - 314$ 制成余弦表格, 加快了迭代求解的速度.

4 LSF 参数的分裂矢量量化

由矢量量化理论知道, 将 10 个 LSF 参数分成一组, 形成 10 维欧氏空间中的一个矢量, 然后对此矢量进行整体量化, 比对每个 LSF 参数分别进行标量量化优越。但整体矢量量化的存储和计算量远大于标量量化, 在现有的硬件条件下, 难于实时实现。权衡现有的各种条件和语音编码应用系统的要求, 我们采取次优矢量量化, 即分裂矢量量化 (SVQ)。

4.1 选择失真测度

我们采用 Paliwal 和 Atal 建议的加权欧氏失真测度^[4], 定义如下:

$$d(\text{LSF}, \text{LSF}_q) = \sum_{i=1}^{10} \{C_i W_i [lsf(i) - lsf_q(i)]\}^2, \quad (15)$$

其中 LSF 和 LSF_q 分别表示未量化的和已量化的线谱频率矢量, $lsf(i)$ 和 $lsf_q(i)$ 分别为矢量 LSF 和矢量 LSF_q 的第 i 个元素, $W_i = [P(lsf(i))]^r$ 为加权因子, $r = 0.15$ 为经验常数, $P[lsf(i)]$ 为对应测试矢量的 LPC 功率谱。这种失真测度考虑了在共振峰区域的量化要好于非共振峰区域的量化, 即对应高幅度共振峰的 $lsf(i)$ 给予更多的权, 对应较小幅度共振峰的 $lsf(i)$ 给予较轻的权。另外还考虑了人耳对高频区的敏感程度低于低频区, 额外附加一权 C_i , 即低频区的权重高于高频区的权重, C_i 如下给出:

$$C_i \begin{cases} 1.0, & 1 \leq i \leq 8; \\ 0.8, & i = 9; \\ 0.4, & i = 10. \end{cases} \quad (10)$$

Paliwal 和 Atal 已经证明^[4], 使用上述失真测度要比传统的欧氏失真测度 (均方误差失真测度) 每帧节省 2 个比特以上。

4.2 SVQ 训练算法

将 10 维 $\text{LSF} = \{lsf(1), lsf(2), \dots, lsf(10)\}$ 矢量分裂为如下三组矢量:

$$\begin{aligned} \text{LSF}_1 &= \{lsf(1), lsf(2), lsf(3)\}, & \text{LSF}_2 &= \{lsf(4), lsf(5), lsf(6)\}, \\ \text{LSF}_3 &= \{lsf(7), lsf(8), lsf(9), lsf(10)\}. \end{aligned}$$

给定 LSF 训练数据和一个体积为 N 、维数为 K 的初始码本, 以类似于常用的 GLA (Generalized Lloyd Algorithm) 算法^[5] 迭代设计上述三组矢量对应的三个码本。每次训练有如下两个关键迭代 (Lloyd 迭代): (1) 用最近相邻准则 (NNR) 围绕码矢量群聚训练矢量。(2) 依据质心准则^[5] 将群聚质心替代旧的码矢量。GLA 算法流程图如图 1。

4.3 保证系统稳定的后处理

采用上述 SVQ 方法有时会出现系统不稳定现象, 这是由于分裂矢量量化造成的。在分裂矢量量化过程中, 虽然每一个矢量内各维能保持有序性, 但每个矢量的衔接处, 例如, LSF_1 和 LSF_2 相连接处的 $lsf(3)$ 和 $lsf(4)$ 、 LSF_2 和 LSF_3 相连接处的 $lsf(6)$ 和 $lsf(7)$ 有可能破坏有序性, 导致系统不稳定。

为保证线谱频率的有序性和系统稳定, 减少计算复杂度, 本文采取了如下两条后处理措施:

(1) 先量化码本 LSF_2 ，之后对 LSF_1 码本以 $lsf(3)$ 为基准进行由小到大的排序，对 LSF_3 码本以 $lsf(7)$ 为基准进行由小到大的排序。

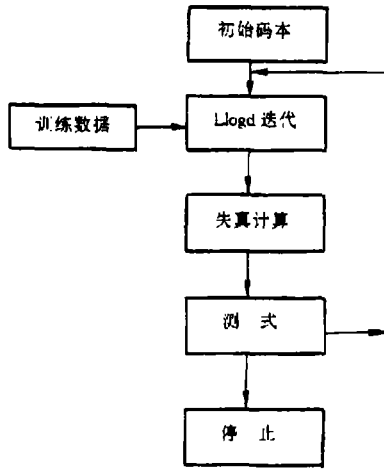


图 1 GLA 算法流程图

(2) 假定 LSF_1 、 LSF_2 和 LSF_3 三个码本的大小分别为 N_1 、 N_2 和 N_3 。在量化 LSF_1 之前，先找到 LSF_1 码本中 $lsf(3) < lsf(4)$ 的码本序号 $Index_1$ ，仅在码本的序号 $1 \sim Index_1$ 间搜索与 $LSF_1 = \{lsf(1), lsf(2), lsf(3)\}$ 最邻近的码矢量；在量化 LSF_3 之前，先找到 LSF_3 码本中 $lsf(7) > lsf(6)$ 的码本序号 $Index_3$ ，仅在码本的序号 $Index_3 \sim N_3$ 间搜索与 $LSF_3 = \{lsf(7), lsf(8), lsf(9), lsf(10)\}$ 最邻近的码矢量。

5 实验及其结论

5.1 实验背景

用于训练 SVQ 码本的语音数据库为汉语语音，长度约 35min。这些矢量来源于 30 余名不同年龄段的男女讲话记录。最初的 2000s 用于训练，最后的 200s 用于测试。语音库经 100~3400Hz 带通滤波，以 8kHz 采样率数字化，每个样值用 16 bit 线性码表示。一个基于自相关方法的 10 阶 LPC 用 25 ms 的 Hamming 窗完成，帧速率为 40 Hz。这样我们粗略地得到 11×10^4 个 LSF 矢量，其中 10^5 个 LSF 矢量用于训练， 10^4 个 LSF 矢量用于测试。

5.2 客观评价

为了测量 LPC 量化性能，这里我们用谱失真 (SD) 作为客观评价标准。对于第 i 帧的 SD_i 定义为^[6]

$$SD_i(\text{dB}) = Db \sqrt{\frac{4}{L} \sum_{n=1}^{L/2} \{\log[S^{(i)}(\omega_n)] - \log[S_q^{(i)}(\omega_n)]\}^2}, \quad (17)$$

其中 \log 为自然对数， $\omega_n = 2\pi n/L$ ， $Db = 10/\log_{10}$ ， L 为 $S^{(i)}(\omega_n)$ 和 $S_q^{(i)}(\omega_n)$ 的计算点数， $S^{(i)}(\omega_n)$ 和 $S_q^{(i)}(\omega_n)$ 分别为第 i 帧语音未量化的和量化的 LPC 功率谱，给出如下：

$$S^{(i)}(\omega_n) = 1/|A^{(i)}(\exp(-j\omega_n))|^2, \quad (18)$$

$$S_q^{(i)}(\omega_n) = 1/|A_q^{(i)}(\exp(-j\omega_n))|^2, \quad (19)$$

这里 $A^{(i)}(\exp(-j\omega_n))$ 和 $A_q^{(i)}(\exp(-j\omega_n))$ 分别为第 i 帧语音未量化的和量化的 LPC 多项式。

LPC 信息的透明量化要求 LPC 信息量化后在编码语音中不引入任何附加的可感觉到的失真。为了达到透明量化，需要平均谱失真接近 1dB，平均谱失真在 2~4dB 范围内的比例

为 2% 左右, 平均谱失真在 $> 4\text{dB}$ 时, 所占比例趋于 0^[4]。

实验中 LSF_1 、 LSF_2 和 LSF_3 三个码本的比特分配分别为 8bit、8bit 和 9 bit, 取 $L = 512$, 在 $0\sim 4000\text{Hz}$ 范围内计算 SD。平均谱失真计算结果表明: 每帧 25 bit 的 LSF 分裂矢量量化器平均谱失真为 1.07dB, 平均谱失真在 $2\sim 4\text{dB}$ 范围内的比例为 2.97%, 平均谱失真在 $> 4\text{dB}$ 的比例为 0。图 2 给出了 25bit/frame LSF 分裂矢量量化的谱失真统计直方图。客观测试结果说明, 本文设计的 25bit/frame SVQ 量化器达到了透明量化效果。

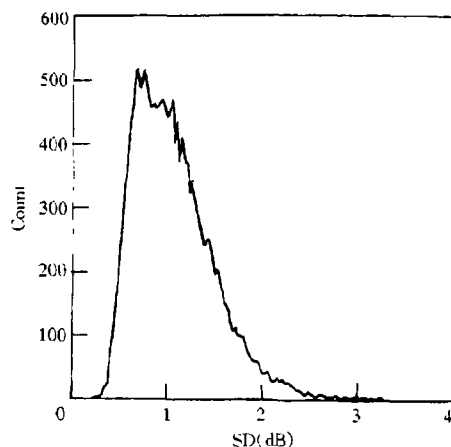


图 2 25 bit 线谱频率 SVQ 的谱失真统计直方图

5.3 主观评价

这里我们给出了一个非正式听力测试结果, 用我们自己开发的 2400bps ICWI(Improved Characteristic Waveform Interpolation) 编码器产生两组编码语音: 一组 (A) 用未量化的 LPC 参数, 另一组 (B) 用 25bit/frame LSF 分裂矢量量化后得到的 LPC 参数。

主观质量评价由 8 个听者对 10 个汉语语音句子 (5 男 5 女讲话) 进行 A-B 比较测试, 8 个听者对 10 个句子的两组编码语音, 没有指出听觉上的差别, 这时我们可以说, 25bit/frame SVQ 量化 LPC 信息具有透明质量。

参 考 文 献

- [1] Soong F S, Juang B H. Line Spectrum Pair(LSP) and speech data compression. Proc ICASSP, San Diego. California: 1984, 1.10.1-1.10.4.
- [2] Kang F K, Franssen L J. Application of line spectrum pairs to low-bit-rate speech encoders. Proc. ICASSP, Tampa, Florida: 1985, 244-247.
- [3] Sugamura N, Farvardin N. Quantizer design in LSP speech analysis- synthesis, IEEE J. of Selected Areas in Communication, 1988, 6(2): 432-440.
- [4] Paliwal K K, Atal B S. Efficient vector quantization of LPC parameters at 24bits/frame, IEEE Trans. on Speech Audio Processing, 1993, 1(1): 3-14.
- [5] Gersho A, Gray R M. Vector Quantization and Signal Compression. Kluwer Academic Publishers, 1992.
- [6] Miguel A, Ferrer-Ballester, Anibal R. Vidal F. Efficient adaptive vector quantization of LPC parameters, IEEE Trans. on Speech Audio Processing, 1995, 3(4): 3-14.

SPLIT VECTOR QUANTIZATION OF LINE SPECTRAL FREQUENCY PARAMETERS

Bao Changchun Fan Changxin* Wang Dusheng*

(Dept. of Electronic Engineering, Beijing Polytechnic University, Beijing 100022)

**(Information Science Institute, Xidian University, Xi'an 710071)*

Abstract In this paper, the definition and properties of Line Spectral Frequency(LSF) parameters are introduced briefly. An efficient solution method of obtaining LSF parameters is presented. A new Split Vector Quantizer(SVQ) of LSF parameters is designed in LPC speech coding. The objective and subjective evaluations show that the proposed scheme offers transparent quantization quality with 25 bit/frame.

Key words Speech coding, Linear prediction, Vector quantization, Line Spectral Frequencies

鲍长春: 男, 1965年生, 博士, 副教授, 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。

樊昌信: 男, 1931年生, 教授, 博士生导师, IEEE会士, IEE会士, 中国电子学会会士, 中国通信学会会士, ISN国家重点实验室学术委员会主任。研究领域为语音和图象信号处理, B-ISDN等。

王都生: 男, 副教授, 研究方向为语音编码。