

汉语普通话辅音的线性预测谱分析

陈培新 李潜生
(南京自动化研究所)(南京工学院)

辅音的谱结构在语音分析、合成、识别以及人工智能等方面有十分重要的意义。由于线性预测理论在语音技术中的应用日益深入,因此本文采用线性预测的方法对汉语普通话辅音进行谱分析。整个谱分析工作在通用的微机系统上进行。本文重点是解决辅音谱分析中几个特别困难的问题,如在取样方面,如何采取措施以保证原始数据的准确性;在谱分析时,对信号帧长、预测阶数如何选择。最后,通过实验,对全部汉语普通话辅音音素确定了在给定取样率情况下的信号帧长及预测阶数,作出了它们的线性预测谱。并与语图仪实测谱图比较,结果令人满意。

一、线性预测谱分析 方法简介

设原始辅音数据序列为 S_n , 该序列可以表征为一个 P 阶的格型模型, 如图 1。

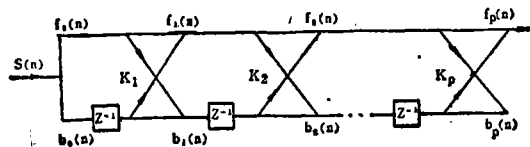


图1 格型模型

其中 f 表示前向预测误差序列, b 表示后向预测误差序列。由参数组 $\{k_i\}$, 通过 Durbin 关系式可得到该序列的线性预测系数组 $\{a_i\}$ 。这时辅音线性预测谱估值为

$$\hat{P}(\omega) = \frac{G^2}{\left| 1 + \sum_{i=1}^p a_i e^{-j\omega i} \right|^2}$$

其中 G 为匹配常数。 p 为预测阶数。详细推导可参考文献[2]、[3]。

用格型模型方法计算线性预测谱有如下优点:

1. 物理概念清晰, 它直接对应于语言产

生的声管模型。

2. 计算速度较快。模型的阶数可有一定的选择, 从而可以方便地控制模型的精度。

3. 计算时利用信号的协方差函数, 可提高精度。

二、实验与讨论

1. 辅音信号的取样

由于辅音一般具有能量小、持续时间短、高频分量多这样一些特点, 因此准确地获取辅音原始数据是一个重要而困难的问题。谱分析结果的好坏首先决定于数据是否准确。

本文中实验用的辅音由广播电视播音员(男女各一名)在消声室内录音, 以增加信噪比, 减小外界噪声对分析结果的影响。由于线性预测谱对噪声敏感, 因此取样时, 为了提高信号噪声比, 各个辅音采用了不同的放音电平。所以下文中各辅音谱的绝对 dB 数不代表它们之间的相对强弱。

每个辅音的发音方式为元音—辅音—元音。这样的组合发音比单独发辅音更加接近自然发音, 同时也为辅音信号的取样带来方

便。由于汉语声调的变化发生在元音部分，可以认为对辅音没有影响，因此，辅音前面的元音都发轻声，后面的元音都发第一声。

取样系统以微机作为控制中心对辅音信号进行取样，简单过程如下：模拟信号（元一辅一元）经低通滤波器滤去10KHz以上的频率分量，由A/D转换为数字信号，送入计算机检查。当计算机发现第一个元音段开始时，则进入取样准备状态。当发现信号进入辅音段时，便开始取样。样点先暂存在内存中。经判别后，再将内存中的辅音信号存入磁盘以备谱分析时用。实际上，在送数据到磁盘之前，先由D/A将取样信号还原为模拟信号，由录放机放音。当听觉满意时，再送数据到磁盘。

上述取样过程中，元音段与辅音段的判别是根据两音能量的大小进行的。从实验中知道，发音过程中，当从元音过渡到辅音时，口腔声道的形状发生变化。这时将出现无声区的背景噪声。由于这个背景噪声的能量与某些辅音的能量大小相当，因此取样时容易误认为是辅音信号。在这种情况下，单从能量大小来判别元音和辅音是不够的。更完善的方法应同时从能量及跨零率两方面考虑。但是，跨零率的考虑又会使算法复杂化，增加运算量。因此为了使取样方法简单易行，本文仍采用能量作为判别元音与辅音的主要依据，辅以下述补偿手段：1)以辅音后接元音的起点作为标准，参考文献[1]中给出的辅音音长，向前取一定长度。2)根据听觉判断是否取在辅音段。3)观察信号波形，根据对辅音时间波形的先验知识判断是否为辅音。

2. 线性预测谱的阶数选择

线性预测的阶数选择是谱分析的一个重要问题。由于线性预测谱反映信号能量的包络，因此，过低的阶数会使谱过于平滑，不能反映出信号的谱峰；而过高的阶数又会使谱过于细致，甚至出现虚假的谱峰，也不能很好地反映出信号的谱特性。

目前，关于阶数的选择问题还没有统一的结论。多靠重复实验来确定“最佳”阶数。一般认为，阶数与信号帧长、取样率等有关。帧长长、取样率高，则阶数相应高些；反之，则阶数相应低些。

本文根据赤驰(Akaike)的“最终预测误差”判阶准则，对汉语普通话辅音作了阶数自适应实验，认为在20KHz取样率的情况下，阶数取16比较合适。尽管各个辅音的阶数可以不同，而且某一特定辅音各帧间的阶数也可以不同，但作为一个统计平均的选择，16阶能较好地反映出各个辅音的主要谱峰，既不过份平滑，也不过份细致。

图2是男声“X”的线性预测谱。取样率为20KHz，帧长为10ms。图中看出，当阶数从16增加到24时，5KHz处的单峰逐渐变为双峰，即谱变得过份细致。

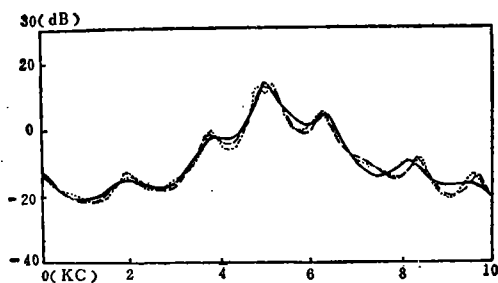


图2 男声“X”的线性预测谱
——16阶 ——20阶 ...24阶

3. 辅音的分帧线性预测谱

谱分析时，信号必须具有平稳性。辅音信号中，最短的音只有几个毫秒，而最长的音可有100多毫秒。因此在分析辅音时，信号往往分成若干帧，使每一帧近似的平稳。一般说，帧长越长，则谱估值的偏差越小，但同时时间特性就越差，因此要在偏差与时间特性之间进行折衷。较长辅音的谱随时间变化的速率不快，帧长可以长些以减小谱估值的偏差；较短的辅音，则帧长取得短些以改善谱估值的时间特性。帧长的选择应以不破坏信号的短时平稳性为准。下表是本文对汉语普

通话辅音进行谱分析时, 信号帧长及阶数的选择。(注: 由于“m、n、l、r、ng”具有某些元音性质, 所以分析时阶数取为20)

辅音	b	p	m	f	d	t	n	l	g	k	h
帧长(ms)	2.5	5	15	10	2.5	5	15	15	2.5	5	10
阶数	16	16	20	16	16	16	20	20	16	16	16

辅音	j	q	x	zh	ch	sh	r	z	c	s	ng
帧长(ms)	10	10	10	5	10	10	15	5	10	10	15
阶数	16	16	16	16	16	16	20	16	16	16	20

除了“b”、“d”、“g”这三个辅音, 帧长的选择应尽量短外, 其它辅音辅音的帧长选择有一定的灵活性。本文一般取为10毫秒。

图3是女声“j”在帧长分别为5、10、15ms情况下的线性预测谱取样率为20KHz, 阶数为16。显然, 帧的长度对谱的影响不大。

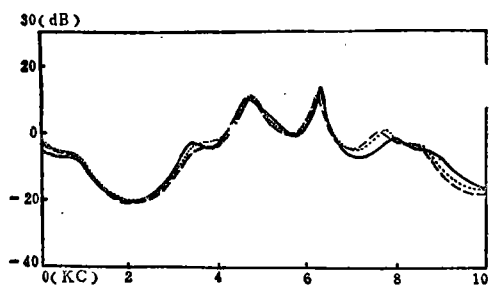


图3 女声“j”的线性预测谱
——帧长5ms ——帧长10ms...帧长15ms

对较长的辅音来说, 它们分帧后的各帧谱基本上是相同的。图4取自女声“sh”的四帧谱图。取样率为20KHz, 阶数为16, 帧长为10毫秒。图第一帧与第四帧时间上间隔20毫秒。图第四帧与第五帧时间上相邻。图第五帧与第八帧时间上间隔20毫秒。显然, 这四帧谱比较接近。特别是图第一帧与第八帧, 它们在时间上间隔60毫秒, 但仍有足够的相似性。

4. 辅音的线性预测谱与语图谱

作者将全部辅音的线性预测谱与语图谱作了比较。结果基本一致。因限于篇幅, 仅举两例。

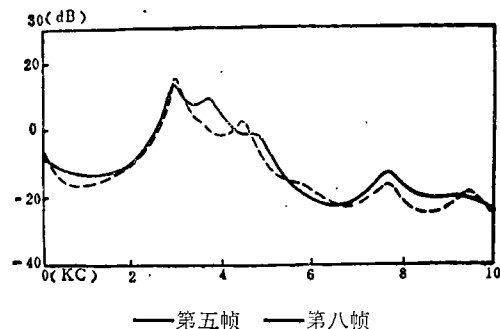
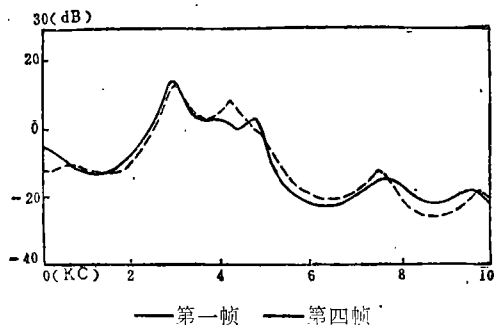


图4 女声“sh”的线性预测谱

比较男声“g”的线性预测谱和相应的语图谱。可以看出, 辅音“g”在1.8KHz、4.6KHz处附近分别有两个谱峰。

比较女声“ch”的线性预测谱和相应的语图谱。可以看出, 辅音“ch”在3.1KHz附近有明显的谱峰, 而在6.0KHz附近处有不明显的谱峰。

从辅音的线性预测谱图与相应的语图比较中, 可看出它们在能量的分布上基本一致。特别在一些能量较强的频率上, 两种谱图吻合得很好。而在较弱能量的频率处, 两种谱图有一些差别。

5. 有关结论

以上对汉语普通话辅音的线性预测谱分析作了一些讨论。限于篇幅, 只对部分辅音作了举例说明。综观谱分析的全部过程, 兹归纳出以下几点结论:

- 1) 在20KHz取样率的情况下, 辅音线性预测谱的阶数取为16较为合适。
- 2) 除了“b”、“d”、“g”三个短音外, 其它辅音的帧长选择有一定的灵活性, 但一般应

(下转第13页)

据输入信号动态大于60分贝, 高端截止频率可根据需要选定。该电路用于浅海波动信号记录分析, 得到满意的结果, 并适合于作有关调频记录器的部件。

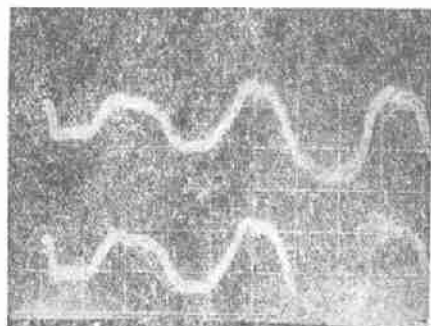


图5 调制输入信号 $V_i(t)$ (上)和解调输出信号 $V_o(t)$ 波形图

图5是调制输入信号 $V_i(t)$ (上图)和解调输出信号 $V_o(t)$ (下图)的波形图。可见, 两者一致性是很好的。

参考文献

- [1] [英]B. S 埃文登, D. R 斯通: 地震勘探仪器——仪器的性能和测试, P. 93~99, 石油化学工业出版社, 1978年7月。
- [2] 展安祥: 模拟磁记录原理 P112~119 国防工业出版社, 1983年6月
- [3] B. & K. Tape Recorders Types 7005 and 7006.
- [4] 王滋庭: 模拟磁带记录仪的应用, 电子技术, 1980年6月 P27~29.
- [5] INTERSIL: IC MASTER P3518~3520. 1984年。
- [6] 卢炳宝、郝鸿安: 单片精密 函数波形发生器, 电子科学技术, 1985年1月。

(上接第25页)

限制在15ms以内。

- 3) 较长辅音各帧间的谱有足够的相似性。
- 4) 汉语普通话辅音的线性预测谱与语图谱基本一致。

另外, 观察所作出的辅音谱, 还有这样一些结果:

- 1) 某些辅音的主要谱峰有一定的分布规律。
“p, t, k, b, d, g, h”的主要谱峰在2 KHz以下, “j, q, x”的主要谱峰在4 KHz以上, “zh, ch, sh”的主要谱峰在4 KHz以下。
- 2) “m, n, l, r, ng”的时间波形具有明显的周期性, 因而具有元音性质。在1 KHz以下, 有一较强的谱峰。
- 3) 男女声的谱有一定的相似性。其中以“b, t, k, zh, ch, sh”最为明显。

6. 使用仪器

- 1) 紫金II微型计算机, 64K内存。
- 2) A/D转换器, 12bit; D/A转换器, 8bit。
- 3) 低通滤波器, 截止频率10 KHz, 10.6 KHz处衰减不小于40 dB。
- 4) 日立牌录放机。

参考文献

- [1] 齐士铃、张家骥“汉语普通话辅音音长分析”声学学报, 第七卷, 第一期, 1982年1月
- [2] Rabiner L. R. and Schafer, R. W. “Digital Processing of Speech Signals” Prentice-Hall, Inc. 1978
- [3] Markel J. D. and Gray, A. H. Jr., “Linear Prediction of Speech”, Springer-verlag, New York, 1976
- [4] Kay S. M. and Marple, S. L. Jr., “Spectrum Analysis-A Modern Perspective”, Proc. IEEE, Vol. 69, No. 11, Nov. 1981