YUYINXINHAO
XIANXINGYUCE

语音信号
线性预测

〔美〕J.D.马卡尔 A.H. 格雷，乔 著

中国铁道

语音信号线性预测

〔美〕J.D.马卡尔　A.H.格雷，乔　著

娄乃英　等译

莫福源　校

中国铁道出版社

1984年·北京

内 容 简 介

本书系统地阐述了语音信号线性预测的基本理论及其应用。

全书共十一章，内容包括：语音信号产生模型，线性预测模型，基本方程式，解法和性质，声管模型，语音合成模型，频谱分析，自动共振峰轨迹估算，音频估算，分析中的计算问题和声码器等。书中附有各种算法的FORTRAN程序和实例。

本书论述严谨，理论联系实际，可供从事语音数字信号处理工作的科技人员参考，也可供高等院校有关专业高年级学生、研究生和教师参考。

Linear Prediction of Speech
J.D.Markel A.H. Gray, Jr.
Springer-Verlag Berlin Heidelberg 1976

语 音 信 号 线 性 预 测
〔美〕J.D.马卡尔 A.H.格雷，亦 著
娄乃英 等译 莫福源 校
中国铁道出版社出版
责任编辑 黄成士 封面设计 翟 达
新华书店北京发行所发行
各 地 新 华 书 店 经 售
中国铁道出版社印刷厂印
开本：850×1168毫米 1/32 印张：10.625 字数：227 千
1 9 8 7年4月 第1版 第1次印刷
印数：0001—2,500册 定价：3.10元

译者的话

近二十年来，语音信号数字处理的一个新领域——线性预测在国外得到了迅速发展。J.D.马卡尔和A.H.格雷，乔合著的《语音信号线性预测》一书是这方面发展的一个系统、深入的总结。

书中除了论述线性预测方法在语音信号处理中的基本理论和在语音合成、基音提取、谱分析、共振峰跟踪、声码器等方面的应用外，还指出了进一步研究的方向和前景。书中引用并列出了该领域中绝大部分的重要文献，为读者深入理解和研究提供了线索。本书是研究语音信号处理的科研工作者、教师的重要参考书，也是该领域研究生和大学高年级学生的良好教材。

参加本书翻译工作的有：丁晓明（第1、2章），王大理（第3章），娄乃英（前言，第4、5、6、7章），马秀莲（第8、9章），林碧琴（第10、11章）。全书译文由莫福源校对。

翻译过程中，我们对原书中的个别错误作了订正。但由于译者的水平有限，译文中难免会有不妥之处，恳望读者批评指正。

译　者

1984年8月

前　言

语音信号处理的一个新领域——线性预测，在过去的十年里，已经发展起来。正如所有的科学研究那样，在内容的合理安排和术语方面总是不够理想，所以研究结果一直没有完全发表。

1974年，我们利用业余时间和周末开始整理有关语音信号线性预测方面的文献，并且力图做到在内容和术语方面的统一。1975年11月完成了这项工作。

如果要用两个词来表达本书的目的，那就是统一和深入。我们用了很大的努力来阐明各种线性预测公式及其解的相互关系，并且逐步深入，例如以统一的一种模式、术语来表示声管模型和合成滤波器的结构。用这种方法叙述本书的内容，以适应公式的推导和理论的证明，同时本书还给出了FORTRAN程序和设计实例。采用这种方法我们期望能使本书内容适应的范围更宽，并且使读者能感到兴趣。

本书内容反映了当前许多重要课题的特殊方法。本书内容正如参考文献上所看到的那样，基本上包括了由B.S.阿塔尔博士，F.T.伊塔库拉博士，J.麦克豪尔博士，S.塞托博士和H.威基特博士所进行的研究成果。我们十分高兴地看到在语音信号处理方面线性预测技术已经产生了重大的影响，以及在这个领域中进行深入研究的重要性。

J.D.马卡尔

A.H.格雷，乔

目　录

1. 绪　论

为了定量描述语音处理所涉及到的某些因素，虽然已经假定了许多不同的模型，但是可以肯定，目前还没有发现一种可以详细描述人类语声中已观察到的全部特征的模型（由于它的复杂性，也许我们不可能找到一个理想的模型）。建立模型的基本准则是：要寻求一种可用以表达一定物理状态下的数学关系，要使这种关系不仅具有最大的精确度，而且还要最简单。方特(Fant) 1960年提出的线性语音信号产生模型是模拟语声特征最成功的模型之一。下面把这种模型简称为语音信号产生模型。

近年来，线性预测的数学技术已用于模拟语声特征的处理。这种线性预测模型与语音信号产生模型有关，因为语音信号产生模型中主要特征的参数，很容易由线性数学方法所得到

本章将给出线性预测模型以及它与语音信号产生模型之间的关系。建立语音信号产生模型的出发点是来自语言生理学（而语言生理学又出自许多有关更好理解语言的不同领域）。然而，本章所讨论的问题，仅局限于语言的基本物理概念，重点阐明语言的声学特征。有关语言的生理特性和声学特性的详细讨论可参见方特(Fant)[39]、弗拉纳根(Flanagan)[41]、彼得森(Peterson)和舒普(Shoup)[123]、[124]的文章。

1.1　基本物理概念

语声波形是一种声压波，它是由图1.1所示的人体发声结构生理运动所产生的。空气由肺部排进声道中，然后在声带间受压。当产生浊音时，例如发“*eve*”中/*i*/*的音，空气由肺部压

---

* 符号/*i*/表示语音音素，是基本语音单位。

入，由嘴唇呼出，从而引起声带的开启与闭合。开闭的速率依赖于声道中的空气压力和声带的生理控制。这种控制包括声带的长度、厚度、反张力的变化。若张力愈高，则所感觉的音调就愈高，从声学上讲也就是在浊音区所测到的基音频率愈高。声带间的开口定义为声门，在声门区内，下声门的空气压力及其随时间的变化决定了压入声道的声门气流的体积速度（亦称声门体积速度波）。声门开闭的速度，在声学测量上近似为所观察到声压波周期的倒数。声门体积速度波决定了输入到声道的声能或激励函数。

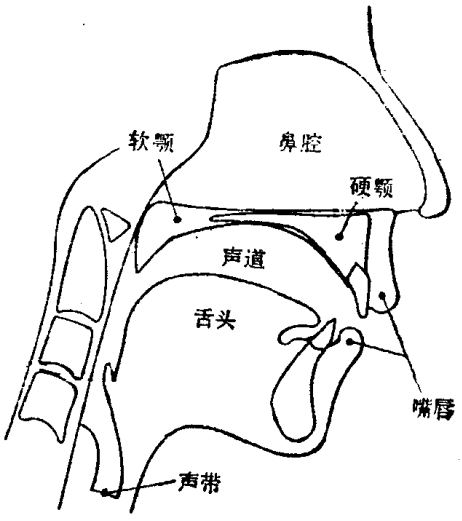

图1.1 主要发音器官剖面图

声道是一根从声门延伸到嘴唇的非均匀截面的声管，它的形状变化是时间的函数，其中能引起时变的最重要的组成部分是嘴唇、腭、舌头和软腭。例如，嘴唇开闭的截面积可以从0变化到20cm²，即嘴唇完全闭合到上、下腭及嘴唇全部张开。在不产生鼻音期间，软腭完全堵塞了通向鼻腔的声道。而在产生鼻音/*n*/、/*m*/、/ŋ/时，例如，在发*run*，*rum*和*rung*时，就要利用鼻腔，把它作为声音传输一根附加的声管。例如在发（*fish*中的清音/*f*/）清音时，首先使声带张开，迫使气流通过声带，把上齿放在下唇上，然后，发音器官产生一次收缩，就发出了*fish*的/*f*/音。当声带不仅有收缩，而且还有振动时，就发出了浊摩擦音，例如*van*中的/*v*/音。对于象*pop*中的/*p*/这样的爆破音，则首先在口腔中建立起气压，然后突然释放，形成爆破音。

1.2 语音信号波形举例

为了在时域和频域上举例说明语音信号的声学含义，我们以短语“线性预测”为例，首先把它输入到话筒，并录在磁带上，然后再进行分析。图1.2（*A*）给出了“*linear prediction*”短语的语音波形（标尺为时间和幅度）。由于在取得这个波形之前，要先进行低通滤波，所以磁带录音输出后的信号带宽仅为5kHz，因此在通过模/数转换到计算机系统时，用10kHz的取样频率。为了显示这种语音取样信号，需要采用线性内插的方法得到它的连续图形。只有在显示很长的一段语音波形时，才能从包络线上大致地看到语音信号随时间变化的特征。

图1.2（*B*）和（*C*）分别从所示短语中截出长度为25.6ms的清音和浊音信号。可以看出，图1.2（*B*）中的波形几乎是一个周期信号。其中两个主峰值之间的距离表示声门振动的音调周期*P*。图1.2（*C*）的波形中辨认不出音调周期来，这是因为单词*prediction*中的清音/ʃ/是由气流直接通过收缩的舌头和牙齿后面形成的湍流所产生的。

图1.2（*D*）给出了图1.2（*B*）中的一个音调周期的波形。在本例中，这个衰减振荡的频率（即振荡周期的倒数）取决于频域中声道主要谐振点的位置，而其衰减速率可近似为这个谐振点的带宽。

使用Kay公司语图仪来描述语音信号频域中的波形称为语图（亦称语音波纹或语谱图），如图1.2（*E*）和（*F*）所示，图中分别使用了宽带和窄带滤波器，语图表示语音能量（作为一种参数）在连续频率轴上随时间变化的规律。图1.2（*A*）中，时间坐标已调整为与语图的时间坐标相一致，在浊音期间，黑色的条带表示共振点的位置，它是一个时间的函数，此外浊音区内，垂直的条纹对应于信号的起始时间或每个音调周期的起点。在清音期间，黑色区域表示能量集中的主要区域。对于宽带语图仪来说，在时域中是可以分辨出音调周期的，但在频域中，需进行大量

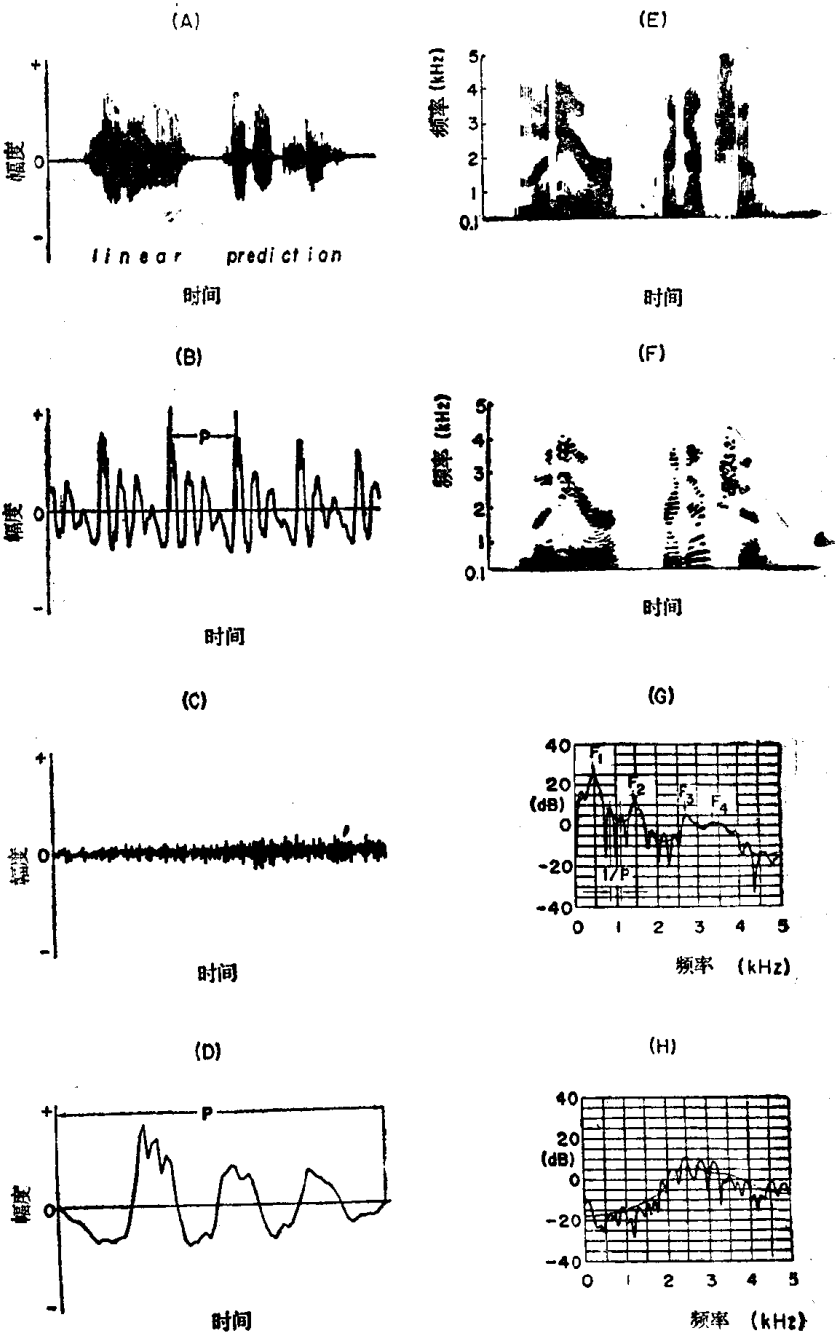

图1.2 短语 *"linear prediction"* 的时域和频域表示

的平均或使频域分辨力下降；而对于窄带语图仪来说，它的频率分辨力是以牺牲时域分辨力的代价获得的。用窄带滤波可得到浊音的谐波结构，也就是说，细细的水平波纹表示了音调频率的谐波分量，在清音期间，根本没有明显的谐波结构。

最后一个重要的图例是根据对数频谱画出的，它给出了随频率变化的对数幅度曲线（以dB为单位）。图 1.2（G）和（H）分别显示了浊音/ I /和清音/ʃ/中部所对应的对数谱，此外在每一个对数幅度谱上都叠加了一个平滑的包络线。图 1.2（G）中包络线上标记了F_1、F_2、F_3和F_4几个主要峰值的位置，确定了共振峰位置的估算值（至此，完全可以把共振峰频率认为是声道的谐振频率）。

从这些图例中清楚地看到语音波形有非常复杂的结构。在语音处理中力图要分析出更多的东西，可先假设一个模型，然后在各种条件下检验它。我们希望模型既是线性的又是时不变的，这是最理想的模型。但遗憾的是语音信号是一个连续的时变过程，人类语音机理，不能精确地满足这两种性质。此外，声门和声道相耦合，还形成语音信号的非线性特性[40]。然而，作出一些合理的假设，在较短的时间间隔内表示语音信号时，可以采用线性时不变模型。

图1.2（B）和（C）中所采用的时间间隔为数十毫秒，这样浊音信号几乎成为周期信号，即每个周期都与前一个周期非常相似。而清音信号，则没有音调周期。在浊音信号的频域描述中，每隔$1/P$个频率单位，出现一个循环周期，而对于清音信号，则显示出随机特性，它们的主要能量集中在3kHz附近。

引入的语音信号产生模型，首先要从实际频谱中分离出平滑的谱包络，然后再对模型的每个组成部分赋于生理学上的意义。下面将会看到，图1.2(G）和（H）中的平滑谱包络结构，很容易由语音信号的线性预测法求得（事实上，图中所给出的平滑谱包络就是对图 1.2（B）和（C）中所示的时域波形，进行线性预测分析计算所得）。

1.3 语音分析与合成模型

线性语音信号产生模型是五十年代末由方特提出的[39]。图1.3给出了这种语音信号产生模型。

声门体积速度波$u_G(t)$可模拟为一个两极点低通滤波器的输出，滤波器截频在100Hz左右。对于浊音来说，滤波器的输入$e(t)$是一个周期为P的脉冲串，而对于清音来说，输入将是一个具有平坦频谱的随机噪声。应该指出，该模型仅描述了一种特殊的情况，它没有考虑输入信号中含有类似浊摩擦音的成分或者接入模拟鼻音的滤波器分支。其声道模型是由一组二极点谐振器级联而成的全极点模型。每一个谐振频率称为共振峰，对应于滤波器的中心频率和带宽。

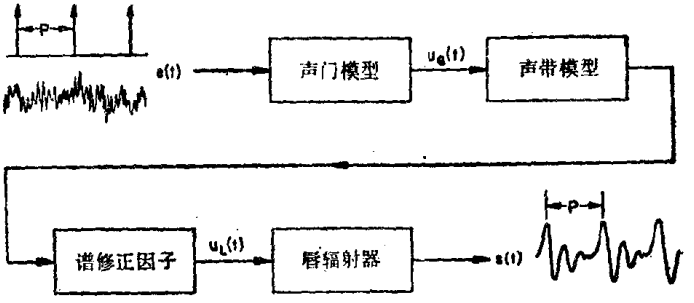

图1.3 线性语音信号产生模型

一种更精确的模型是由无限个谐振器所构成，这些谐振器对低频的作用是提高它们的谐级。因此，只对系统低频部分即话音从20Hz到几kHz最重要的频段进行精确模拟时，其频谱形状可以由高阶极点修正因子来描述。这高阶极点修正因子提供了所有高阶极点的效果，而和高阶极点的全部特性无关。由唇辐射模型的唇部体积速度波$u_L(t)$可变换为离开唇部某一距离的声压波（亦称语音波$s(t)$）。这些模型的假设、数学证明和详细推导由方特[39]和弗拉纳根[41]给出。同时，弗拉纳根还给出了一些关于

语音辐射的详细实验数据，以便确定这个语音信号产生模型。

为便于计算机的使用，以上可以用 z 变换符号来表示[69]，公式如下：

$$S(z)=E(z)G(z)V(z)L(z) \tag{1.1}$$

式中

$$S(z)\leftrightarrow s(nT)=s(t)\mid_{t=nT} \tag{1.2}$$

上式定义了连续波形$s(t)$与通过每隔T时间间隔取样后的离散信号$s(nT)$与z变换 $s(z)$ 之间的对应关系。作为一种简化表示法，通常假设归一的取样间隔，$T=1$，这样$s(n)$表示了$s(t)$的取样形式。一种类似的对应关系也用于其它变量。如声门模型的激励函数$E(z)\leftrightarrow e(n)$，这是一组按基音周期$P=IT$（其中I为正整数）为间隔的单位取样脉冲串，即

$$E(z)=\sigma\sum_{n=0}^{\infty}(z^{-I})^n=\frac{\sigma}{1-z^{-I}} \tag{1.3}$$

$|z|>1$，声门模型$G(z)$具有下列形式

$$G(z)=\frac{1}{(1-e^{-cT}z^{-1})^2} \tag{1.4}$$

唇辐射模型$L(z)$为：

$$L(z)=1-z^{-1} \tag{1.5}$$

这些都是简化了的，不需语音信号实际特性的假设。

由K个共振峰组成的表达式为

$$V(z)=\frac{1}{\prod_{i=1}^{K}\left[1-2e^{-c_iT}\cos(b_iT)z^{-1}+e^{-2c_iT}z^{-2}\right]} \tag{1.6}$$

其中第i个共振峰频率和带宽可分别由$F_i=b_i/2\pi$和$B_i=c_i/2\pi$，计算得到。拉宾纳 (Rabiner)[129] 指出，在数字表达式中，高阶极点的修正项可以消去。

应该指出 $z=0$ 时的零点对全极点或全零点滤波器定义是没