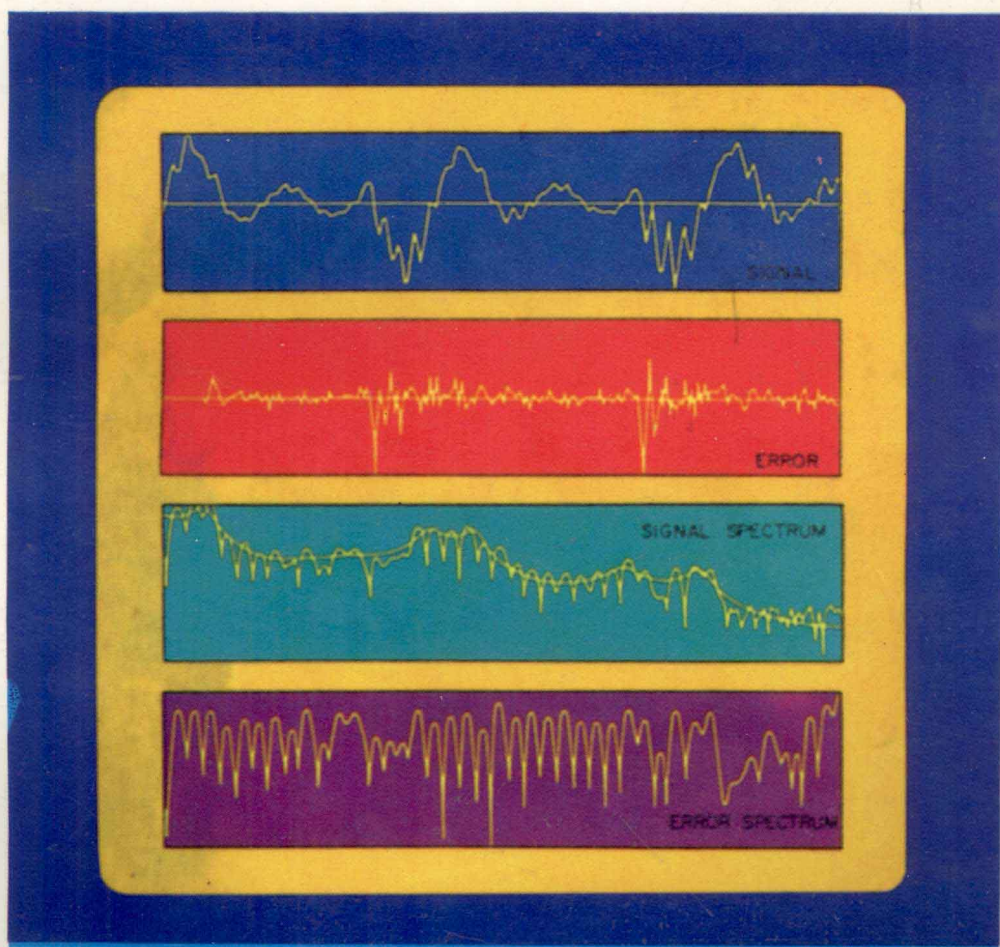


计算机实用技术系列丛书(二)

语音信号数字处理技术

谢依兰 编著



学苑出版社

计算机实用技术系列丛书(二)

语音信号数字处理入门

谢依兰	编著
天 奥	改编
熊可宜	审校

学苑出版社

1993

(京)新登字 151 号

内 容 简 介

本手册共分九章,首先向您介绍了语音的产生及其数学模式,接着对语音的波形进行微分量化,然后对语音线性预估进行分析,对语音的合成和人机语音辨认进行阐释,最后对语音与计算机网络的结合、语音与类神经网络的结合进行科学的分析。

欲购本书的用户,请直接与北京 8721 信箱联系,邮编:100080,电话:2562329。

版 权 声 明

本书繁体字中文版原书名为《语音讯号数位处理》,由松岗电脑图书资料股份有限公司出版,版权归松岗公司所有。本书简体字的中文版版权由松岗公司授予北京希望电脑公司和学苑出版社独家出版、发行。未经出版者书面许可,本书的任何部分不得以任何手段复制或传播。

计算机实用技术系列丛书(二)

语音信号数字处理入门

编 著:谢依兰
改 编:天 奥
审 校:熊可宜
责任编辑:甄国宪
出版发行:学苑出版社 邮政编码:100036
社 址:北京市海淀区万寿路西街 11 号
排 版:北京天奥科技公司激光照排中心
印 刷:双育印刷厂
开 本:787×1092 1/16
印 张:8.5 字数:196 千字
印 数:1~5000 册
版 次:1993 年 12 月北京第 1 版第 1 次
ISBN7-5077-0760-1/TP·7
本册定价:12.00 元

学苑版图书印、装错误可随时退换

目 录

第一章 语音的介绍	1
1.1 语音的通讯	1
1.2 信号处理	1
1.3 数字信号处理	2
1.4 数字语音处理	3
1.5 应用范围	4
第二章 语音的产生及其数学模式	5
2.1 发音器官	5
2.2 声学语音学(Acoustic Phonetics)	9
2.3 语音的声学理论	16
2.4 无损管模式	23
2.5 语音信号数字模式	27
第三章 时域中的语音处理	30
3.1 时域关连的语音处理	30
3.2 短时段能量与平均音量	32
3.3 短时段平均过零点率	34
3.4 基频周期的估计	36
3.5 短时段自相关函数	37
第四章 语音波形的表示法与编码	39
4.1 语音信号的取样及其统计模式	39
4.2 量化	41
4.3 适应量化	44
4.4 微分量化	48
4.5 Delta 调度	51
4.6 微分 PCM (DPCM)	54
第五章 语音线性预估编码	57
5.1 线性预估分析的原理	57
5.2 LPC 公式的解	63
5.3 LPC 分析公式的比较	66
5.4 预估误差	66
5.5 线性预估的频域意义	76
5.6 线性预估与无损管模式的关系	80
5.7 LPC 系数与其他语音参数的关系	81
5.8 线性预估参数语音合成	82
5.9 LPC 参数的应用	83

第六章 语音合成自动化	87
6.1 语音合成的概念	87
6.2 激发源	89
6.3 发音腔道的共振模拟	91
6.4 语音技术	91
6.5 送气音与擦音的产生	95
6.6 线性预估式的语音合成	95
第七章 人机数字语音处理	97
7.1 语者辨认与语音辨认	97
7.2 相似性判断法	106
7.3 LPC 距离测量	107
7.4 马可夫隐藏层模型(Hidden Markov Model)	109
第八章 人机语音诊断	115
8.1 “听”的介绍	115
8.2 耳部结构	115
8.3 听觉神经感应	116
8.4 听觉神经影响	118
8.5 声谱图	118
8.6 仪器分析与评估	120
8.7 语音的突想	123
第九章 语音辨认与类神经网络	125
9.1 使用的科学基础	125
9.2 语音辨认的大小程度	126
9.3 语音辨认的技术	126
9.4 语音特质分析	126
9.5 类神经网络	129
9.6 类神经网络与语音辨认商品的应用	130
9.7 技术的评估	131
9.8 内神经与类神经	132

第一章 语音的介绍

当今的语音逐渐成为 21 世纪的宠儿。使用激光的凝聚或扫描可得知远处或隔间的谈话内容。由玻璃的震动声可得知邻窗的谈话内容。甚至人类已开始研究狗的发音转换/模拟成为人类可知的内容。可笑的是,以后不知道将会有多少人因此而失业(因为狗可以代替人类)。当然运用语音的成功例子也越来越多,最早又最成功的例子算是 1983 年语音和金融结合的 NTC(日本电信电话公司)。因此,语音与计算机网络的结合乃是必然的趋势。

语音的目的在于通讯,通讯中的语音信息是经过原始大自然景象分歧而来,此景象直接传达的媒介就是声波(ACOUSTIC WAVEFORM)。语音的构成是由固定的音素(PHONEME)数目组合成的。音律的产生大约是 10 个音素/秒,而每一种音素代表不同的发音状态,所以也用各种不同的符号表示。

1.1 语音的通讯

在语音的通讯系统中,语音可以运用各种方式传送存储及处理。一般而言,在任何系统中,有两点必须是要注意的:(1)要能保留语音信号的信息,(2)语音信号所表示的格式是为了便于传送或存储或有灵活的修正以产生语音信号而不会影响信息内容的递减。

语音信号的表示必须能够由说话者或机器自动地分析或抽取信息内容。在设计与建立过程中,信号处理的数据方法扮演着一个基本的角色。本书所探讨的语音信号(不包含信息),每秒约是 500 至 100 万个位。

1.2 信号处理

现在正式进入我们的主题。一般而言,信号的操作与处理问题可由图 1.1 说明。开始的方格中,信息来源是来自于说话者。测量与观察是声学的音波工具。而信号处理首先包括基本模型的信号例子,然后是一些较高层次转换的应用,以便能发出信号到另一个更为便利的格式。处理的最后一个步骤则是信息的输出与利用,这个步骤可由听者或机器来执行。依照例中的方法,系统的功能会自动地从已给予的说话者中使用时间独立的语音信号声谱表示。一个可能的信号转换是沿着整个句子找出平均的声谱,对于每一个可能的说话者可比较平均声谱,以便储存在声谱图上。然后根据一个相似测量的声谱,选择说话者的身份。例子中的“信息”信号也就是代表说话者的身份。

因此,语音信号处理通常包括两项工作:第一,它是一项工具,是为获得波形或参数格式的表达法;第二,信号处理有助于将信号转换到特制或特殊的格式中,但更为恰当的是可应用到特定的实例中。

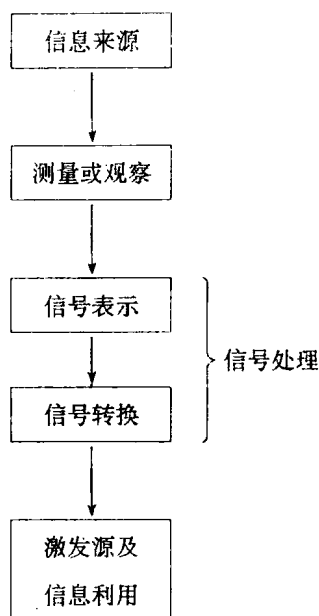


图 1.1 一般信号处理与操作

信号的处理首先是根据已定的模型,然后以高级转换求取更简易的格式。激发源的信息利用可由人类的听觉或自动化的机器执行。最重要的是,信号处理包括两项工作:(1)语音信号的表示方法可用波形或参数格式,(2)语音信号处理的辅助功能可处理信号的转换格式,以配合特定的应用。

波形表示法,顾名思义就是模拟语音信号经过取样(取样理论,在固定的时间段区间内,显现样品的频宽信号)与量化处理后能将“波形”保留。参数表示法是直接表示语音信号输出的模型。而参数表示法的激发参数可当成是语音的来源,声道的响应参数可当成是个别的语音。

1.3 数字信号处理

在学习数字语音处理之前,应先了解电子数字信号处理(DFT,Z-transform,Digital Filters),取样原理(Sampling Theory),语音学(发音方法,注音符/音标,声谱图)等。

现在先让我们探讨一下何谓数字信号?电子信号表示有两种:模拟信号(Analog Signal)与数字信号(Digital Signal)。目前语音学上使用数字表示方法居多,因其可转换成数据,提供较方便且可迅速了解的信息。

数字信号处理关系到信号分歧的获取与理论、设计及设定处理分歧的数字。而数字信号处理的主要目的是验证这些模拟信号处理。所以,我们可以问为什么对于语音网络中的数字信号处理技术需要被显示出来呢?其中,有很多理由是可以联想到的。最直接的事实则是可使用数字的技术设定信号处理的功能。理论上,也可运用到时间的分割上,其余则较不适合。

数字信号处理技术首先可应用到语音处理的问题上,就如同复杂的模拟系统。当应用模拟的数字模拟时,计算机则要求一段长时间与短时间的处理。例如,在一小时内,仅要求几秒的语音处理。实际上,目前的实验室所设立的语音处理系统观点是为运用计算机系统做特殊的数字

硬件模拟。另外经过模拟系统,可加强数据处理技术的理论发展和并存发展。

在语音网络的系统中,有许多原因是说明为何使用数据处理的技术。例如,如果使用适当的编码,语音数字的格式则可经过干扰频道转换,而且,如果语音信号是在数字的方格内,则它可证明格式数据的转换。因此,通讯网络能使用在语音与数据的转换上,而无需区分解码的特殊例子。

在有声信号的传输上因需要安全性,经过模拟系统则能产生更大的帮助。近来,将数字技术应用在语音处理上也非常广范。

1.4 数字语音处理

数据的格式是语音信号所考虑的基本点。“取样理论”是个很著名的理论,在一定的时间内,显示样品的频宽信号。图 1.2 是语音的两种表示法:(1)波形表示法——LDM、PCM、DPCM、ADM、ADPCM、DVSD;(2)参数表示法——激励与声道参数。波形表示法,顾名思义就是模拟语音信号经过取样与处理之后能将“波形”保留。参数表示法也就是表示语音信号输出的模型。而参数表示法的激励参数可当成是语音的来源,声道的响应参数则可当成是个别的语音。

数字语音处理的主要工作有:(1)语音的数字传输与储存,(2)语音合成,(3)语者辨认,(4)语音辨认,(5)音质的改善。总体上而言,处理语音沟通的问题应着重在于:(1)语音所产生的数字信号格式,(2)建立一个智慧型处理技术,(3)区分应用等级。当然,数字信号处理主要也是运用取样原理完成。

基本的语音传输率由图 1.2 可显示波形传输的频率范围与参数表示法的频率范围:

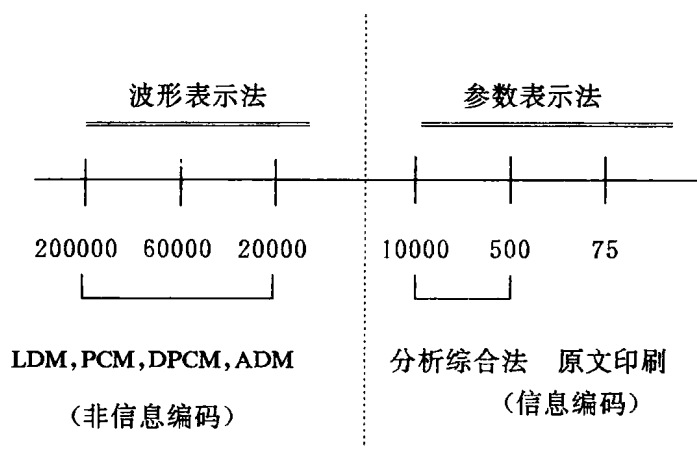


图 1.2 语音信号表示频率范围

虚线点是 11115,000 位/秒,左边表示高频,右边表示低频。参数表示的基本信息传输是 75 位/秒,波形表示的最高频是 200,000 位/秒,两者之比约为 1:3000 倍。当然其间的差异也因成本、品质、可行性等因素影响频率的高低运用。

1.5 应用范围

语音的应用例子很多,下面是一些早期的应用例子:

(1) 数字传输与语音存储

早期的语音处理是应用有声编码器,1930年由 Homer Dudley 发明。其目的是为减低语音信号传输所需的频宽。且使所有的数字电话工厂为了提升对数据语音系统的一个可能低的位比率、继续低成本的端子应用。

(2) 语音综合系统

计算机的有声响系统,基本上全部是数字化,由人的键入字到终端机或键盘上,再由声音回答所需的信息。1876年,美国的贝尔发明电话,至今一般的按键电话仪器无需额外特殊的仪器,可自动的传达信息的功能。所以,语音是综合所有系统在学习所有的人类语音后合成的处理,同样扮演着一个基础学科的角色。

(3) 说话者验证与证明系统

说话者验证与证明系统的技术包括在一堆可能类似的说话者中,验证说话者的身份。基本的系统是应用到信息的控制处理或限制领域,而且自动地转换各种信号的状况。

(4) 语音辨认系统

语音辨认是最普遍的格式,此格式是声学波形转换成对等的数字信息。而语音辨认系统的潜在应用有很多种,例如,语音操作打字机及语音计算机网络。

(5) 残疾辅助功能

语音信号处理的应用是为制定特殊的格式,以便配合残疾者所需。例如,预处理记录带倒回器提供盲人“读者”,经过已给予的语音器材有机会读到所想要的数据。还能提供感测帮助,做视觉显示,做聋人的教学。

(6) 信号品质的加强

信号品质加强是为改进说话的品质,如除去语音的响应,除去杂音或可使潜水者在氮氧混合的水气中,仍能重新存储说话的记录。

第二章 语音的产生及其数学模式

语音的产生过程是由人体的发音器官形成。人体内可产生若干不同的声音。此章主要是介绍声学语音学所形成的英文发音的语音学。根本上,声道的声音产生过程,类比的传输线形,单一声道系统内的产生行为稳定状态等。都为我们所考虑的范围。研究人类沟通应用的管道,几乎都是靠语言学的规则加上信号处理的问题。

2.1 发音器官

图 2.1 是人体发音器官的 X 射线侧面图。发音的管道开始于声带之间或喉门,结束于嘴唇。声道包括咽腔(从会厌到嘴部)及嘴巴或口腔。男性的平均声道长度是 17 公分,加上声道所经过区域面积——舌、唇、上下齿及小舌,整个部分约 20 平方公分。综合喉头以上的发音器官部份包括:双唇、上下齿、齿龈、硬腭、软腭、小舌、口腔、鼻腔、喉腔、气管、舌、声门 12 种(如图 2.2)。

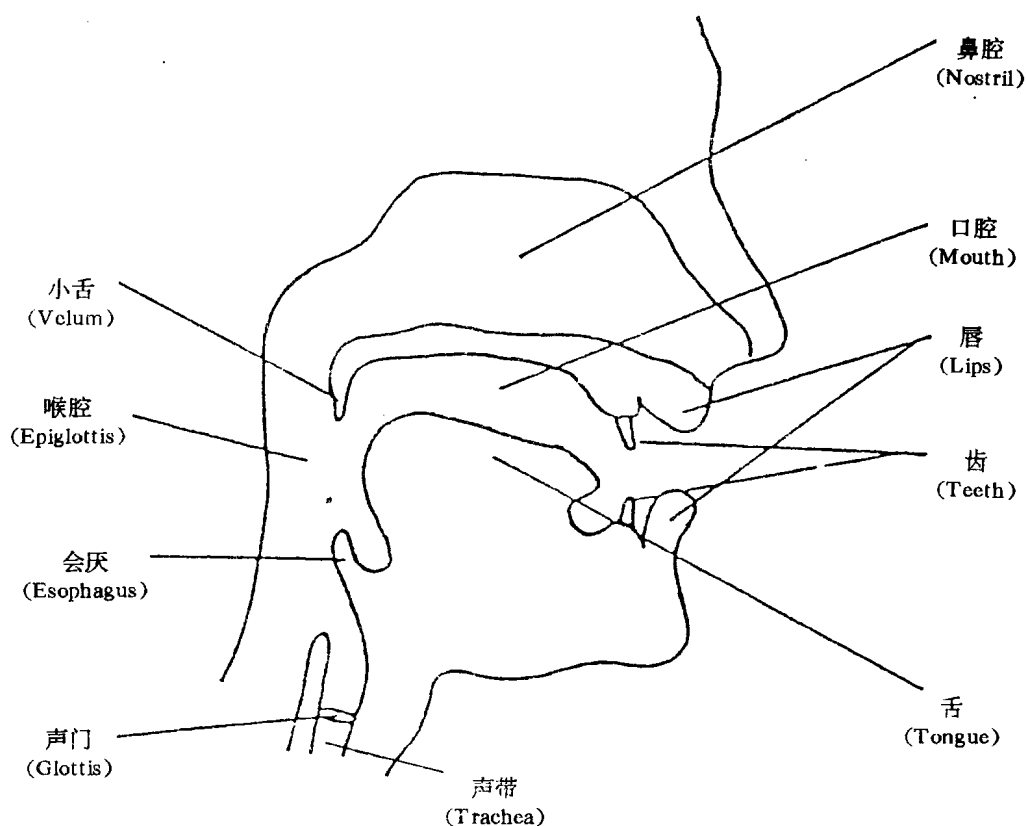


图 2.1 人体发音主要器官

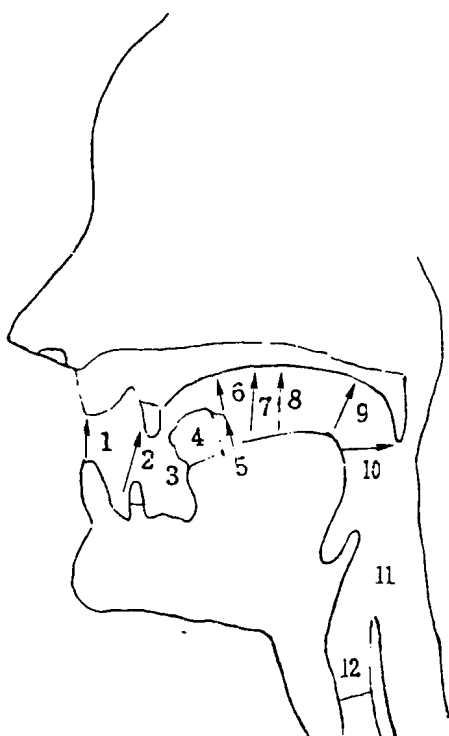


图 2.2 十二种发音部位

图 2.3 则可帮助我们了解实际上发音器官的数学模型,由这些物理的系统导出数据处理公式。一个完整的图形包括——肺、支气管与气管——语音产生的热能来源。简而言之,声学波是从肺脏所扩散出的音波,并通过空气的流动而产生震荡不平的语音。图 2.4 是语音波的例子,此图(a)是男生发出的声音“Should We Cha(se)”,图(b)是相对应的声谱图。

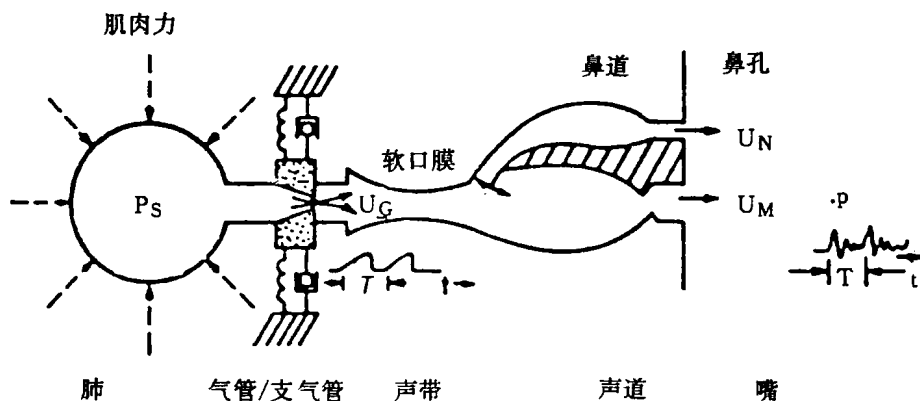


图 2.3 发音器官数学模型

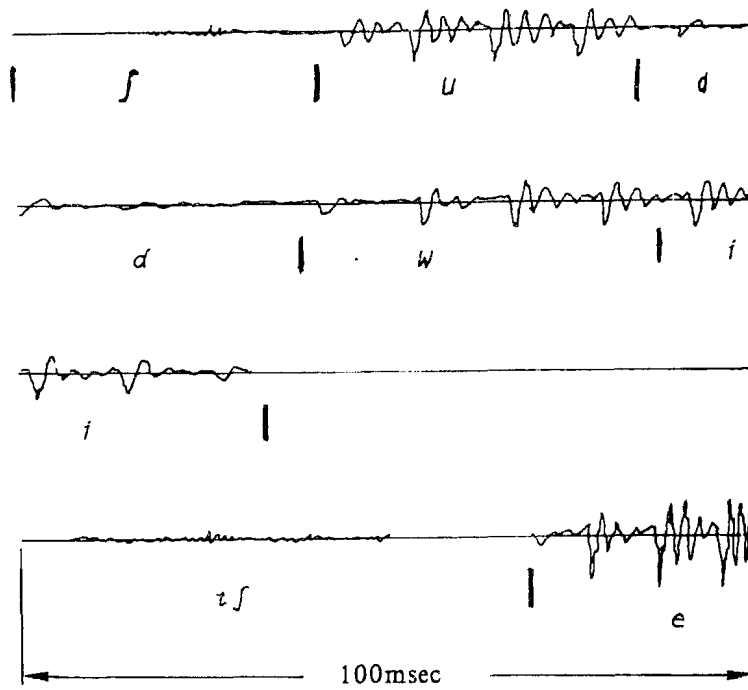


图 2.4 (a)波形发音"should we cha(se)"

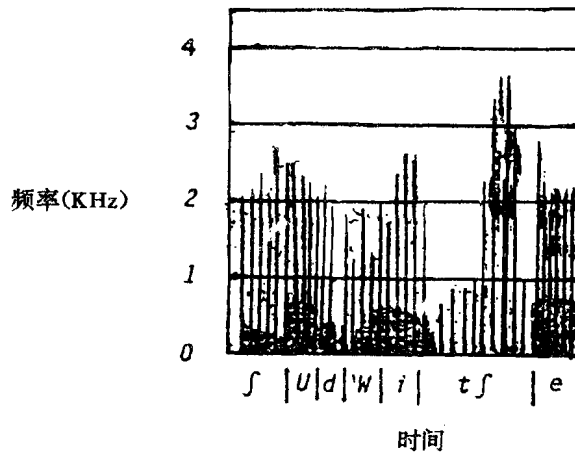


图 2.4 (b)波形发音所对应的声谱图

一般波纹形成的部份,可由舌与软腭语音活动构成机械式的语音产出。

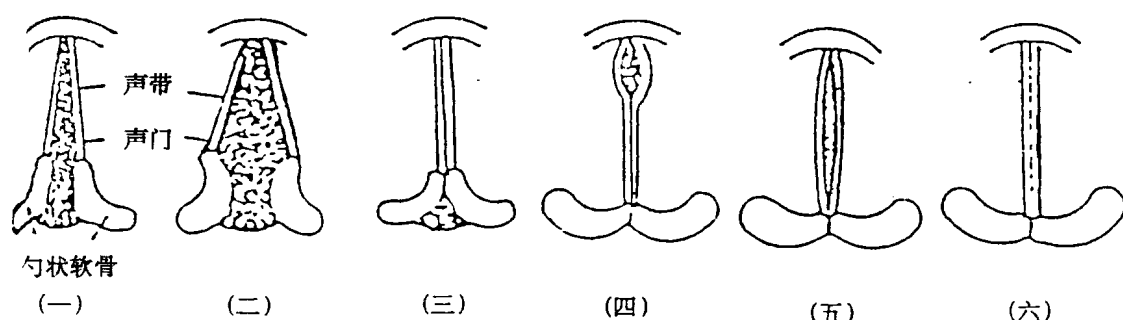


图 2.5 声带发音比较图

现说明语音活动部份:

(1) 舌——可分为舌尖、舌叶、舌前、舌后、舌根等部位,各部分与其它发音器官接近或接触,因而构成发音时的气流装饰。

(2) 软腭——舌后与软腭接触时,可发腭音;软腭提升,口腔与鼻腔分开,发口腔音;软腭下垂,空气由鼻腔溢出,发鼻音。

从图 2.1 中可看出声带是位于会厌下半部,而声带之间的空间称为声门(glottis),如图 2.2 的第 12 种发音部位。而声门的开口状况大小,也决定波动峰层的高低与密度。声门的开合状态与发音类型有六种:(如图 2.5)

(一) 平常呼吸状态——声带不振动。

(二) 气息热重时——声带分得更开。

(三) 耳语时——两条声带几乎完全合起来。

(四) 发出叭嘎声——下降语调结束时,状软骨紧闭,声带只在前端振动。

(五) 平常发声时——两条声带近乎于平行,空气挤出,产生振动。

(六) 提举重物时——声带全部紧闭。

(1) 双唇(bilabial)

(2) 唇齿(labiodental)

(3) 齿间(interdental)

(4) 齿龈(alveolar)

(5) 卷舌(retroflex)

(6) 腭龈(palatoalveolar)

(7) 龈腭(alveolopalatal)

(8) 硬腭(palatal)

(9) 软腭(velar)

(10) 小舌(uvular)

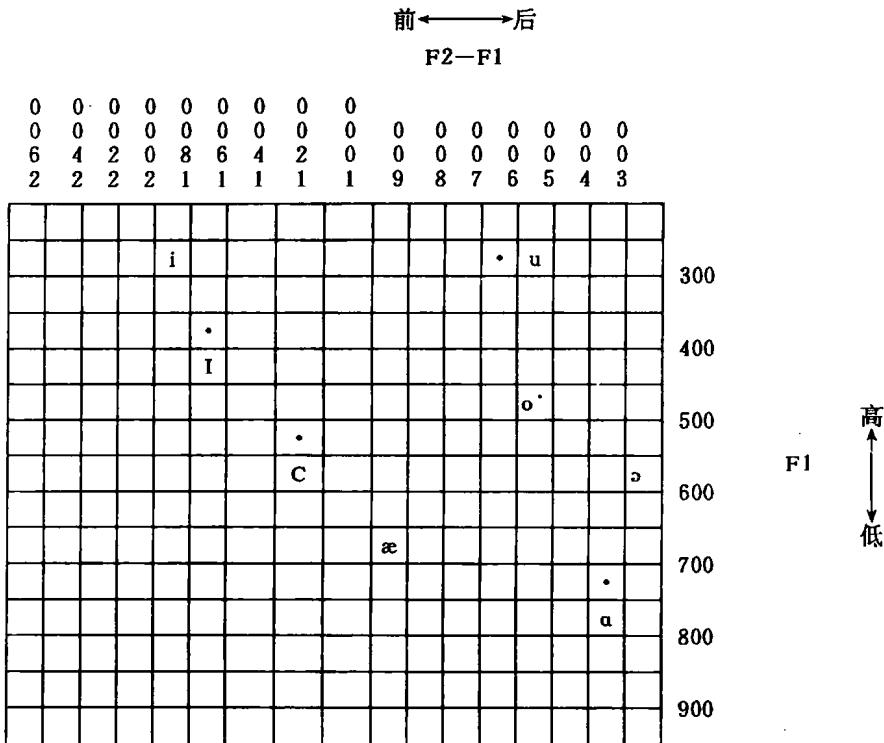
(11) 喉腔(pharyngeal)

(12) 声门(glottal)

2.2 声学语音学(Acoustic Phonetics)

2.2.1 声波

语音的来源,主要是靠声波的传递,此声波是采用空气中传播的纵波。声波中的基础频率(Fundamental Frequency, F_0)与音调有关。而基频(pitch)乃是声音在听者听觉上所引起声音高低的感受。图 2.6 表现母音与舌面所反应的最高点的高低与反舌面最高点的前后。



F1 反应舌面最高点的高线

F2—F1 反应舌面最高点的前后

图 2.6 母音在舌前最高点的位置

基频与基频之间的一个距离是一个共振峰(format)。共振峰是基础频率的信数,是能量中的频带。如图 2.7 可显示共振峰与基频间的关系。强音(intensity)与声波的振幅(amplitude)成正比。重音(stress)与强音有关,也有所差别。

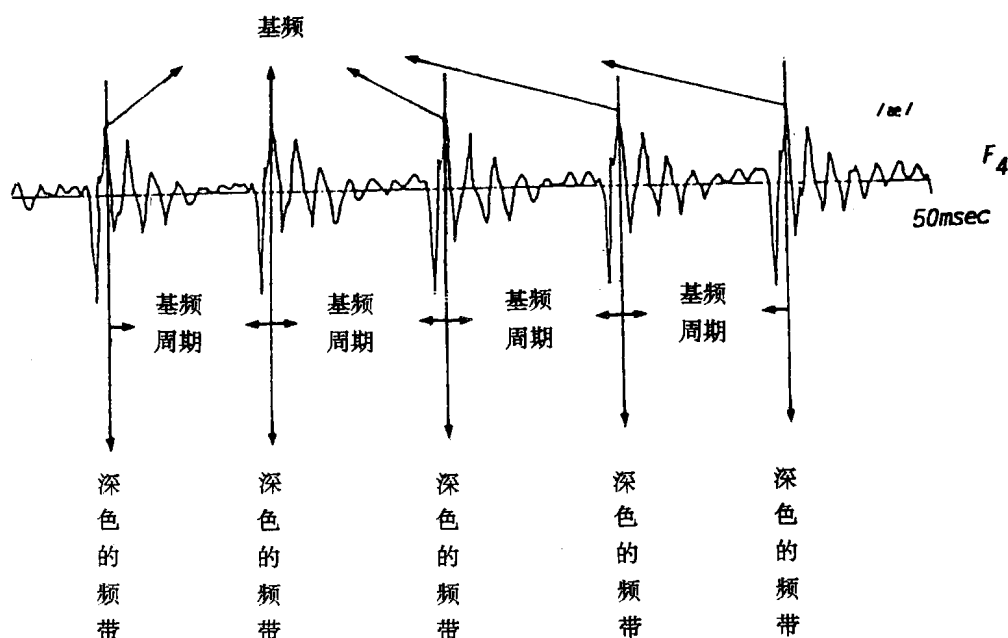


图 2.7 “æ” 母音的基频周期

2.2.2 母音

大部份的语音,包括英文,可借助一系列特殊的声音或语音来描述。特别是美式英语,有 42 种语音,包括母音、双母音、半母音及子音等四个主音。图 2.8 是语音的等级,语音大致分为四个主音,每个主音又有其相对的次音。语音又可区分为连续音与非连续音,连续音包括:母音、鼻音、擦音;非连续音包括:双母音、半母音、塞音、塞擦音。母音的产生是通过固定声道加上声带的振动激发,并透过空气中周期性的脉冲而产生。母音的发音特征是发音器官不紧接,口腔内无阻碍,发音气流畅通,声带振动。母音与另一母音之间的分别,是因为声带振动时,发音腔道中空气共鸣而产生不同的共振峰,因而产生不同的声音效果。发音腔道的共鸣可用发音腔道的大小及类型来控制,这种控制主要靠舌头、双唇及下巴的开合。

母音的声谱图可用几个音来比较“i、I、ε、æ”四个英语母音的声谱图(如图 2.9),每个母音都有若干个共振峰,图中较为深色的频带即共振峰所在,由上而下。此四个峰定名为 F1、F2、F3、F4……不同的母音。F1 与 F2 的分布有明显的不同。比较图 2.10 的语音波形,以同样的“i、I、ε、æ”四个母音做左右方式绘出,若以图 2.9 与图 2.10 比较,基频乃是深色频带的所在,基频与基频之间,是一个周期,如图 2.7 是 æ(F4)母音的例子。

再看母音的声谱图可分成宽带与窄带两种。宽带声谱图在时轴上较精确,共振峰也较清楚,窄带声谱图则是在频率的记录上比较精确,可以看到一条条横向的线,正好是基频的信数,即其信音。从窄带声谱图可以看到共振峰率牵涉了好几个信音,而且基频对共振峰影响不大。以下是同一段语音的宽带声谱图与窄频带声谱图。图 2.10 的句中的是“Is pat sad or mad?”的意思。

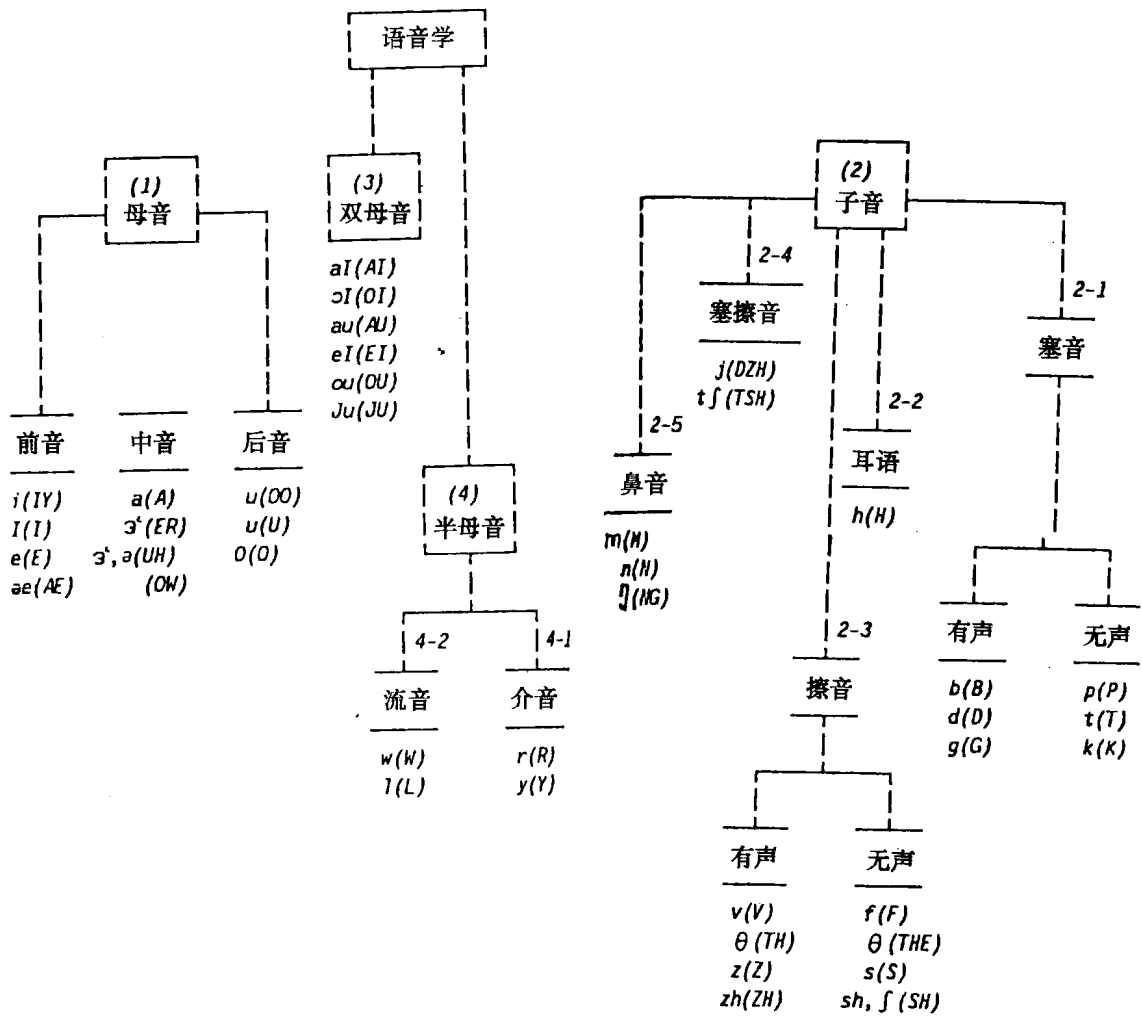


图 2.8 美式英语语音学等级

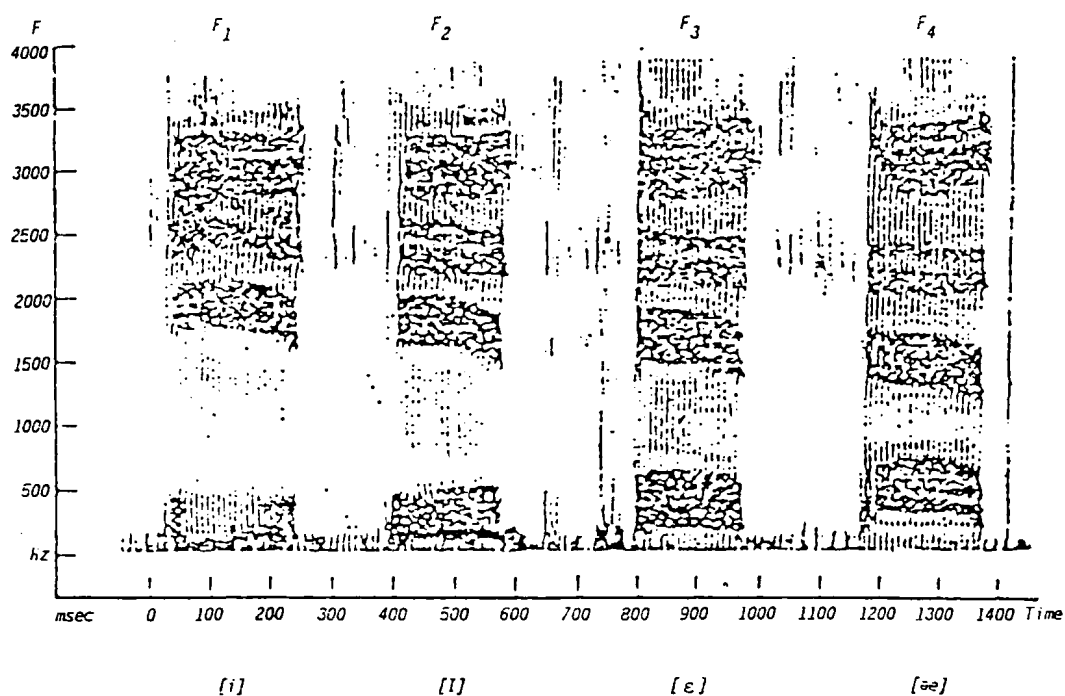


图 2.9 “i、I、ε、æ”声谱图

