

# 语音信号处理

陈永彬 编著

上海交通大学出版社

12.5

## 内 容 提 要

本书是按机械电子工业部制定的工科电子类专业教材 1986~1990 年编审出版规划, 由无线电技术与信息系统教材编审委员会推荐出版的。全书共分五章。内容包括: 语音学与语音信号模型; 语音信号分析; 语音通信; 语音合成; 语音识别。

本书主要是供大专院校学生作教材或教学参考书使用, 同时也可供从事语音信号处理工作的工程技术人员阅读。

### 语音信号处理

出 版: 上海交通大学出版社

(淮海中路 1984 弄 19 号)

发 行: 新华书店上海发行所

排 版: 上虞科技外文印刷厂排版

印 刷: 江苏太仓印刷厂

开 本: 787×1092(毫米)1/16

印 张: 10.5

字 数: 259000

版 次: 1990 年 12 月 第一版

印 次: 1991 年 1 月 第一次

印 数: 1—1300

科目: 238—497 交大版

ISBN7-313-00789-2/TP·39

定价: 2.10 元

## 出版说明

根据国务院关于高等学校教材工作分工的规定,我部承担了全国高等学校、中等专业学校工科电子类专业教材的编审、出版的组织工作。由于各有关院校及参与编审工作的广大教师共同努力,有关出版社的紧密配合,从1978年至1985年,已编审、出版了两轮教材,正在陆续供给高等学校和中等专业学校教学使用。

为了使工科电子类专业教材能更好地适应“三个面向”的需要,贯彻“努力提高教材质量,逐步实现教材多样化,增加不同品种、不同层次、不同学术观点、不同风格、不同改革试验的教材”的精神,我部所属的七个高等学校教材编审委员会和两个中等专业学校教材编审委员会,在总结前两轮教材工作的基础上,结合教育形势的发展和教学改革的需要,制订了1986~1990年的“七五”(第三轮)教材编审出版规划。列入规划的教材、实验教材、教学参考书等近400种选题。这批教材的评选推荐和编写工作由各编委会直接组织进行。

这批教材的书稿,是从通过教学实践、师生反映较好的讲义中经院校推荐,由编审委员会(小组)评选择优产生出来的。广大编审者、各编审委员会和有关出版社为保证教材的出版和提高教材的质量,作出了不懈的努力。

限于水平和经验,这批教材的编审、出版工作还会有缺点和不足之处,希望使用教材的单位,广大教师和同学积极提出批评建议,共同为不断提高工科电子类专业教材的质量而努力。

机械电子工业部教材办公室

## 前 言

本教材系按机械电子工业部制定的工科电子类专业教材 1986~1990 年编审出版规划,由无线电技术与信息系统教材编审委员会通信编审小组征稿、评选、推荐出版,责任编辑樊昌信。

本教材由西安电子科技大学樊昌信担任主审。

本课程的参考学时数为 40 学时,其主要内容为:

一、语音学与语音信号模型:汉语语音学、发音与听觉器官、语音信号的数学模型。二、语音信号分析:时域、频谱、倒谱、线性预测等分析,音调检测与汉语声调判决,矢量量化压缩。三、语音通信:波形编码,声码器及其硬件,新一代声码器。四、语音合成:按参数合成,按规则合成。五、语音识别:DTW 技术,FSVQ 技术,HMM 技术,讲话者识别,孤立字(词)识别,连续语音识别。各章之后都附有参考文献、复习题和练习题。

本教材由陈永彬编写。参加审阅工作的除了主审人之外,还有西安电子科技大学的刘德修同志,他们都为本书提出许多宝贵意见,这里表示诚挚的感谢。由于编者水平有限,书中难免还存在一些缺点和错误,殷切希望广大读者批评指正。

陈永彬

1989.5

# 目 录

<b>第一章 语音学与语音信号模型</b>	
1.1 语言学概述 .....	1
1.2 汉语语音学 .....	4
1.3 发音器官与听觉器官 .....	10
1.4 语音信号的数学模型 .....	14
参考文献, 复习题, 练习题 .....	23
<b>第二章 语音信号分析</b>	
2.1 语音信号的时域分析和预处理 .....	26
2.2 语音信号的频谱分析 .....	35
2.3 语音信号的倒谱分析 .....	45
2.4 语音信号的线性预测分析 .....	53
2.5 音调检测与汉语声调判决 .....	65
2.6 语音信号的矢量量化 .....	69
参考文献, 复习题, 练习题 .....	78
<b>第三章 语音通信</b>	
3.1 语音编码 .....	83
3.2 声码器 .....	91
3.3 新一代声码器 .....	97
3.4 语音数字处理硬件 .....	103
参考文献, 复习题, 练习题 .....	108
<b>第四章 语音合成</b>	
4.1 用波形或参数来合成语音 .....	114
4.2 汉语按规则合成 .....	120
参考文献, 复习题与练习题 .....	127
<b>第五章 语音识别</b>	
5.1 语音识别传统方法 .....	130
5.2 有限状态矢量量化技术 .....	134
5.3 隐马尔柯夫模型技术 .....	138
5.4 讲话者识别 .....	144
5.5 孤立字(词)识别系统 .....	147
5.6 连续语音识别系统 .....	155
参考文献, 复习题与练习题 .....	159

# 第一章 语音学与语音信号模型

语音信号处理是研究用数字信号处理技术对语音信号进行处理的一门学科。处理的目的是要得到一些语音参数以便高效地传输或储存；或者是通过处理的某种运算以达到某种用途的要求，例如人工合成出语音、辨识出讲话者、识别出讲话的内容等等。

人类社会愈来愈显示出信息社会的特点。不但在人与人之间，而且在人与机器之间每时每刻都需要进行大量的信息交换。通信或信息交换，已成为人类社会存在的必要条件，正如衣食住行对于人类是必要的一样。

虽然，嗅觉、触觉等也是人类固有的感觉，人们也可以从中得到某些外界信息，但是可以指出，最重要的、最精细的信息源只有图像和语言两种。而且，语言是人类最重要、最有效的、最常用的和最方便的通信形式。

随着现代科学和计算机技术的发展，除了人与人之间的上述自然语言的通信方式之外，人机对话及智能机器人等领域也开始使用语言了。这些人工语言同样有词汇、语法、句法结构和语义内容等。控制论创始人维纳在 1950 年就曾指出过：“通常，我们把语言仅仅看作人与人之间的通信手段，但是，要使人向机器，机器向人以及机器向机器讲话，那也是完全办得到的。”这个预见实现的日子终于即将到来。

我们认为，语音信息交换大致上可以分为三大类：（1）人与人之间的语音通信（包括语音压缩与编码、语音增强等）。（2）第一类人机语音通信问题，指的是机器讲话、人听话情况。这就是“人工嘴巴”的研究，或者用技术术语来讲，这就是：语音的人工合成的研究。（3）第二类人机语音通信问题，指的是人讲话，机器听话情况，这就是“人工耳朵”的研究。同样，用技术术语来讲，这就是：语音的人工识别和理解的研究。

语音信号处理又是一门新兴的边缘学科。如上所述，它是“语言-语音学”与“数字信号处理”两个学科相结合的产物。由于前者已有很多参考书籍，而后者是本课程的先修课程；所以下面仅简短地介绍：在本课程中要用到的语言学和汉语语音学的有关内容，介绍发音器官和听觉器官的有关知识以及讨论语音信号的数学模型。这些内容是学习本课程的基础知识。

## 1.1 语言学概述

语言学是以人类的语言为研究对象的一门科学。语言(Language)是从千百万个人的言语(Speech)中概括总结出来的规律性的符号系统。所以，研究语言首先要了解一下人的言语过程。

言语过程可以分为五个阶段，如图 1.1.1 所示：

(1) 想说阶段 人的说话首先是客观现实在大脑中的反映，经大脑的决策产生了说话的动机；接着讲话神经中枢选择恰当的单词、短语以及按语法规则的组合，以表达他想说的内容和情感。这个阶段与大脑中枢的活动有关。

(2) 说出阶段 由上阶段中枢的决策，以脉冲形式向发音器官发出指令，使舌、唇、颚、声

带、肺等部分的肌肉协调地动作,发出声音来。当然,与此同时,大脑也发出其他一些指令给其他有关器官,使之产生各种动作来配合言语的效果,如:面部表情、手势、身体姿态等。另外,还开动了另一个“反馈”系统,来帮助修改言语。这就是:他不但发出言语,而且他自己的听觉系统也在听自己的言语。但是,在这个阶段中,主要的是与发音器官的活动有关。

(3) 传送阶段 说出来的言语是一连串声波,凭借空气为媒介传送到听者的耳朵里。当然,有时遇到某种阻碍或其他声响的干扰,使声音产生损耗或失真。这阶段中,主要是传送信息的物理过程起作用。

(4) 接收阶段 从外耳收集到的声波信息,经过中耳的放大作用,到达内耳。经过内耳基底膜的振动,激发柯替氏器官内的神经元使之产生脉冲,将信息以脉冲形式传送给大脑。在这个阶段中主要是与听觉系统的活动有关。

(5) 理解阶段 听觉神经中枢收到脉冲信息之后,通过一种至今尚未完全了解的方式,辨认出说话的人及其所说的信息,从而听懂了讲话者的话。

从五个阶段来看,言语的过程包含着相当复杂的因素,其中有心理的、生理的、物理的以及个人的和社会的因素。这里,个人的因素是指讲话者的口音和用词造句的特色以及听话者的听音和理解能力;社会的因素则是指讲话者和听话者对用于进行交际的手段有共同的理解的社会基础。

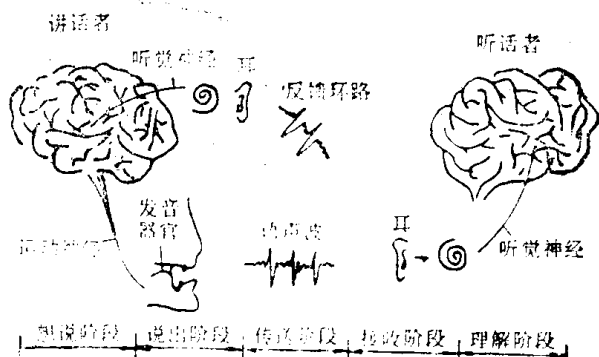


图 1.1.1 人的言语过程

语言是从言语中概括出来的一个符号系统。包括形式和内容两个方面,即语音的形式和语义的内容。将这两个基本要素结合起来,可以构成语言的语素、词、短语和句子等的不同层次的单位;这个构成规则就是语法。

以汉语的一个句子“他热爱祖国”为例:

语音形式: tā rè ài zǔ guó.

语义内容: 他 热 爱 祖 国。

这个句子切分出来的三个词的语音形式就相应地表达了这三个词的语义内容。

语言学是一门既是最古老的、又是最年轻的学科之一。研究表明,自从人类从能劳动而分化出来时就开始有了语言。大约有几十万年、甚至 100 万年的历史了。顺便提一下,语言的出现要比文字早得多,文字至多只有一万年的历史。但是,世界学者们公认,从 19 世纪开始,人们才开始从各个角度认真地去研究语言,才出现了现代语言学这门科学。最近十几年来还出现了这么个词,叫做“连接号语言学”(Hyphenated Linguistics),它们就是语言学与其他学科相交叉的一些边缘学科。例如:

(1) 语言(语音)-信号处理,研究语言(语音)信息的编码、传递、储存、合成和识别等技术。(2) 语-音学,研究语言声波的各种性质。尤其其中的一个分支——实验语音学,用各种实验方法来研究语音,取得了不少成果。(3) 数理-语言学,研究语言符号系统的数理逻辑,尤其注意语义在推理中的作用。(4) 心理-语言学,研究言语时,大脑和神经系统的心理活动。(5) 社会-语言学,研究语言的应用对于各种社会因素的依从关系。(6) 人类-语言学,研究古代的遗迹和人类头骨及发音器官的化石,探讨语言的起源和历史演变的过程。本课程就是上述(1)的边缘学科。

人类的语言是如何起源的?又是如何演变的?

1976年,美国纽约州的科学院主持召开了一次规模很大的学术会议——语言和言语的起源与演变的学术讨论会。有几千人参加,内容涉及到十几个学科,并发表了一本论文集<sup>[1]</sup>。从这个论文集中可以看出如下三个问题的研究成果:

(1) 语言起源的时间。大约1000万年前开始人和黑猩猩与猿类的分化。42万年前北京周口店猿人开始用火。所以大致可以说,语言大约是在几十万年至100万年前发生的。

(2) 语言与语音的关系。由于进化的结果,猿人开始站立并用两只脚走路。这一点与语言发生有密切关系。因为站立,喉头受地心吸力的影响逐渐下移,除口腔之外形成了咽腔,增加了发音的能力。例如,元音[i],[u],[a]都需要声门相当低才能发出来,[k],[g],[ŋ]这一类舌根音也是这样,有了发音的条件,才可能有语言的产生。

(3) 语言与大脑的关系。大脑是产生语言的根本因素。人的大脑最大、最重(约1500g),大脑皮层面积最大(有皱纹,使脑神经细胞达140亿个左右),人的大脑发育最快(婴儿大脑350g,成人1500g)。

下面简单讨论语言的演变问题。

大家都承认,在人类社会的演变过程中,语言是在变化的。语言的演变包括语音的变化、语义及词汇的变化以及语法的变化等。举一些例子来说明如下。

(1) 语音的变化。例如,汉语音位的变化,原来m、g,这两个音位在词头词尾都可以用,现在m只能在词头,g只能在词尾使用。又如,古汉语中所有浊塞音、浊塞擦音和浊擦音,在现代北京话中(中国一些方言除外)都已全部清化了。英语中,语音的变化也很多。如:动词被名词化之后,其读音发生变化。又如,非重读元音的丢失现象、英语中不规则动词的过去式,开始向规则动词转化,也用词尾-ed表示了等等。

(2) 词义的变化。这有扩大、缩小、转移三种情况。

1) 扩大:如汉语中“江”,原仅指长江,现泛指“江”。如英语中manuscript(底稿)[ˈmæn.jʊskript]原指手写的底稿,现泛指所有底稿,包括打字机打的底稿。

2) 缩小:如汉语中“瓦”,原指一切土器,现仅指屋顶上的瓦。如英语中的meat [mi:t]原指一切食物或用餐(如green meat 青菜;before meat 餐前),现仅指“肉”。

3) 转移:如汉语中“兵”,原指武器,现转指人——战士。如英语中book,原指一种可用来写字的山毛榉(beech)树的树皮,现转指书本。

此外,语法上也有许多演变,这里就不详细介绍了。

下面,让我们讨论语言学的发展与近况。

中国是世界上最早进行语言研究的国家。早在汉代(公元前206年至公元220年,许慎在《说文解字·序》里,就已经在许多处提到汉语的演变的有关问题了。



语言学的发展大致上可以分为四个时期。

(1) 历史比较语言学时期 本世纪之前,就不断有人研究语言问题。其方法是从世界各地搜集和整理各种语言材料,进行分类和具体的分析、归纳与总结,从而得到某些结论。例如,证实了印度、伊朗的语言与欧罗巴语言之间的有机联系等等。

(2) 结构语言学时期 以瑞士的语言学家索绪尔(Ferdinand de Saussure,1857~1913)为代表,开始了对各种语言的结构以及各种语言之间的关系进行了研究;同时开始了各种实验,建立了实验语音学这个学科;提出了语言的符号系统。他们的最主要贡献在于:语言在结构上被科学地划分为句、词、音节、音素等不同层次,也就是说,在语音方面,建立了音位学和区别特征理论;在语义方面,建立了语素(morpheme),语素音位(morpho-phoneme)等。这一时期一直延续到1957年。

(3) 转换生成语言学时期 这个时期的代表人物是美国的乔姆斯基(Noam Chomsky 1928~),其代表作是《句法结构》(Syntactic Structures)。在此期间诞生的两门学科:信息论(C.E.Shannon提出,1916~)和控制论(Norbert Wiener提出,1894~1964),渗透到语言学中来了。Chomsky提出并建立一套语言演绎系统,他认为语法是一套规则,根据这些规则,可以推导生成出无限多的合法句子。这就是转换生成语法(Transformation Generative Grammar,简称TG)。这些规则如:短语结构规则、助动词转换规则等。这种用有限规则生成无限句子是人的内在能力的表现。所以,这也促进了对人脑与人工智能的研究。

(4) 目前时期 到了60年代,Chomsky已发现他的理论的缺陷。于是他不断进行规则的修补。到1968年之后,不断有人评论这些缺陷。尤其,发现规则和修补规则都是很抽象的研究,与现实社会的语言之间脱节十分严重。目前,美国是语言学研究的中心,投入的人力最多,发表的文章最多。1973年在美国密执根(Michigan)办了一次语言学讲习班,是一个重要的转折点。会上提出要改变抽象的研究,要研究语言在时间和地理上的变化以及它与社会的关系。从此,思想活跃起来了,各种人有各种不同的想法。如Willian Labov提出变项规则(Variable Rule),企图用它判断语言变化的趋向。此外也出现了一些新刊物,如《大脑与语言》(Brain and Language),《语言与社会,社会语言学》(Language and Society, Sociolinguistics)。不少语言学家开始走入实际,注重语言的应用(如应用语言学 Applied Linguistics 这个学科),有的与医生合作,研究失语症、阅读困难症等病症等。

## 1.2 汉语语音学

汉语是世界上最大的语种,使用人口达十亿。汉语标准语音指的是北京语音,在中国又称为普遍话。

汉语语音的特点如下。

(1) 音系简单。这是指音素少、音节少(大约有64个音素,但只有400个左右音节,如考虑每个音节有五个声调,也只不过1200多个有调音节)。音节的结构也比较简单。例如:1) 尾音只有i、u、o、n、ng五种;2) 没有不除阻的m、b、d、g等音;3) 没有喉塞音[ʔ];4) 声母都为清音,浊声辅音只有m、n、l、r四个;5) 没有复合辅音;6) 元音的长度没有区分语义的作用;7) 音素的发音方法和发音部位都是一般的,没有边擦音等。

(2) 听感上有清亮、高扬和舒服、柔和的感觉。其原因是:1) 清辅音多,而且多是弱清

音。因而没有肌肉紧张、气流快促的感觉。2) 没有入声这种短促的发音。由于平、上、去三调都是舒声,而且调值的高频成份较多。3) 开口呼的音节占全部音节的一半以上;用 a 这个音素为主要元音的音节就占 40%。由上三点,汉语具有好的听感。

(3) 有鲜明的轻重音和儿化韵,所以字词分隔清楚,语言表达准确而丰富。轻音与重音配合使语气活泼,语义明显,感情流露。儿化韵能起适当的语法修词作用。

汉语的拼音方法如下。

汉语由音素构成声母或韵母。有时,将含有声调(汉语通常认为有五个声调)的韵母称为调母。由单个调母或由声母与调母拼音成为音节。汉语的一个音节就是汉语一个字的音,即音节字。由音节字构成词(其中主要是两音节字构成的两字词,约占74%),最后再由词构成句子。

国际上,都是用音标来描述拼音过程的。汉语也不例外。而且有些音标是汉语所特有的符号(如,1、1、ə、ɛ 等)。汉语拼音的音标包括:声母表、韵母表和声调符号等。

汉语声母共有 22 个,其中包括零声母。零声母是指没有声母、只有调母的情况。除零声母之外,其他所有的声母全部都是单辅音。可用字母、汉语名称和国际音标符号“[ ]”表示如下。

b[p]玻、p[p']坡、m[m]摸、f[f]佛、d[t]得、t[t']特、n[n]纳、l[l]勒、g[k]哥、k[k']科、  
h[x]喝、j[tɕ]基、q[tɕ']欺、x[ç]希、zh[ʒ][ts]知、ch[ç̌][ts']蚩、sh[ʃ̌][s]诗、r[ʒ]日、  
z[ts]资、c[ts']雌、s[s]思。

汉语韵母共有 38 个,其中有三个(-i, er, ê)是特殊韵母:(1) -i 有两种发音,即[ɿ]资韵,[ɨ]知韵。由于它们是互不重叠的,所以只需用一个韵母符号来表示。例如,在[ɿ]前的声母只能有 z、c、s;而在[ɨ]前的声母只能有 zh、ch、sh、r。(2) er(ɛ)是儿化韵,很少用到。(3) ê 的发音为[E],常在 ie 这个韵母的韵尾中用到它。

除了这三个韵母之外,其余的 35 个韵母的拼音字母及其汉语名称和国际音标符号如表 1.1 所示。

表 1.1 韵母表

	单元音(6个)		复合元音(13个)				复合鼻元音(16个)					
	a [a] [A] [ɑ]	o [o] [ɔ]	e [ɛ] [ə]	ai [aɪ]	ei [eɪ] 欸 飞	ao [aʊ]	ou [əʊ]	an [an]	en [ɛn]	ang [ɑŋ]	eng [ɛŋ]	ong [ʊŋ]
				哀(韵)	熬	欧	安	恩	昂	亨(韵)	轰(韵)	
i 行	i [i] 衣	ia [iA] 呀	ie [ie] 耶	iao [iaʊ]	iou [iəʊ] 忧			ian [iæn]	in [in]	iang [iaŋ]	ing [iŋ]	iong [iʊŋ] 雍
u 行	u [u] 乌	ua [uA] 蛙	uo [uɔ] 窝	uai [uaɪ] 歪	uei [ueɪ] 威			uan [uan]	uen [uɛn]	uang [uaŋ]	ueng [Uɛŋ] 翁	
ü 行	ü [y] 遇			üe [yɛ] 约				üan [yen] 冤	ün [yn] 晕			

汉语的声调只有阴平、阳平、上、去以及“轻声”等五种声调。它们的符号如下：阴平—，如：mā 妈；阳平 / ，如：má 麻；上声 ∨ ，如：mǎ 马；去声 \ ，如：mà 骂；轻声 ma 吗。

汉语的音素可分为辅音、单元音、复元音、复鼻尾音四种。音素是发出各不相同的音的最小单位；也就是说，它们都有自己独立的各不相同的发音方法和发音部位。除了这四种音素之外，元音中还有几个借用国际音标的音素，附在单元音中介绍。现讨论汉语的各种音素如下。

1. 辅音。共有 22 个，除了 21 个声母（因为它们都具有各不相同的发音方法和发音部位）之外，还有一个韵母中的鼻韵尾音 ng[ŋ]。如表 1.2 所示。

表 1.2 辅音音素表

部位 方法		双唇阻	齿唇阻	舌尖前阻	舌尖阻	舌尖后阻	舌面阻	舌根阻
		上唇与下唇	上齿与下唇	舌尖与上齿背	舌尖与上牙龈	舌尖与前硬腭	舌面与硬腭	舌根与软腭
塞音	不送气	b[p] 玻			d[t] 得			g[k] 哥
	送气	p[p'] 坡			t[t'] 特			k[k'] 科
塞擦音	不送气			z[ts] 资		zh[tʂ] 知	j[tɕ] 基	
	送气			c[tʂ'] 雌		ch[tʂ'] 蚩	q[tɕ'] 欺	
擦音	(清音)		f[f] 佛	s[s] 思		sh[ʂ] 诗	x[ç] 希	h[x] 喝
	(浊音)					r[ʐ] 日		
鼻音 (浊音)		m[m] 摸		n[n] 讷, 声母或韵尾				ng [ŋ]
边音(浊音)				l[l] 勒				

汉语中浊辅音很少，只有 m、n、l、r 四个。其他都是声带不振动的清辅音。但 ng 则是浊辅音的韵尾音。表中送气是指发音后释阻，由气流冲出而发出声音来。这时应在右上方加符号 ' 表示之。

2. 单元音。共有 13 个，此外还有 7 个借用音。13 个单元音如表 1.3 所示。图 1.2.1 示

汉语单元音的发音舌位图。

表1.3 单元音音素表

类别 舌位 高低, 口形 开闭	舌面元音					舌尖元音		
	前		央	后		前	后	央
	展唇	圆唇		展唇	圆唇			
高, 闭	i [i] 衣	u [y] 遇		u [u] 乌	-i [ɿ] 资韵	-i [ʅ] 知韵		
半高, 半闭				e 鹅 [ɤ]	o 喔 [o]			
中, 中			e 的韵 [ə] (轻声)					er 儿 [ɛr]
半低, 半开	ê [ɛ̃] 别韵							
低, 开	a [a]		a [A] 啊	a [a]				

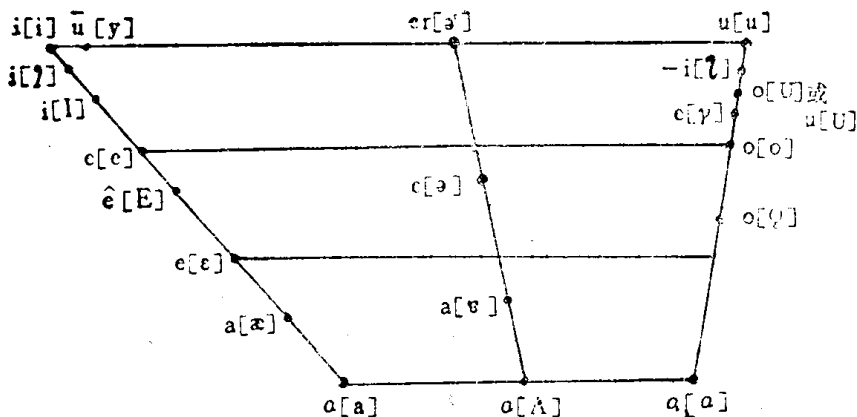


图1.2.1 汉语单元音的发音舌位图

应该注意,元音并不等于韵母。元音、辅音是按音素的发音特征来分类的;而声母、韵母则是按音节结构来分类的。这是两种不同的概念,尽管它们之间有一定的联系。

下面再说明几点：

(1) 13个元音中，列入韵母表的只有10个：i[i]、ü[y]、u[u]、-i(包括互不重叠的[ɿ]和[ʅ]两个元音，都用拼音字母-i表示)、e[ɤ]、o[ɔ] (这里[o]下面的符号c，表示它与国际音标的[o]不同，其舌位稍低些，而c表示汉语语音的意思)、er[ɛ]、ê[E]、a[A]等。

(2) 除了上述10个音素之外，还有三个单音：e[ə]、a[a]、a[ɑ]。其中第一个e[ə]只用于轻声，如：de[tə]“的”。而后两个则也只能用于轻声或拼音音素，不能单独用。

(3) 所有汉语元音都没有鼻化的成分，发音时软腭上升，挡住鼻腔；即没有[ã]、[õ]、[ỹ]、[ũ]、[ü̃]等音。

(4) 舌尖前元音-i[ɿ]，舌尖后元音-i[ʅ]和卷舌元音er[ɛ]是汉语语音所特有的元音音素。

(5) 元音舌位图中的每一个点，写在左边的一个音素表示展唇、写在右边的一个音素表示圆唇。

(6) 在舌位图中，左边为前元音区、中部为央元音区、右边为后元音区。上下的顺序为：高元音区、次高元音区、次低元音区、低元音区。

(7) 除了这13个单元音之外，汉语中还借用了7个国际音标的发音单元音音素。它们只在拼音中使用，不能单独作为韵母使用。它们是：i[I]、e[e]、e[ɛ]、a[æ]、a[ɐ]、o[o]、o[U]或u[U]。

3. 复元音。共13个，它们都是韵母表中的韵母。从元音舌位图可见，单元音只占据某些点的位置。也就是说，人的发音器官实际上是可以发出无限多个单元音的。复元音的发音的方法是：按复元音中的单元音的顺序连续地移动舌位、唇形而发出声音。但是这不等于加法(如：ei ≠ e + i)，而是一种新的“动态”的声音，成为一种固定的音组，并被视为一个语音单位。这对于复鼻尾音也适用。这13个复元音如表1.4所示。

表1.4 汉语复元音音素表

ai	[aɪ] 哀	ei	[eɪ] 飞韵	ao	[aʊ] 熬	ou	[əʊ] 欧	前响
ia	[iA] 鸦	ie	[ie] 椰	iao	[iaʊ] 腰	iou	[iəʊ] 优 (iu)	
ua	[uA] 娃	uo	[uɔ] 窝	uai	[uaɪ] 歪	uei	[ueɪ] 威 (ui)	
	后	响			中		响	

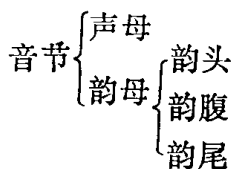
现说明几点如下：(1) 二合元音有9个，三合元音有4个。(2) 复元音的发音有前响、中响、后响之分。例如：ai[aɪ]、ei[eɪ]、ao[aʊ]、ou[əʊ]这4个二合元音是前响的。即：a、e、a、o要发的重些、长些，而i、i、o、u4个尾音要发得轻些、短些。其余5个二合元音，则是后响的。至于三合元音，可视为由首元音加后面的双元音组成。此时由于双元音都是前响的，所以这4个三合元音就都是中响的了。(3) 三合元音中的iou(iu)和uei(ui)，当它自成音节，或与声母拼音并且声调是阴平或阳平时，中间的o[ə]或e[e]音消失；而当它与声母拼音且声调是上声、去声时，中间的o[ə]和e[e]要发音。

4. 复鼻尾音。共16个，它们也都是韵母表中的韵母。在汉语中，鼻韵尾只有两个：-n、-ng。它们与元音复合之后也成为不可分隔的音组。复鼻尾音音素如表1.5所示。

表 1.5 汉语复韵尾音韵素表

an [an] 安	en [ɛn] 恩	ang [aŋ] 肮	eng [ɛŋ] 半韵	ong [Uŋ] 东韵	前响
ian [iɛn] 烟	in [in] 因	iang [iaŋ] 央	ing [iŋ] 英	iong [iUŋ] 拥	
uan [uan] 弯	uen [uɛn] [un]温	uang [uaŋ] 汪	ueng [uɛŋ] 翁	中响	
uan [yɛn] 渊	un [yn] 晕	前响			

下面,讨论汉语的音节的结构。这可写成:



在全部汉语音节字中,这四个部分有全备的,也可能是缺项的。汉语音节字的构成规律如下:(1)必须有韵腹。它必是主元音。单元音都可以作为韵腹。(2)22个辅音中,除ng外都可作为声母;而韵尾鼻音则只有-n和-ng。(3)韵头只有由i,u,ü三个元音,其他元音不能作韵头。(4)韵尾除-n及-ng外,只有i,u,o三个元音,其他元音不能作韵尾。

汉语的韵母还有一种分类法,即可分为:开口呼类(以a,o,e开头及er,-i在内),齐齿呼类(以i开头的),合口呼类(以u开头的)和撮口呼类(以ü开头的)等四类。但其中有两个例外:ong算是合口呼,iong算是撮口呼。韵母的这种分类法对于拼音规律的描述很有用。具体分法如下。

- (1) 开口呼类(共14个): a, ai, ao, an, ang, o, ou, e, ei, en, eng, er, e, -i。
- (2) 齐齿呼类(共9个): i, ia, ie, iao, iou, ian, in, iang, ing
- (3) 合口呼类(共10个): u, ua, uo, uai, uei, uan, uen, uang, ueng, ong。
- (4) 撮口呼类(共5个): ü, ue, uan, un, iong。

我们知道,声母有22个,韵母有38个。如果全部都拼成音节字的话,将有 $22 \cdot 38 = 836$ 个无调音节字。但实际上,即使加上语气词,汉语也只有412个无调音节字。也就是说,有一半以上的拼音音节对于汉语是没有意义或没有字的。另外,汉语主要有四种声调,如果22个声母与 $4 \cdot 38 = 152$ 个调母相拼音,将有 $22 \cdot 152 = 3344$ 个音节字。实际上,汉语全部音节字连同轻音字在内,大约也只有1282个左右,这比其他语言都要少。但用它们组成词以及组成句,却比较灵活、表达力也比较丰富。这是汉语的一个重要优点。

在汉语拼音的过程中,从声母到调母之间存在着“过渡音段”。这是指声母已经除阻、而调母的舌位尚未就位时的情况。经研究,过渡音的构成是:(1)送气的塞音和塞擦音(p,t,k,c,ch,q)作为声母时,过渡音是这些声母的送气余波。(2)清擦音(f,s,sh,x,h)作为声母时,由于它们也有气流送出,所以,过渡音也是这些声母的送气余波。(3)浊辅音(m,n,l,r)作为声母时,由于拼音时它们在除阻后声带仍不停止振动,所以过渡音就是它们的声带振动音。(4)不送气的塞音和塞擦音(b,d,g,z,zh,j)作为声母时,由于它们除阻时的唇形是被其后的调母

的发音唇形所同化(如:bin 宾和 ban 班,这两个字的声母 b,在除阻时它们的唇形是不同的);也就是说,从声母除阻至调母发音之前,唇形已经是后面的调母的唇形,声带已经振动。所以,过渡音就用后面调母元音的声带振动音(预振动)。

汉语语音一些特性,如:辅音音长,声母韵母及声调的频谱和出现概率与结合概率,汉语词长及词频统计以及语音质量(即清晰度、可懂度)测试标准等,请参阅有关文献<sup>[2][3][4][5]</sup>,这里就不再赘述了。

### 1.3 发音器官与听觉器官

先讨论发音器官。它包括:肺、气管、喉(包括声带)、咽、鼻和口。这些器官共同形成一条形状复杂的管道。其中喉以上的部分称为声道,它的形状随着发出语音的不同其形状是变化的。喉的部分称为声门。

产生语音的能量,来源于正常呼吸时肺部呼出的稳定气流,喉部的声带既是一个阀门又是一个振动部件。在说话的时候,声门处气流冲击声带产生振动,然后通过声道响应变成语音。由于发不同音时,声道的形状不同,所以听到不同的语音。应注意辅音有两种发音方法。一种是声门打开,但声道的某处收紧而形成湍流,这种高速湍流导致发出摩擦音。另一种声门也打开,但声道的某处完全阻塞,然后突然放出气流,这种高压气流导致发出阻塞音。以上就是用发音器官发出语音的概况。如图 1.3.1 所示。

“肺”内可容纳约 1.75 l 容积的空气。正常呼气时大约能呼出 275 ml 的空气,在讲话时肺的气压比大气压大百分之一左右。不讲话时,呼吸的时间大致相等。在讲话时可使呼气时间达到整个呼吸周期的 85% 左右。气管是由一些环状软骨组成的,讲话时它将来自肺部的空气送到喉部。

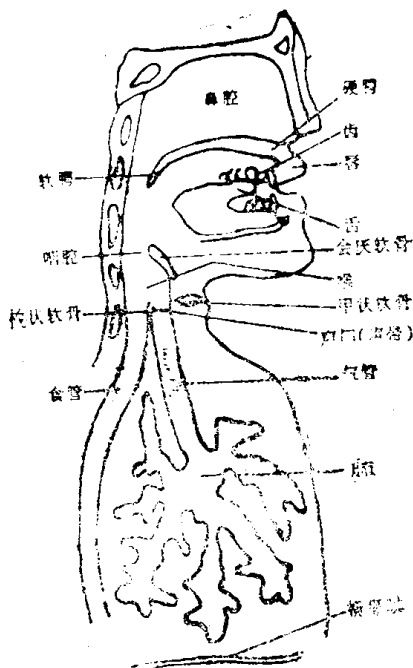


图 1.3.1 发音器官的生理解剖

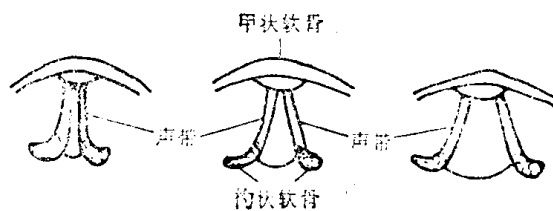


图 1.3.2 声门开度的控制情况

“喉”是由许多软骨组成的。突出在颈部的喉结称为骨甲状软骨,喉的顶部是梨状的会厌软

骨。会厌软骨的作用是：在吞咽食物时不让它进入气管。对发音影响最大的是从喉结至杓状软骨之间的韧带褶，称为声带。声带的长度仅约 10~14 mm，比指甲还小。呼吸时左右两声带打开，讲话时则合拢起来。两声带之间的部位称为声门。声门的开启和关闭是由两个杓状软骨控制的，它使声门呈 $\Delta$ 形状开启或关闭。讲话时声带合拢因而受声门下气流的冲击而张开；但由声带韧性迅速地闭合，随后又张开与闭合……。不断地张开与闭合的结果，使声门向上送出一连串喷流。图 1.3.2 示声门开度的控制情况。声带每开启和闭合一次的时间就是音调周期。它的倒数称为音调频率。

图 1.3.3 示声带开启的面积与时间的曲线。由图可见，声带开启大约用了 4ms 的时间，声带开启的面积约达最高峰 8 mm<sup>2</sup> 左右；随后大约用了 3 ms 的时间闭合起来；然后受气管的气流冲击 1 ms 时间，又重复开放。所以，音调周期为 8 ms，或音调频率为 125 Hz。这个频率是一般成年男子的音调频率。通常，音调频率取决于声带的大小、厚薄、松紧程度以及声门上下之间的气压差的效应等。其范围约为 60~350 Hz 左右(约两个半八度音左右)。但就某个人来说，仅约为一个半八度音左右。音调频率范围随发音人的性别、年龄而定。老年男性偏低，小孩和青年女性偏高。

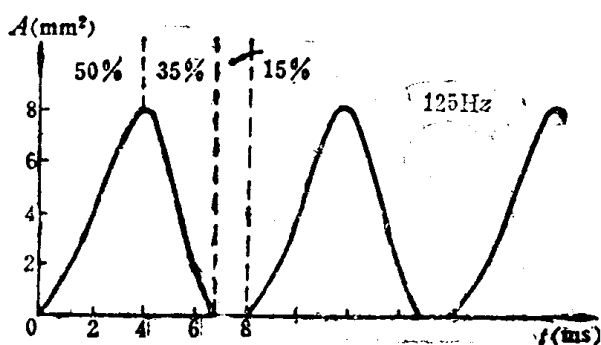


图 1.3.3 声带开启的面积  $A$  与时间  $t$  的关系曲线

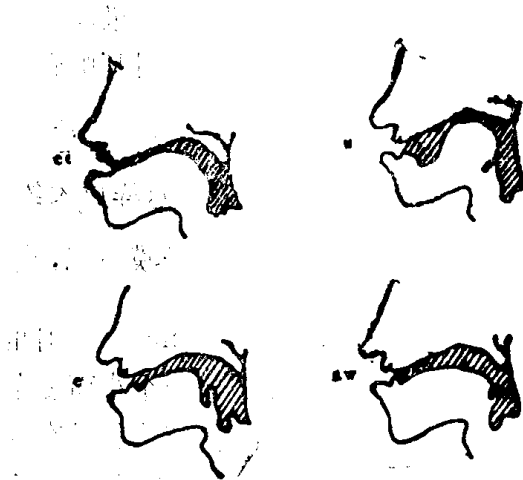


图 1.3.4 发不同音时咽腔形状的变化

“声道”由咽腔、鼻腔和口腔三个空气腔体组成，起于声门，止于两唇。一般成年人声道的长度大约是 17 cm 左右，最大截面积可达 20 cm<sup>2</sup> 左右。咽腔是连接喉和食管与鼻腔和口腔的一段管子。在讲话时，咽腔的形状是有变化的，如图 1.3.4 所示。咽腔与口腔一起使得声道形状的变化增多，因而能发出较多的不同的声音。鼻腔从咽腔一直沿伸到鼻孔，约 10 cm 长。发鼻化语音时，软腭下垂；如果它上抬，则完全由口腔发出语音了。口腔是声道最重要的部分，它的大小和形状可以通过调整舌、唇、齿和颚来改变。舌最活跃，它的尖部、边缘部、中央部都能分别自由活动；整个舌体也能上下前后活动。双唇位于口腔的末端，它也可活动成展开的(扁平的)或圆形的形状。齿的作用是发齿化音的关键，如 [θ] 音等。最后，颚中的软颚如前所述，是发鼻音与否的阀门。至于硬颚及齿龈则是声道管壁的构成部分，也参与了发音的过程。由上所述可见，声道是自声门声带之后的、最重要的、对发音起决定性作用的器官。

下面讨论人的听觉器官。它包括外耳、中耳和内耳三个部分，让我们顺序介绍如下。整个听觉器官的纵剖面如图 1.3.5 所示。

外耳。外耳是听觉器官的第一层，虽然比较简单，但在听觉系统中起着放大和保护的重要作用。外耳由耳廓、外耳道和鼓膜组成。整个长度约 25 mm 左右。可以用简单公式计算出



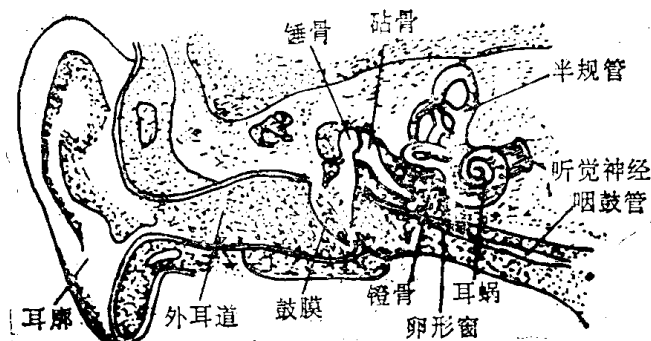


图 1.3.5 听觉器官的纵剖面

它的谐振频率,大约 3400 Hz 左右。所以,在 3~4 kHz 范围内,外耳可使输入的声压放大约 2~4 倍,也就是使鼓膜处的声压增大了约 3~6 dB 左右。在生理上,外耳还起着保护鼓膜的作用。

中耳。中耳包括三块听小骨,即锤骨、砧骨和镫骨。它们组成了鼓膜至内耳卵形窗之间的机械链。锤骨的柄固定地附着在鼓膜且覆盖着一半以上的膜面,镫骨的踏板则覆盖着内耳入口的卵形窗膜上。中耳房还通过耳咽管与外界空气相通,以免压力差而损害鼓膜。中耳具有两个主要用途。一是放大声压,三块听小骨起着机械杠杆的放大作用(由公式  $F_2 = \frac{d_1}{d_2} F_1$ ,  $d_1$  是砧骨长度,  $d_2$  是镫骨长度。这样放大约 1.5 倍左右),更重要的放大作用是骨传导产生的。

(由公式  $P_2 = \frac{A_1}{A_2} P_1$ ,  $A_1$  是鼓膜面积,  $A_2$  是卵形窗面积。通常,  $A_1 \approx 22 A_2$ , 故可增强声压 22 倍左右。鼓膜面积约  $0.065 \text{ cm}^2$ )。中耳的第二个用途是保护内耳免受特强声音的损害。

内耳。内耳是颅骨腔内一个小而复杂的系统。其中对听觉作主要贡献的是充满液体的耳蜗。从声振动到神经冲动的转化就是在耳蜗中完成的。耳蜗高约 5 mm,直径约 3 mm,自身绕成两圈半,拉直后约 30~32 mm 长。图 1.3.6 示拉直后的耳蜗的纵剖面、横剖面以及耳蜗的横断面。可以看到,除了尖端部分之外,整个耳蜗由基底膜和瑞士膜隔开成三个区域。中部区域称为耳蜗导管,上下两分区域分别称为前庭阶和鼓阶。前庭阶与鼓阶在尖端部分相通。耳蜗导管中充满高粘度的胶状的内淋巴液。而相通的前庭阶和鼓阶内则充满粘度为水 2 倍的淋巴液。当有高频的声振动刺激镫骨时,淋巴液就产生压力的变化,传到鼓阶内侧的基底膜上。基底膜在耳蜗底部最窄(宽度约为 0.04 mm),在耳蜗尖端部最宽(约 0.5 mm 左右)。1960年, G. Von. Bekesy 用正弦信号对基底膜进行了详细的研究,得出一些重要的结果,揭开了听觉的奥秘,获得了诺贝尔奖金。原来,基底膜听觉响应是与刺激的频率有关的。频率较低时,靠近耳蜗尖部的基底膜产生响应;反之,频率较高时,则靠近圆形窗的窄而紧的基底膜产生响应。基底膜的频率响应的空间分布,导致基底膜上不同位置的柯替氏器官的纤毛细胞对不同频率的声音引起弯曲,从而刺激其附近的听觉神经末梢,产生电化学脉冲,并沿听觉神经束传递到大脑。大脑对送来的脉冲进行至今尚未充分研究的一系列分析与判断,就可以识别并理解该语音的含义。

下面,还必须简短介绍一下神经系统。因为发音,尤其是听觉,都牵涉到人的神经活动。在发音时,将观念变成单词和句子并发出指令,控制发音器官使作适当运动;在听音时,将内耳的柯替氏器官发出的脉冲,经过神经系统的处理,才能使大脑感知这些编了码的信号,变成词