

多学科学术讲座丛书

5

语言信息和语言通信

马大猷著

知识出版社

上海

责任编辑：陈荣乐

多学科学术讲座丛书

5

语言信息和语言通信

马大猷 著

知识出版社出版发行

(上海古北路650号)

(沪版)

新华书店上海发行所经销 上海东方印刷厂印刷

开本 850×1156毫米 1/32 印张 8.75 插页 2 字数 210,000

1987年6月第1版 1987年6月第1次印刷

印数：1—3,500

统一书号：ISBN $\frac{7-5015-5173-1}{H \cdot 17}$ 定价：2.10元

内 容 提 要

本书作者是著名的声学专家。在本书里作者首次把社会科学的语言学与物理学的语言声学及技术科学的信号处理技术很好的结合起来,是一部自然科学与社会科学有机组合的创新著作。

本书总结了语言信息和语言信号的处理技术方面的已有成果,展望在语言信号的传输的储存的前景。是定量地研究语言和语言通信工作者较理想的参考和学习材料。语言是人类最重要的交际工具,是人类社会的基础,本书也是广大读者了解语言、运用语言的有益读物。

作者职务

中国科学院学部委员、中国科学院声学研究所研究员

中国科学院数理学部副主任

中国声学学会名誉理事长、《声学学报》主编

美国声学学会会士

美国噪声控制工程协会通信会员

序 言

1982年冬，参加中国人民政治协商会议全国委员会五届五次会议的民盟小组委员，讨论如何开发盟内智力资源的问题，研究了由民盟中央发动盟内力量，筹备组织全国性的多学科学术讲座的倡议，深得与会委员的积极响应。随后中国民主同盟中央常务委员会决定从1983年暑期开始举办多学科学术讲座，并责成民盟中央文教委员会和科技委员会主持具体工作。经过半年的努力和筹备，讲座于1983年6月12日开学，开学典礼由民盟中央副主席楚图南同志主持，中共中央统战部杨静仁部长、李定副部长亲临指导。李定同志在讲话中肯定了民盟中央举办多学科学术讲座的重要意义，他指出，这不仅符合民盟盟员拥有大量学有专长的专家学者的智力集团的特点，而且也是民主党派工作中的一个创举，深得全国学术界的重视，是党的知识分子政策的胜利。知识分子通过讲座的形式，积极主动地把自己的智慧贡献给祖国的社会主义建设事业，反映了知识分子在党的教育下的又一次觉醒。

民盟中央1983年举办的讲座，分10个专题，每题十讲，每一专题，一般由主讲教授一人负责，也有少数专题由两位或三位主讲

教授共同负责。听讲者一千余人。他们来自全国各地,有不少人本身就是专家教授,有的已年逾花甲仍孜孜不倦为四化远涉千里来到北京,和中青年共同学习和学术交流,这也是前所罕见的现象,具体地反映了党中央自三中全会以来的正确领导,动员了千百万知识分子,初步实现了团结奋斗的新局面。

1983年的讲座内容,涉及人文科学、社会科学和自然科学的一些方面。人文科学方面有朱光潜、黄药眠、常任侠三位教授的《美学和中国美术史》,吴组缃和张毕来教授的《谈红楼梦和红学四论》,商承祚、陆宗达教授的《中国文字学和训诂学》。在社会科学方面,有关梦觉教授的《陈云同志的经济思想》,千家驹教授的《中国经济问题》及徐铸成教授的《新闻艺术》。在自然科学方面,有马大猷教授的《语言信息和语言通信》,叶培大教授的《光纤理论》,钱伟长教授的《广义变分原理》等。准备在1984年参加讲座的主讲教授有唐敖庆、余瑞璜、张文佑、费孝通、陶大镛等20余位。

1983年主讲教授14人,平均年龄76岁,最高年龄86岁。他们以年逾古稀的高龄,冒着酷暑,一丝不苟地为学员认真讲解,亲切座谈,深受广大学员的欢迎和爱戴。

这次讲座努力贯彻了百家争鸣的学术方针,提倡严肃的学术民主。主讲教授都能在尊重不同意见的同时,深入透彻地讲解自己的学术观点,有些主讲教授对那些学术上的不正之风,进行了认真、严肃而又满腔热忱的批评和教育。这反映了老一辈学者对当前学术界不正之风的否定而又负责的态度。殷切期望我们的中青年学术接班人,发扬良好学风。有的主讲教授就在同一讲座上,以友好的态度各自讲解分析了双方不同学术观点的矛盾,而不以自己的观点强加给对方,更不以自己的观点来打倒别人的不同观点。这样就能以人之长补己之短,就能达到不同观点的相互融化,逐步走上更高水平的学术境地,从而更有利于为我国社会主义建设服务。

讲座的讲解，都是各主讲教授长期或毕生从事的学术工作，还有的是当前在四化建设第一线战斗岗位上总结提出的主要贡献。主讲人对讲稿都做了充分的准备，在讲座中又通过听讲学者的学习讨论，再次进行增删修改，才最后定稿。现蒙知识出版社编为丛书，按讲题分别出版。希望本丛书对于我国学术工作，产生有益的影响。

钱 伟 长

1983年7月26日于北京

目 录

第一章 语言信息	1
§ 1.1 语言通信	1
§ 1.2 信息量——离散过程	5
2-1 定义	6
2-2 英语用字母的信息量	3
2-3 多余度	10
2-4 通道的信息容量	11
§ 1.3 信息量——连续过程	12
§ 1.4 语言信号表示方法	14
§ 1.5 信号的数字化和基本处理方法	16
5-1 采样	16
5-2 量化	18
5-3 傅里叶变换	21
5-4 π 变换	21
第二章 语言发声模型	24
§ 2.1 语声模拟	25
§ 2.2 语言发声理论	28
2-1 发声系统的研究	28
2-2 语言发声模型	32
§ 2.3 语言参数	38
3-1 面积函数和频谱	39
3-2 参数的变化范围	45
3-3 语言信号的主要参数	47
§ 2.4 语言信号的统计模型	47
2A 附录 声道共振	52

2A-1 波动方程	53
2A-2 均匀无损失管道	54
2A-3 两段均匀管道的组合	57
第三章 语声	59
§ 3.1 汉语普通话语声	59
1-1 元音	60
1-2 双元音和三元音	61
1-3 元音加韵尾	61
1-4 辅音	62
1-5 声调	64
§ 3.2 语声的物理特性	65
§ 3.3 普通话频谱	72
3-1 元音	72
3-2 辅音	75
3-3 语声长度	77
3-4 普通话平均频谱	79
§ 3.4 普通话语声的统计特性	81
4-1 音位统计	82
4-2 音节统计	84
4-3 语词统计	88
§ 3.5 语言形式的统计理论	92
5-1 Zipf 定律	92
5-2 离散瑞利分布	94
5-3 泊松分布	98
5-4 测不准原理	99
第四章 语言感知和质量评价	101
§ 4.1 人耳	101
§ 4.2 听觉神经系统	110
§ 4.3 清晰度试验	114
§ 4.4 清晰度和可懂度	119

4-1	句可懂度	119
4-2	音位清晰度和词可懂度	120
4-3	清晰度指数	121
4-4	影响清晰度的因素	124
§ 4.5	可懂度理论	127
第五章	语言信号处理	130
§ 5.1	短期平均处理	131
§ 5.2	短期平均零交叉率	138
§ 5.3	音调周期估计的平行处理方法	140
§ 5.4	短期自相关函数	143
§ 5.5	短期平均差值函数	146
§ 5.6	采样率的增减	147
6-1	减少采样率	148
6-2	增加采样率	149
§ 5.7	量化噪声	149
§ 5.8	自相关声码器	155
第六章	频谱分析与合成	157
§ 6.1	短期傅里叶变换	157
1-1	傅里叶变换观点	158
1-2	线性滤波观点	163
1-3	语言的短期分析与合成	165
§ 6.2	滤波器组	167
2-1	实际要求	167
2-2	模拟滤波器	171
2-3	数字滤波器	176
§ 6.3	音调检测	188
§ 6.4	用合成进行分析	189
4-1	与音调同步的频谱分析	190
4-2	极点零点分析	191
4-3	声门波	191

§ 6.5	分析合成系统	193
5-1	短期傅里叶变换的数字编码	194
5-2	通道声码器	195
5-3	相位声码器	199
5-4	语声声码器	202
5-5	共振峰声码器	203
§ 6.6	倒频谱	206
第七章	语言的线性预测	210
§ 7.1	线性预测分析	210
1-1	自相关法	212
1-2	协方差法	214
§ 7.2	LPC 在频率域中的意义	215
2-1	均方预测误差	215
2-2	预测误差	216
§ 7.3	线性预测声码器	219
第八章	计算机人机口语对话	223
§ 8.1	口语应答系统	224
1-1	设计口语应答系统的一般原则	226
1-2	一个多路输出的口语应答系统	228
1-3	联接共振峰编码词的语言合成	231
§ 8.2	讲话人识别系统	233
2-1	讲话人证实系统	236
2-2	讲话人辨认系统	241
§ 8.3	语言识别系统	244
3-1	单个数字识别系统	245
3-2	连续数字识别系统	246
3-3	大词汇单词识别系统	248
3-4	连续语言识别系统	250
§ 8.4	人机对话的前景	255
第九章	聋哑人的语言通信	257

第一章 语 言 信 息

§ 1.1 语 言 通 信

语言通信是说话人向听者传递信息的过程。语言是人类独特的功能、人类社会的神经系统。事实上,语言作为传递信息的工具是我们的祖先在从猿到人的过程中发展起来的,并且是这一过程的必要条件。恩格斯说^①：

我们的猿类祖先是一种社会化的动物,人,一切动物中最社会化的动物,显然不可能从一种非社会化的最近的祖先发展而来。随着手的发展、随着劳动而开始的人对自然的统治,在每一个新的进展中扩大了人的眼界。他们在自然对象中不断地发现新的、以往所不知道的属性。另一方面,劳动的发展必然促使社会成员更紧密地互相结合起来。因为它使互相帮助和共同协作的场合增多了,并且使每个人都清楚地意识到这种共同协作的好处。一句话,这些正在形成中的人,已经到了彼此间有些什么非说不可的地步了。需要产生了自己的器官:猿类不发达的喉头,由于音调的抑扬顿挫不断加多,缓慢地然而肯定地得到改造,而口部的器官也逐渐学会了发出一个个清晰的音节。

这里所指的非说不可的“什么”,就是信息;由于互通信息的需要发展了语言和语言器官,而发展了的语言器官更进一步发展了语言通信,这个道理讲得非常透彻、明了。因此,恩格斯又概括道^②：

① 《马克思恩格斯全集》，人民出版社，第20卷，511(1971)。

② 《马克思恩格斯全集》，人民出版社，第20卷，512(1971)。

语言是从劳动中并和劳动一起产生出来的。

自从有了人类就有语言通信和信息传递，而语言通信和信息传递是人类区别于其他一切动物并取得不断发展的主要功能。语言功能在人类历史上当然是逐渐发展起来的，开始可能只会叫喊，经过千百年才开始向系统的语言发展。有人猜想，早期人类在发明语言之前，头一个使用的通信工具可能是手势，在狩猎、农业、工艺、艺术发展以后，手不够用了，才转而使用可以听得到的、明确的发音声道的“手势”。但这是不对的。人类在形成中可能用过手势，但只能是个别手势，不成系统，不能称为“手势语言”，正好象动物也可以用某种叫声表示惊慌、痛苦、求爱等等，但那不能称为语言。语言需有基本语法规律，需要有基本词汇，只有这样才能系统地传达信息，进行抽象思考。系统地使用手势只是在语言形成以后，并且是以语言为根据的。斯大林说：有声语言或词的语言始终是人类社会的唯一语言（马克思主义与语言学问题）。

传递信息并不限于语言通信，在人类发展过程中不断出现新的信息传递方式。几千年前就出现了文字，这是记录有声语言的符号，是书面语言，有了文字，信息可以传递得更远，并且可以保留下来，传之后世。还有表示有声语言的手势、旗语、盲文等等，用这些信息可传递到听不见的远处、到聋哑人、盲人等。100年前发明了电报和电话，使通信距离不受限制，而且传递的信息不仅是语言和文字，图象和资料也可以远距离传送（传真）。近年电子技术和计算机技术的发展，使无线电通信、广播、电视、载波通信、电传、光纤通信、电子信件等，都一一成了现实，信息传递技术的发展真是一日千里，远远超过历史时期人们的想象力。现在人们已在谈论信息时代的问题了。但是信息的范围虽广（除语言信息、文字信息、图象信息等以外，还有生物信息、遗传信息、地震信息、宇宙信息等），语言通信还是最基本、最重要的信息传递方式。从每个人的日常生活直到国家大事、重大科学技术的发展，语言通信都起着

决定性的作用。列宁说：语言是人类最重要的交际工具，这是过去、现在和将来都不可否认的颠扑不破的真理。

随着语言通信的日益发展，如何在语言通信中更好地传递信息的问题就提到科学技术发展的日程上来了。什么叫做信息、如何度量，信息通道的容量如何度量，经济有效的编码过程，噪声特性和抗干扰问题，离散过程（文字材料）和连续过程（语言、音乐材料）的处理等，都是极待解决的问题。1948年山农（Claude E. Shannon）首先研究了这些问题，创造了崭新的学科——信息论（或称通信理论），开辟了新的科学领域。基本在同时温纳（Nobert Wiener，曾任清华大学教授）在创立“控制论”中也对信息论的基本原理做了重要贡献。现在信息论的应用范围已不限于通信问题，电子计算机、纯粹物理以及科学实验过程中，信息论也非常重要。

什么是信息？现代汉语词典说：“指用符号传送的报道，报道的内容是接收符号者预先不知道的。”韦氏大词典则说：“知识或情报的传递或接收。即可传递的事实，不同于整套思想或知识中的事实。数据、新闻、情报等，是由研究或观察所得的知识，……”任何报道、音信则称为消息。有人说，信息是原材料，只是数据的积累；而知识却是经过一定思考与讨论，用比较和分类方法把数据加以组织的。再进一步就可形成科学知识并形成科学定律。这是对上述定义的解释。但这些定义都是定性的，要它成为科学的定义，必须经过一定的思考、整理工作。在通信中谈语言和文字（书面语言）的信息，则要从信号（语言的波形和文字符号）中的信息多少（信息量）出发，进而研究信号源、信道、接收以及信号处理所起的作用。

在现代语言通信中，除对面谈话（包括厅堂中报告和文艺表演）、电话、载波电话、无线电通信等以外，语言信号处理占很大重量。这些包括语言的统计分析和频谱分析，这是对语言信号和语言通信质量的重要研究内容。还有语言信号的合成，包括按文字合

成,后者实际就是说话机器,现在有的火车站已用合成方法作出自动应答系统,回答旅客的问题了。手表、照相机以及一些计算机说出简单词句也已实现。把分析与合成联合起来就成为声码器,它把输入的语言信号分析为频谱分量或其他决定信号的参量,编为电码,经远地传送以后再合成语言信号。它的优点是因为所传送的是电码,比较简单,节省信道,抗干扰能力强,此外还可以把电码改变成为保密系统,大大提高信道的使用价值,在国防上和工商业中都很重要。以上都是已实现的系统,现在正在发展中的还有语言自动识别系统,已作成一些雏形。用机器自动识别语言(或辨认说话人)进而控制自然、控制机器是人们多年的幻想,中国有呼风唤雨的寓言,《一千零一夜》中还有“开门吧,芝麻”的故事。用自动识别指挥机器自动化所需词汇小,最简单,现在已有条件实现。根据讲话录音以辨认犯人,现在已有不少应用。不限人、无限词汇的语言自动识别,现在还是研究中的课题,将来如能实现,则语言打字机、语言电报机、排版机,甚至翻译机都不成问题了,人们的文化、政治生活将完全改观。这是进一步研究的目标。此外,语言通信工作还涉及聋哑人教育和设备问题,读书机(涉及图象识别问题)则对盲人是重大帮助。

语言通信涉及的学科很广,很多方面最后成果都是仪器设备,但语言通信工作决不仅是电子线路和电子计算机问题。这些方面的重大发展大大促进了语言通信的工作,但是光靠电子线路和计算机是作不出语言机器来的。只有充分进行与语言有关的声学、语言学、语音学、心理学、生理学等方面的研究工作,才能提供制造机器设备的基础。

本书的目的,就是从信息传递角度讨论语言信号和语言通信所涉及的问题,包括以下各方面:

- (1) 信息论的基本概念
- (2) 语言的产生和特性

- (3) 语言感知——听觉
- (4) 语言信号的数字化
- (5) 语言波的传输与储存
- (6) 语音分析
- (7) 语言合成
- (8) 声码器
- (9) 自动语言识别
- (10) 讲话人的识别
- (11) 语言和听觉缺陷的校正
- (12) 语言信号质量的提高

§ 1.2 信息量——离散过程

上述现代汉语词典和韦氏大词典中有关“信息”的定义,完全是为外行人写的。在科学上,一个名词要有一个严格的定义,名词本身可能是现有的,也可能是新起的,但必须重新给它确切的定义,这样才可以避免在某一科学范围内发生误会或含糊不清。两个学科也可能用同一名词,但各有各的明确定义,不致混淆。例如“根”在数学范围内和在植物学范围内都有确切的定义,数学中的根不会长而植物学中的根永远也不会是虚数!词意的唯一性,这是科学方法的特点。在科学中使用普通词时,这一点更重要,因此在科学中信息的定义是什么更要明确。

在科学语言中,给一个名词下定义有两种方法。数学家从一些仔细选择并陈述的公设出发,更复杂的东西则由这些初始公设导出并用同它们的关系确定。实验科学则用另一种,称为“运算的”定义。力、质量、速度等的定义,都是测量这些量所需的实验的简单描述。很多重要科学家都推重实验科学中的运算观点。一般说来,在科学语言中只宜引进有运算定义的量。无法给运算定义

的词,往往用不了多久就无用了。现在就从信息的确切定义开始。

2-1 定义

信息的定义是从统计分析得来的。先考虑一个非常简单的例子:

掷一个骰子。出来的结果可以有 6 种,可以推出这几种的概率都是相同的。开始时,对这个系统毫无信息。如果对这系统得到信息渐多,就可以肯定这六种结果中的一种要实现。初始问题中不确定度越大,作选择需要的信息量越大。总结:

初始情况: $I_0 = 0$, 概率相同的结果 $P_0 = 6$

最后情况: $I_1 \neq 0$, 选择结果 $P_1 = 1$

用 I 代表信息量,信息量的定义可取为

$$I_1 = \log_2 P_0 (\text{位}) \quad (1.1)$$

在骰子的例中 $I_1 = \log_2 6 = 2.6$ 。 $\log_2$ 表示以 2 为底的对数。信息量用对数的原因是只有这样,信息量才可以相加。例如,掷一个骰子,其结果可有 6 种,信息量为 $\log_2 6$ 。如果掷两个骰子,其结果可以有 $6 \times 6 = 36$ 种,信息量是 $\log_2 6 \times 6 = 2 \log_2 6$ 。掷两个骰子的信息量是一个骰子的两倍,这从常识上看来是完全合理的。所以信息量要用对数表示。

采用以 2 为底的对数则是较为简单的。在信息论中把信息量当作无量纲量(纯数)看待,可用二进制数的位数表示。例如,在一个问题中,一共有概率相同的几种选择,每一种相当于二进制数的一位,0 或 1 (阴或阳,正确或错误,开或关,……,事实上二进制数的起源是和中国的阴阳有关的)。选择的结果就有 $P = 2^n$ 种不同,而信息量

$$I = \log_2 P = n$$

不同的结果可用 n 位二进数辨别。例如,镍币一面是字,另一面是国徽,可用 0 和 1 代表。掷一个镍币的结果可以是 0 或 1,掷两个则可能是 00、01、10 或 11,如此类推,掷 n 个镍币,结果就可用 n 位

二进制数表示。用二进制的位 (bit) 表示很简单。

对数的底采用 2 并不是必然的,必要时也可以采用其他底值。在计算数学中有时采取 8 或 16 作底,而以八进制或十六进制的数字表示。如果底是 α ,信息量就是

$$I_{\alpha} = \log_{\alpha} P = \log_2 P / \log_2 \alpha = n / \log_2 \alpha \quad (1.2)$$

如 α 为 16, $\log_2 16 = 4$

$$I_{\alpha} = I / 4$$

十六进制所需的位数为二进制位数的 $1/4$ 。在计算机上可以 4 个二进制位成为一节,以节为单位就成为十六进制了。

日常使用的十进制, $\alpha = 10$, $\log_2 10 = 3.32$,

$$I_{10} = I / 3.32$$

二进制数的位数为十进制数的位数的 3.32 倍,知道 P 的常用对数 (以 10 为底的), 乘以 3.32 就成为二进制数的位数。二进制数折合成十进制数的方法如下。设二进制数为

$$\dots edcba$$

$a, b, c, d, e, \dots$ 都是 0 或 1。这个数的等值十进制数就是

$$a + b \cdot 2^1 + c \cdot 2^2 + d \cdot 2^3 + e \cdot 2^4 + \dots$$

倒过来,把十进制数折合成二进制数则连续用 2 除,得到的余数就组成 $\dots edcba$ 二进制数。一些对应关系列于下表:

二进制数	0	1	10	11	100	101	110
十进制数	0	1	2	3	4	5	6
二进制数	111	1000	10000	10100	100000	110100	
十进制数	7	8	16	20	32	52	

二进制数的加法很简单,相应的位加起来就是了,遇到 $1+1$ 则向上一位进 1。

如上所述,一个问题有 P 个同样可能出现的结果或选择时,每