

声码器通信

〔苏〕M.A. 萨波日科夫 B.Γ. 米哈伊洛夫 著

王世福 张志明 译

方再根 审校

机械工业出版社

声码器通信

[苏]M.A. 萨波日科夫 B.Γ. 米哈伊洛夫 著

王世福 张志明 译

方再根 审校

宇航出版社

内 容 简 介

本书是一部有关声码器通信理论和技术的专著。书中所讨论的问题均以近二十年来一些著名学者所发表的论著为依据。内容包括：语言信号理论和变换方面的典型材料，新的语言数字处理方法和压扩方法（线性预测、同态变换等），用于通信、电子计算机、医学以及语言控制的声码器部件（分析器与合成器）的技术说明和电路。

本书可供从事通信、语言信号处理以及其他有关专业的科技人员参考。

ВОКОДЕРНАЯ СВЯЗЬ

М.А. Сапожков

В.Г. Михайлов

Москва «Радио и связь» 1983

声码器通信

〔苏〕 М.А. 萨波日科夫 В.Г. 米哈伊洛夫

王世福 张志明

方再根 审校

宇航出版社出版

新华书店北京发行所发行 各地新华书店经售

空军学院印刷厂印装

开本：787×1092 1/32 印张：12.25 字数：267千字

1988年4月第1版 1988年4月第1次印刷 印数：1—4,000册

ISBN 7-80034-078-8/TN·005 定价：2.60元



译 者 的 话

《声码器通信》是莫斯科无线电与通信出版社1983年出版的一部有关声码器通信理论和技术方面的最新专著。

本书涉及内容广泛，从发音理论、语言信号处理到声码器的发展与应用，都有比较详尽的论述。书中还介绍了几种新的语言数字处理方法和压扩方法（线性预测、同态分析等），列举了许多试验数据。本书对从事声码器通信、语言信号处理以及其他有关科学技术领域的研究人员来说，不失为一部难得的和非常有价值的参考书。

本书由王世福和张志明共同翻译，由王世福负责全书的统一和校对。

凡原文中用俄文表示的单位和缩略语，译文中一律改用国际表示法。例如：Гц(赫)—Hz, дБ(分贝)—dB, бит/с (比特/秒)—bit/s, мс(毫秒)—ms, мкс(微秒)—μs, вх(输入)—in, вых(输出)—out, Г(元音)—V, С(辅音)—C, 等等。

在考虑专业术语的译名时，主要参考了以下工具书：R. R. K. 哈特曼、F. C. 斯托克著，黄长著等译，《语言与语言学词典》；汪德昭主编，《英汉声学词汇》。

北京工业学院方再根同志对本书内容进行了认真的审校，国防工业出版社王细李同志为本书的加工付出了辛勤的

劳动。此外，本书在翻译和出版过程中，还得到了其他有关领导和同志多方面的关心与帮助。在此，一并表示感谢。

由于译者水平有限，舛误和不妥之处在所难免，欢迎读者批评指正。

1986年3月

目 录

第一章 发音理论.....	(1)
第一节 概述.....	(1)
第二节 语言振动源.....	(6)
第三节 声道.....	(12)
第四节 语言信号的基本特性.....	(21)
第五节 语言信号的统计参数.....	(34)
第二章 语言的感知.....	(45)
第一节 概述.....	(45)
第二节 几种听觉特性.....	(49)
第三节 语言元素的感知.....	(58)
第四节 语言可懂度理论.....	(61)
第三章 语言信号的处理.....	(73)
第一节 概述.....	(73)
第二节 动态范围的压扩和限幅.....	(77)
第三节 频段的压扩和限幅.....	(93)
第四节 时间压扩.....	(94)
第五节 语言信号的数字处理法.....	(96)
第四章 声码器的构造原理.....	(105)
第一节 概述.....	(105)
第二节 语言信号模拟分析与合成的基本原理.....	(110)
第三节 语言信号数字分析与合成的基本原理.....	(124)

第四节	声码器的信道形成装置·····	(149)
第五章	语言振动源的分析与合成方法·····	(154)
第一节	基音和声-噪信号的统计特性·····	(154)
第二节	基音和声-噪信号的感知特点·····	(159)
第三节	基音和声-噪信号的提取方法·····	(164)
第四节	激励信号的形成方法·····	(183)
第六章	频谱-频带法·····	(191)
第一节	概述·····	(191)
第二节	声码器的主要部件·····	(194)
第三节	几种声码器信号的有效编码方法·····	(207)
第四节	几种声码器的电路·····	(212)
第七章	共振峰法·····	(223)
第一节	概述·····	(223)
第二节	共振峰分析法·····	(228)
第三节	共振峰合成法·····	(241)
第四节	共振峰声码器·····	(257)
第五节	共振峰声码器的质量和效率·····	(271)
第八章	正交法和频谱-时间法·····	(276)
第一节	正交法·····	(276)
第二节	相关法·····	(288)
第三节	同态分析·····	(295)
第四节	线性预测法·····	(297)
第五节	几种线性预测声码器的电路·····	(304)
第九章	语言元素法·····	(308)
第一节	语言元素变换的一般原理和特性·····	(308)
第二节	音位分析法·····	(311)

第三节	音位合成法·····	(325)
第四节	几种语言元素语言分析-合成法·····	(331)
第十章	声码器的应用·····	(337)
第一节	概述·····	(337)
第二节	声码器在电话通信系统中的应用·····	(339)
第三节	声码器在控制论、医学和科学研究中 的应用·····	(351)
总参考文献	·····	(361)
各章参考文献	·····	(362)

第一章 发 音 理 论^①

第一节 概 述

语言由声音、音节、单词、句子等组成。此外还采用了一些辅助性术语，例如，词素——具有语法和词汇意义的单词的一部分；句段——具有句法和语调-语义意义的句子的一部分（不少于一个单词）。就正常的语速而言，通常只在每一个句段之后出现不同长度的间歇。因此，句段中的所有话音都是连起来的。这是语言的一个重要特性。

每一个人都有自己的发音特点，即自己特有的发音方法。所有各种各样的话音，却可以被划分为有限个（30~50个）标准音位。确切地说，音位是存在于一系列具体话音——音位变体——中的该种语言的最小语音单位。谢尔巴院士是这样说明音位与话音间的关系的：“音位是人想要发出的音，而话音则是实际发出的音。”^②由于发句段时的连续性，出现了邻近音位间的相互干扰。这时所获得的音位发声

①本章内容以G.范特〔1〕、Л.В.邦达尔科〔10〕、Л.А.瓦尔沙夫斯基〔1.10〕等人的研究成果为依据。

②在本书中，除需要研究的是音位而不是声音的场合外，我们将一律采用“话音”这个术语。

变体称为位置变体。这种变体的数量比音位本身的数量多2~4倍。音位变体之所以如此种类繁多，是由于每个人发这些音位都具有自己特有的风格的缘故。

在同一种语言的字母和音位之间并没有一一对应的联系——尽管在许多情况下字母和音位是一致的。因此，在所有的语种中，音位数量和字母数量都是不一样的。

和字母一样，音位分为元音音位和辅音音位。辅音音位又分为浊辅音音位和清辅音音位。俄语中共有41个音位，其中有6个元音音位：а、о、у、э、и、ы。元音я、ю、ё、е在词首是й和与之对应的а、у、о、э的复合音，在其他位置则用于使其前面的一个辅音音位软化。

俄语中的大部分辅音都具有两种形式：硬辅音和软辅音（腭化音）。其中每一种形式都有与之对应的音位，因为它们不仅在发音上听起来有明显的差别，而且其客观特性也各不相同。辅音又可分为响声辅音和非响声辅音。响声辅音包括м、н、л、р的硬音和软音两种形式，非响声辅音包括其余所有的音。按照发音方法分类，辅音可分为擦音（ф、с、ш、х、в、э、ж）、塞音（п、т、к、б、г、д）、塞擦音（ц、ч）鼻音（м、н）和边音（р、л）。这些音中的ф、с、ш、х、п、т、к、ц、ч属于清辅音，其余的属于浊辅音。仅有硬音形式的音有ш、ж、ц，仅有软音形式的音有ч。й属于中性音位。辅音字母ш相当于ш和ч的双重音位。其余15个辅音各有15个硬音音位和15个软音音位。为了方便起见，我们把所有的软音音位都用软音符号表示（如фь、ть等）。这样分类不仅仅是为了准确地理解语音，重更要的是为了便于研究那些对语言进行技术与合成来说所必需的基本特性。

表1.1所列,为词汇量未加限制的俄语中音位出现概率的数据。但多数情况下利用的词汇量是有限的,因而这些词的音位配列可能与此表中的数据完全不同。

表 1. 1

音位	概 率 ($\times 10^4$)	音位	概 率 ($\times 10^4$)	音位	概 率 ($\times 10^4$)	音位	概 率 ($\times 10^4$)
А	1316	и	243	ль	162	ф	85
ь	977	ъ	240	ы	159	сь	85
т	602	и	232	у	153	ч	59
а	539	р	230	рь	133	мь	56
й	457	рь	221	з	130	бь	52
н	392	л	212	дь	126	пь	50
о	379	л	207	б	119	кь	36
с	359	м	202	х	102	зь	21
э	343	ц	197	г	91	фь	8
к	284	ть	196	ж	89	гь	7
в	273	д	177	вь	89	хь	5

注: А—非重读а; ь—非重读ä; ъ—非重读о。

不同音位的出现概率数据,可以说明其信息率:不常出现的音位比经常出现的音位含有较多的信息。我们知道,元音含有的信息就比辅音少得多,而辅音中的清辅音则含有更多的信息。例如,如果已知一个声音序列为“а.я.а”,则很难猜出这是一个什么词。但如果给定的序列为“п.м.тк”,

则不难猜出，这个词是“памятка”或者“пометка”。元音的主要作用在于重新调整发音器官以发出它后面的辅音。如果取消元音，则说话就会结结巴巴，并且会不由自主地在辅音之间出现一个不稳定元音。例如，在单词“Влтава”中的в后出现一个ы，变成了“Вылтава”。

人说话时在其周围空间所形成的声波可表示为由相应声源参数和人的口腔与鼻腔中复杂的共振系统参数所决定的函数。属于这种参数的有作为声辐射器的口腔和鼻腔的负载阻抗^[1]。声道可利用图1.1所示的电路类比。在这种表示法中，电路中的电流相当于声道中的体积速度 $U = vS$ ，式中， S ——声道截面面积； v ——该截面上的媒质微粒振动速度。四端网络输出端电压相当于声道输出端声压，即在声场中： $p = \phi(U_0, Z_{in})$ ，式中， Z_{in} ——声道输入声阻抗； U_0 ——四端网络输入端体积速度。

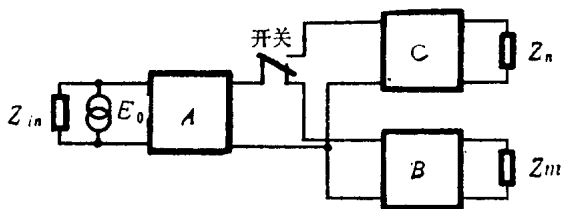


图 1.1 分别相当于喉腔(A)、口腔(B)和鼻腔(C)的四端网络形式的声道模型

对振动源的离散（音频）频谱来说，拉普拉斯变换形式

的声压可表示为^[1]：

$$p(s) = E(s)F(s) \quad (1.1)$$

式中， $s = -\delta \pm j\omega$ ——振动的复数角频率； δ ——衰减指数^①； E ——振动源生成函数； F ——传递（滤波）函数，即声道传输系数。傅里叶变换比拉普拉斯变换在形式上更为直观，但拉普拉斯变换能为分析语言过程提供更多的可能性。

公式（1.1）也可表示为振幅谱的形式

$$|p(f)| = |E(f)| |F(f)| \quad (1.2)$$

式中， f ——振动频率。对连续（噪声）频谱来说，需要一个相当窄的频带的中心频率代替公式（1.1）中的纯话音频率。

相应地，还可求出公式（1.1）中频率 f 的相位关系： $\Phi_p(f) = \Phi_E(f) + \Phi_F(f)$ 。在大多数情况下相位关系对语言的感知和识别没有明显的作用^[2]。

由公式（1.1）和（1.2）看出，声压可根据频率的变化而具有不同的值。一般说来，在一个语言频段内可以有几个音频成分和噪声成分。频谱包络也可以具有几个最大值和最小值，而且在许多情况下还具有零值。图1.2所示为生成函数、滤波函数和声场内的声波。这里既以频谱形式，也以时间形式对它们进行了表示。决定具体话音感知和识别的频谱最大值附近的频段称为共振峰。

①我们以G.范特有关言语构成的一般理论^[1]为依据。

② δ 通常为正值。

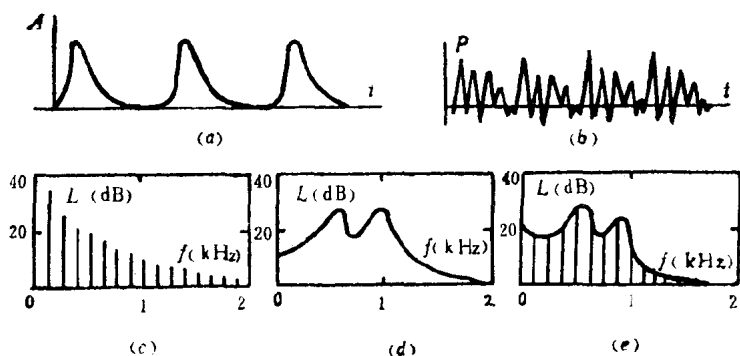


图 1.2 声道中的几个发音过程

- (a) 声带脉冲; (b) 辐射波; (c) 源频谱 $E(f)$;
 (d) 发送系统特性 $F(f)$; (e) 辐射音频谱 $P(f)$
 $= E(f)F(f)$ 。

第二节 语言振动源

语言振动源是包括声带在内的喉和肺中气流所通过的声道中的狭小缝隙。语言振动源按频谱形式分为离散振动源和连续振动源，按时间结构分为起伏振动源和脉冲振动源（包括单脉冲振动源）。

作为谐波振动源的喉，用来产生语音中的元音和浊辅音[⊖]。所有这些音有时也称为噪音[⊕]。应当指出，元音和浊

⊖ 正如以后将要证明的，喉是一个准周期振动源，因为在发音过程中脉冲起点间的间隔是不断变化的。

⊕ 以后我们将把元音和浊辅音统称为声音。

辅音虽然是由声带产生的,但声带并不象琴弦那样产生声波,而只是调制通过声带的气流。在声带下部形成的气压比在声带上部的气压高,于是声带就振动,一部分气流通过声带振动时所形成的小孔,使声带下部的气压降低,于是声带封闭,声带下部的气压再次升高,声带再次振动,照此类推,就形成了一个气流脉冲序列。声带开放的周期取决于声带的固有振动频率,因此可把声带振动和琴弦振动等同看待。气流脉冲的平均周期取决于声带的张力及其长度和厚度,也可以说取决于琴弦固有振动的参数。不过声带除基本振动外还具有一种附加振动(造成这种附加振动的原因之一是血液的流动)。

参考文献[1.1]中表明,不应当把声带看作是只具有一个质量的单共鸣腔系统,而应当看作是具有两个质量的系统。在这种情况下,由喉所产生的脉冲更接近于真实脉冲。这样就造成了偶脉冲和奇脉冲在波形和长度上的不同。至今还存在着一种观点,认为声门是发音器官滤波系统的一个组成部分,而振动源则位于声门下面。正如参考文献[1.2]中所证明的,这种观点并不是没有根据的。同时,认为声带是有声音振动源的观点也被一系列研究所证实。参考文献[1.2]中列举了足以证实这一观点的数据,并给出了喉的数学模型。从模型看出,话音形成的理论数据和试验数据是完全一致的。

由声源所产生的离散频谱的一次近似为纯谐波频谱。对应于单位时间内脉冲数的振动基频 $F_0 = 1/T$ 称为基音, T 为基音周期。在由声带所产生的脉冲中,除基本成分外,还含有大量谐波(40个以上)。谐波强度级随着频率的增加(即

随着诸波顺序号的增加)而平稳地降低。语言基音的特性将在第五章中详细讨论。这里仅指出基音的频段是相当宽的:最低男声为40~70Hz,最高童声为400~500Hz。

包括公式(1.1)在内的发音理论的主要假设条件是,发音器官生成函数和传递函数彼此间的独立性。如果喉阻抗相当大和相当稳定(不取决于时间和肺压),则对声源来说这一论点就可以完全被证实。但事实上,平均值相当大的喉阻抗在相当大的范围内和在每一个脉冲的发送过程中不断地变化着。因此,认为振动源与发音器官共振系统之间相互独立并不是十分准确的,尽管在一次近似中这种表示法比具有无限大内阻抗的振荡器表示法(即发电机表示法)精确得多。

根据参考文献[1],声门中空气微粒的振动速度为 $v = (2\Delta p / \rho k)^{1/2}$, 式中, $k = 0.875$; Δp ——喉内压差; ρ ——空气密度。喉对气流的阻抗可表示为两个成分(狭缝成分 R_s 和扰动成分 R_T)的和。狭缝成分 $R_s = 12\mu l^2 d / S^3$ 不取决于振动速度 v ,而是对应于声带间矩形狭缝的粘着阻抗,而扰动成分 $R_T = k_0 / v S$ 则与速度成比例。在以上公式中, μ ——粘度系数; $S = bl$ ——缝隙面积; b ——缝隙宽度; l ——缝隙长度, d ——缝隙深度。在2000Hz左右的频率上 R_s 和 R_T 两个成分的阻抗相等。因此,在低频上将产生擦音的分流效应^[1]。参考文献[1.1]中指出, R_s 成分并不十分精确,因为缝隙的宽度是与其长度一起变化的。

所有这些假设都已由多次试验和计算所证明。参考文献[1.2]中证实了喉头下腔对整个发音过程所起的作用。气流的固定成分和低频成分通过声门产生分流,频谱其余部分的功率进入声门时也有一部分损耗,因此在喉头等效电路中应

包括漏阻抗 $R_L = \rho v / S_0$ 。

根据这些数据，图 1.3 给出了声源脉冲的各种频谱包络形状。

响声辅音（有时称为半元音）主要由声带形成，即只有音频成分，而没有噪声成分。在非响声辅音的浊辅音中，噪声成分就功率来说近似于音频成分。

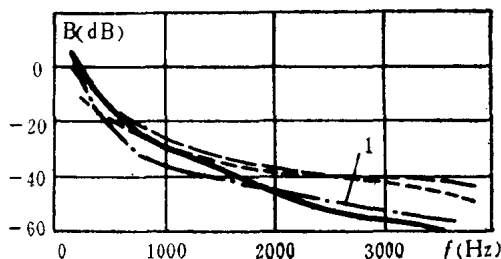


图 1.3 几个由声带产生的脉冲的频谱包络图
(曲线 1 表示斜率为 12dB/oct 的频谱包络)

所有的无声音和塞擦音仅由噪声成分构成，即当发这些音时声带实际上并不产生声音——尽管许多研究者都指出了存在不超过 30 Hz 的声带慢振动。我们知道，噪声源位于声道的收缩部位或其附近，噪声是由于气流的湍动而产生的。声源的精确位置取决于许多因素。气流通过声门时所产生的湍动既可产生噪声，又可产生复合音（后者是在发浊辅音时出现的情况）。

如果振动源正好位于声腔上，则振动具有周期性。由于