
张 刚 张雪英 马建芬 编著
语音处理与编码
兵器工业出版社
control

语音处理与编码

张　刚　张雪英　马建芬　编著

兵器工业出版社

内 容 简 介

本书分上、下两篇。上篇主要介绍有关语音信号数字处理的基本知识，包括：语音发声和听觉机理；数字模型；线性预测方法以及矢量量化等内容。并介绍了语音处理的三个应用领域：语音合成、语音识别及语音编码。下篇着重介绍了 90 年代以来语音编码领域的主要成果，包括三个主要国际标准 FS-1016、ITU-T G.728 和 ITU-T G.729。其中也有近年来作者的一些研究成果。

本书适用于高等院校电子工程、通讯工程和计算机专业本科生、研究生的教学参考书或教材。也可用作工程技术人员和科研人员的参考书。

图书在版编目（CIP）数据

语音处理与编码/张刚，张雪英，马建芬编著. —北京：兵器工业出版社，2000.8
ISBN 7-80132-835-3
Ⅰ.语 … Ⅱ.①张… ②张… ③马… Ⅲ.语音数据处理 Ⅳ.TN912.3
中国版本图书馆 CIP 数据核字(2000)第 67835 号

出版发行：兵器工业出版社
责任编辑：常小虹
责任技编：赵哲峰
社　　址：100089　北京市海淀区车道沟 10 号
经　　销：各地新华书店
印　　刷：太铁三校校办工厂
版　　次：2000 年 8 月第 1 版第 1 次印刷
印　　次：1—1000

封面设计：李增俊 吴国英
责任校对：程永强
开　　本：787×1092　1/16
责任印制：王京华
印　　张：22
字　　数：500 千字
定　　价：35.00 元

前　言

语音信号处理是一门涉及面很广的交叉学科，主要用于现代通信领域。语言是人类交换信息最方便、最快捷的一种方式，在高度发达的信息社会中，用数字化的方法进行语音的传送、储存、识别、合成和增强等是整个数字化通信网中最重要、最基本的组成部分之一。随着人类步入信息社会步伐的加快，越来越多的地方需要用到语音信号处理知识。目前，高校中的计算机和电子类专业已陆续开设这门课。编写本书的目的，就是为高校高年级本科生、研究生及有关科技人员提供一本参考书。

本书主要以研究生和高年级本科生为读者对象，注重语音信号处理基础知识的描述，并在此基础上，重点讲述语音编码方面的知识和最新成果。全书分上、下两篇。上篇前五章主要介绍语音信号产生的数字模型、语音处理的各种方法及矢量量化原理。上篇后三章是对前述内容的总结，介绍了语音处理的三个主要应用领域：语音合成、语音识别和语音编码。上篇内容可以用作工科高校电子专业 40～50 学时选修课教程。下篇着重介绍了最近十几年语音编码领域里发展起来的一个重要方法：码激励语音编码方法，详细阐述了 FS-1016、ITU-T G.728 以及 ITU-T G.729 三个码激励编码国际标准，同时将我们这几年在语音编码领域的最新科研成果及国际上该领域的发展状况融入其中。读者可以从中了解到 90 年代末话音编码领域的最新成果和主要概况。下篇内容可以作为通信专业及计算机应用专业硕士研究生 60 学时的专业课教程。通过阅读本书，读者可以掌握语音信号处理的基础知识，了解语音编码的最前沿知识。

参加本书编著工作的主要有张刚、张雪英和马建芬三位同志。其中，第 1、2、6、7、10 章及附录 A 主要由张刚编著，约 20 万字。第 14 章由张雪英和马建芬合作编写。第 5、8、11、12、13 章主要由张雪英编著，约 16 万字。第 3、4、9 章和附录 B 主要由马建芬编著，约 15 万字。由于编著者水平有限，书中难免存在错误之处，敬请读者批评指正，以在此书再版时更正。

本书的部分研究工作得到了山西省自然科学基金、山西省青年科学基金的资助。

编著者

2000.6

目 录

上 篇

下 篇

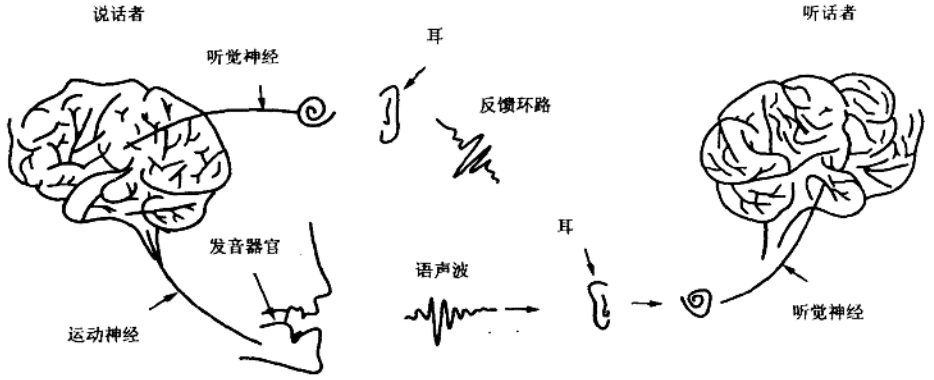
上　篇

§1　绪论

人们知道，声音和振动是一个事物的两个方面。动物通过发声与自然界沟通。如：虎、豹用怒吼震慑猎物敌手，蝙蝠利用超声波在黑暗中活动，鲸鱼等许多动物通过发声与同类交流。伴随着由猿到人的进化，人类的声音所承载的信息也不断丰富。到今天，语言是人类最重要、最有效、最常用和最方便的通信形式。

人类语言(language)主要表现形式有文字和声音。语音学是研究人类不同语言中发音与语义之间的相互关系及规律的学问。以数字信号处理方法为工具进行语音研究，是语音处理的基本任务，也是本书介绍的主要内容。

人的言语(speech)过程可以分为五个阶段。如图 1.1 所示。



想说阶段		说出阶段	传送阶段	接受阶段		理解阶段
机器 翻译	文语转换 语音分析成	语音编码 语音加密	公用电信网 移动通信网	语音解码 语音增强	语音识别 语音理解	机器 翻译

图 1.1　人的语言过程

(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。

⑵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。

⑶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。

⑷接收过程　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。

⑸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。

图 1.1 描述了人在言语过程五个阶段的主要生理关系。同时列出了与之相对应的科研内容和应用领域。除机器翻译属计算机人工智能学科外，其余三个主要应用领域都与数字信号处理有密切的直接关系。

§1.1　语音分析与合成

以语言信息压缩、存储为主要目的对语音信号数字模型进行研究，同时研究音素、音节、词组与句子的发音规则。最终恢复出自然流畅的语音来。例如文语转换系统(Text-to-speech)。语音分析与合成将赋予计算机说话的功能。也是进行话音编码、语音识别研究的基础。

§1.2　语音识别和理解

研究如何使计算机能够听懂人类的语言。以汉语语音为例：汉语约有 400 个音节，加上声调约 1200 个音调节，把这些语音信号的特征存储到计算机内，并与计算机接收到的汉语发音进行比较，找到特征相同的音节或音调节，这个过程就是语音识别。将识别出的音节序列转换成文字，就是语言理解。许多算法将理解过程溶入到识别中来提高识别的准确性。因此可以将语音识别与理解归入同一类应用。

§1.3 语音编码

为了充分利用频率资源有限的传输环境，在人与人之间直接进行语言交流的同时以实施语音数字传输为目标进行的计算机实时数据压缩和解压缩就是语音编码、解码。语音编码与文本到语音转换有两个主要区别。第一、前者是人与人之间的话音交流，要保留说话人的声音特征。后者是文本到声音的转换即计算机发声。它可以是标准播音员或其它声音。此外，前者不仅对压缩率和音质有要求，而且要求较低的编、解码延迟。而后者对处理帧长没有什么太严格的限制。

由图 1.1 可以看出，在每个阶段的交界处，科学研究表现出异常地活跃。由此也产生现代信息社会的朝阳产业。例如：本世纪后半叶到下个世纪中叶，世界范围发展最快，规模最大的当属通信产业。以语音信号数字处理为基础的语音编码解码研究,是通信产业的重要基础。

通信产业起源于 1874 年电话的发明。从那时起，通信产业大致发生了三次重大变革。第一次变革产生于七十年代初。1972 年 CCITT 组织公布了第一个语言编码标准 G.711。即对数 PCM 编码，由此开始，数字程控交换网络逐步淘汰了传统的模拟交换传输方式。第二次重大变革产生于八十年代末。1988 年欧共体 13 个国家数字移动特别工作组(GSM)制定了采用长时预测规则码激励的编码标准(13k bps RPE-LTP)。1989 年美国蜂窝通信工业协会(CITA)宣布了北美数字移动通信话音编码标准(8K bps 矢量和激励 VSELP)。从而确立了全球范围第二个传输网移动通信产业的崛起。第三次变革发生在世纪之交。以新兴的计算机因特网为基础的信息高速公路在全世界范围迅速发展。如何在 INTERNET 网上有效地传输话音成为产业界关注的焦点。IP 电话将使因特网成为第三个话音通信传输网。目前 IP 电话所用的话音编码标准有 G.723.1、G.728、G.729 等。这些标准各有长短。人们正在努力研究适合 IP 电话的新的编码算法。毫无疑问，一个具有低延迟、低码率、低复杂性、高音质的话音编码算法将是未来 IP 电话网络的奠基石。

本书主要介绍话音编码基础知识的基本原理。同时也包括在该领域内，国际国内一部分主要研究内容和我们近几年在话音编码领域的最新科研成果。全书分上下两篇。上篇主要介绍语音信号数字处理的基础知识。第二章到第五章介绍语音信号产生的数字模型、语音信号数字处理的各种方法和矢量量化的基本原理。上篇后三章从应用角度对这些基础知识进行了总结，介绍了语音处理三个主要领域：语音分析与合成、语音识别及语音编码。其中第八章可以作为下篇内容的一个引言或概论。下篇着重介绍了最近十来年语音编码研究领域发展起来的一个重要方法：码激励(CELP：Code Exciting Linear Predication)语音编码方法，详细阐述了 FS-1016、ITU-T G.728 及 ITU-T G.729 三个码激励编码国际标准。读者可以从中了解到 90 年代末话音编码领域的主要概况和最新成果。

本书上篇可以用作通信专业和计算机专业 40—50 学时的基础课教材。也可供自动化专业，电子技术专业高年级本科生作选修课用。下篇可以作通信专业高年级或研究生的专业课 60 学时教程。

§2 语音学与语音信号模型

本章讨论语音的发音和听觉机理，汉语语音的特点及语音的信号模型。这些都是进行语音处理研究的基础知识。

§2.1 语音的发音机理

我们可以先体会一下乐器的发音机理。有两类主要乐器：一类是小提琴、二胡等弦乐器。其发音机理是先由物体(弦)产生振动，引起空气振动产生声波。提琴手通过改变振动弦的长度来调整音调。另一类是笛子、小号等乐器，其发音机理是由气源(演奏者的嘴)产生气流并穿过一条狭长的有孔管道，气流引起笛膜振动产生声波，声波经过笛孔引起谐振。通过改变谐振孔开启的位置调整音调。语音的发音机理类似于管乐器。

§2.1.1 人的发音器官

人的发音气官由三个子系统组成：(1)肺和气管产生气源；(2)喉和声带(相当于笛膜)称为声门；(3)由咽腔、口腔、鼻腔组成的声道(相当于笛管)。肺的发音功能主要是产生压缩气体，通过气管传送到声音生成系统。

喉 喉是控制声带运动的软骨和肌肉的复杂系统。它主要包括下述几部分。

1．环状软骨
2．甲状软骨
3．杓状软骨
4．声带

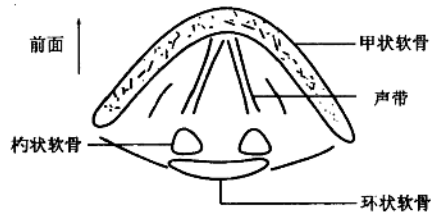


图 2.1.1 喉的平面解剖示意图

环状软骨和甲状软骨是主要结构。环状软骨结构上与构成气管的软骨环相似，但是后部则高得多，以支撑声带的后端。甲状软骨位于前端，基本上与环状软骨较高部分相对，其形状使它具有足够的强度以承受声带的拉力。位于喉前端呈圆形的甲状软骨称为喉结。

声带是伸展在喉前、后端之间的褶肉，如图 2.1.1 所示，前端由甲状软骨支撑，后端由杓状软骨支撑，而杓状软骨又与环状软骨较高部分相联。这些软骨在环状软骨上的肌肉的控制下，能将两片声带合拢或分离。声带之间的间隙称为声门。声带的声学功能主要是产生激励。

声道　声道包括喉以上的所有发音器官。图 2.1.2 所示径向截面图说明了声道的一般结构。声道通常分为以下几部分：

1．咽喉(位于会厌以下)；

2．口咽(位于舌之后，和软腭之间)；

3．鼻咽(位于软腭之上，鼻腔的后端)；

4．口腔(位于软腭之前，包括唇、舌和硬腭)；

5．鼻腔(位于软腭之上，从咽延伸到鼻孔)

此外声道还涉及下述器官：

1．会厌；

2．下腭；

3．舌；

4．软腭；

5．硬腭；

6．牙齿；

7．唇。

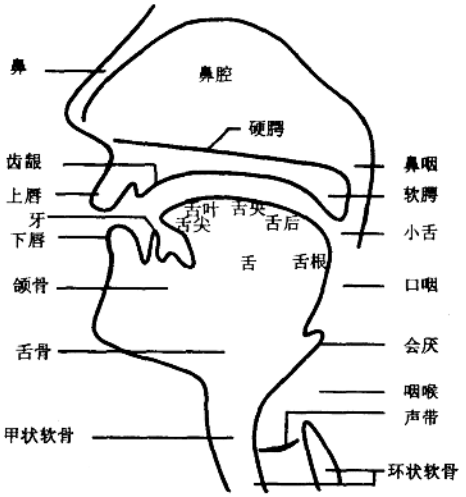


图 2.1.2　声道结构

会厌是位于声带之上舌之后的软骨床。下腭的作用是阻嚼食物和支撑舌的前端。口腔的上腭分为前上腭和后上腭两部分。前者由腭骨的硬上腭组成，起到支撑上牙和分隔口的作用，后者称为软腭由肌肉组织构成。其声学作用是封闭声道与鼻腔隔离。硬腭的前端是齿龈。

舌是一个大的肌肉系统。前端连接下颌，后端与喉骨和头骨相连。按发音功能，大体可分为几个区。能够伸出口腔外的部分称为舌叶，舌叶的正前端是舌尖；舌叶之后，硬腭下方部位称为舌央；软腭下方为舌后；与咽相对的部位称为舌根。

成年男性声道的长度为 17cm。当声波通过声道时，其频率高低受声腔共振的影响。这种共振与声道不同区段形状有关。声道的形状变化由舌、软腭、唇、牙所决定。

§2.1.2　语音生成

语音的生成动作可分为两种功能，即激励和调制，如图 2.1.3 所示，激励一般在声门处完成。当声波离开声道时还受到嘴唇幅射的作用。

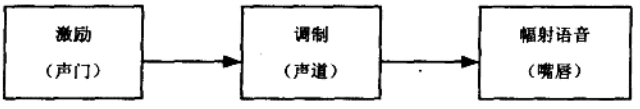


图 2.1.3　语音生成模型

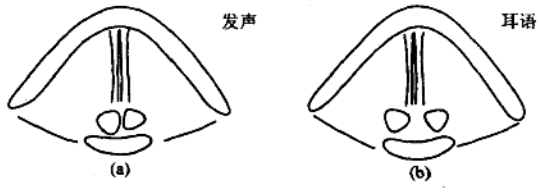


图 2.1.4 声带和软骨的位置

(a) 发声时的位置 (b)耳语时的位置

激励功能可以通过发声、耳语、摩擦、压缩和振动几种方式实现。

发声 由声带振动而产生的最重要的激励方式。杓状软骨受肌肉控制产生运动形成声带的闭合与张开，如图 2.1.4(a)所示。当气体被迫通过声带时就会产生振动。声门的闭合开启会产生周期性的脉冲气流，脉冲的频率称为基音。由声带振动发出的语音称为浊音，其它方式发音称为清音。

耳语 这时声带呈闭合状态，但杓状软骨之间开一个小三角口，如图 2.1.4(b)所示，空气通过小孔时产生湍流，从而产生一个宽带噪声作为激励信号。

摩擦 如果声道在任一处出现阻塞，通过该阻塞点的气流会引起湍流而产生宽带噪声，其频谱能反映阻塞点的位置。这样产生的声音称为擦音或咝音。

压缩 吸气后将声道完全关闭，形成压力，当声道突然打开会导致压力突然消失从而产生爆发气流。如果除阻突然发生，称为塞音或爆破音(例如[p]、[t])。如果除阻逐渐发生或产生湍流形成与擦音相联的爆发，称为塞擦音。

振动 如果空气流受迫点不是声带而是阻塞点，则产生振动如卷舌音 r。

对激励产生的气流再加入某些不同的信息，就是调制。调制过程主要通过口腔、鼻腔和咽腔不同位置和形状变化而产生 的。不同的声道形状具有不同的固有频率。语音的调制就是通过改变声道形状来产生不同的共振峰，从而产生不同的元音和辅音。

§2.2 语音的听觉机理

听觉是接受声音并将其转换成神经脉冲的过程。大脑受到听觉神经脉冲的刺激感知为确定的含意是一个非常复杂的过程，至今尚不清楚。

§2.2.1 听觉器官

人听觉器官分为三个部分：外耳、中耳和内耳，如图 2.2.1 所示外耳由耳廓(可见卷曲软骨)、外耳道和中耳(鼓膜)组成。耳廓的作用是保护耳孔。外耳道是一个直径约 0.7cm,长约 2.7cm 的耳管，声音通过耳管到达鼓膜。耳管也有许多共振频率，其中只有一个(约 3kHz)属音频

范围。鼓膜系统位于外耳道内端的韧性锥状物。它能随外界声音的振动将声音发送给内耳听觉神经。鼓膜将中耳与外耳分隔开。中耳是一个充气腔，并通过卵形窗和圆形窗这两个小孔与内耳相连。中耳还通过咽鼓管与外界相连，以便使中耳和周围大气之间的气压得到均衡。

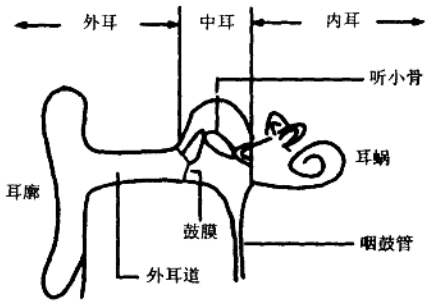


图 2.2.1 人耳结构示意图

中耳含有三条听小骨，分别称为锤骨、粘骨和镫骨，这三条听小骨的作用是建立鼓膜与卵形窗之间的声音耦合。锤骨与鼓膜相连，镫骨与卵形窗相连而粘骨则介于锤骨和镫骨之间。听小骨的功能有两个，即阻抗变换和限幅。

阻抗变换可以很有效地完成由空气至充液内耳的声能转换。三块听小骨机械连接优点起着类似杠杆的放大作用。可将声压增强约 22 倍。限幅可在大音量时保护耳免受损害。限幅由内耳肌拉动听小骨产生收缩使声音传输得以衰减。尤其是镫骨肌的收缩作用能改变镫骨的振动方向，从而降低了卵形窗的激励。内耳由前庭、圆形窗、卵形窗及耳蜗组成。前庭包括半规管和有关器官，它们用于平衡听觉方向。

对听觉起主要作用的器官是充满液体的耳蜗。声音从振动转换成神经脉冲就是在耳蜗内完成的。耳蜗高约 5mm，直径约为 3mm，自身绕成两圈半，拉直后长约 30~32mm。图 2.2.2 所示为拉直后耳蜗的纵剖面，横剖面和耳蜗的横断面。可以看到，除尖端外，整个耳蜗由基底膜和瑞士膜隔开成三个区域。中部区域称耳蜗导管，上、下两区域分别称为前庭阶和鼓阶。两区域在尖端部分相连。耳蜗导管中充满高粘度的胶状的内淋巴液。而相连的前庭阶和鼓阶内则充满粘度为水的 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。1960 年，G.Von.Bekesy 用正弦信号对基底膜进行了详细研究，发现基底膜对听觉的响应与刺激的频率有关。当频率较低时，靠近耳蜗尖部的基底膜响应。当频率较高时，靠近圆形窗的窄而紧的基底膜产生响应。基底膜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，从而刺激附近的听觉神经末梢，产生电化学脉冲，并沿听觉神经束传送到大脑。Bekesy 的研究成果最终使他获得了诺贝尔奖金。

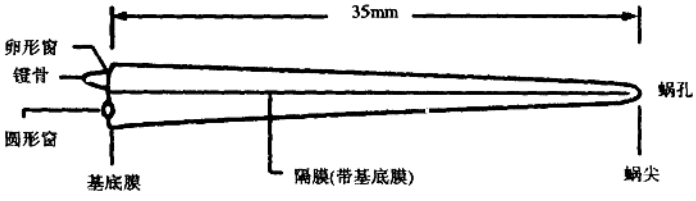


图 2.2.2 耳蜗伸展后的外观图

§2.2.2 语音的感知

语音的听觉感知是一个复杂的人脑——心理过程。对听觉感知的研究还很不成熟。听觉感知的试验主要还在测试响度、音高和掩蔽效应等。人耳听觉界限的频率范围大约为16Hz-16kHz。其上限随着年龄增大而降低。年青人可达 20kHz，老年人可低到 10kHz。在频率范围低端，感觉声音变成低频脉冲串，在高端感觉声音减小直至完全听不到一点声响。语音感知的强度范围是 0－130dB 声压级(基准声压级为 $10^{-10}w/cm^2$)，声音强度太高，感到难以忍受，强度太低则感到寂静无声。

响度 这是频率和强度级的函数。通常用响度(单位为宋)和响度级(单位为方)来表示。

人耳刚刚可以听到的声音强度，称为“听阈”。此时响度级定为零方。测量表明听阈值是随频率变化的。通常，人们把 1kHz 纯音听阈值定为零方。此时声强为 $10^{-16}w/cm^2$，这样的声波振动几乎不能使鼓膜离开它静止位置，可见人耳对声音是非常灵敏的。另一方面，加大声音的强度，使听起来令耳朵感到疼痛，这个阈值称为“痛阈”。测试表明对 1kHz 的纯音，当声强级大到 120dB 时，即声强为 $10^{-4}w/cm^2$ 会达到痛阈。可见人耳的听觉范围相当宽，相差 10^{12} 倍。

响度是同响度级有区别的。60 方响度级比 30 方响度级的声音要响，但没有响了一倍。响度是刻划数量关系的。2 宋响度要比 1 宋响度的声音响一倍。一宋响度被定义为 1kHz 纯音在声强级为 40dB 时(声强为 $10^{-12}w/cm^2$)的响度。

音高 音高也叫基音。物理单位为 Hz，主观感觉的音高单位是美。当声强级为 40dB(或响度级为 40 方)，频率为 1kHz 时，设定的音高为 1000 美。

响度与音高之间具有互为补充的关系。例如可以用频率补充声强使人们感觉到响度相同。也可以用声强补充频率使人感觉音高相同。

掩蔽效应 一个声音的听觉感受性受同时存在的另外一个声音的影响，这个现象称为人耳的“掩蔽效应”。此时前者称为被掩蔽音，后者称为掩蔽音。在掩蔽情况下，被隐蔽音的听阈会提高，即加大被掩蔽音的强度才能听到。此时听阈称为掩蔽听阈。

低频的纯音可以有效地掩蔽高频的纯音。对于中等掩蔽强度来说，纯音最有效的掩蔽出现在它的频率附近。

如果把噪音视为许多纯音组成的宽带音，掩蔽作用最明显的是被掩蔽纯音频率附近的一个窄带掩蔽分量。

利用人耳的掩蔽效应，在进行语音压缩时，让量化噪音的频谱跟随语言信号频谱包络变化。这时共振峰的频率成分就会掩蔽掉量化噪声。这个技术称为噪声整形或听觉加权处理。

§2.3 语音信号模型

由 §2.1 节介绍的发音机理和图 2.1.3 所示的语音生成模型可知，有三部分作用施加在语音的声波上。分别是由声门产生的激励函数 $G(z)$；由声道产生的调制函数 $V(z)$和由嘴唇产

生的幅射函数 $R(z)$。语音信号的传递函数由这三个函数级联而成，即：

$$H(z)=G(z)V(z)R(z)$$

下面各小节，我们将建立这三个函数的数学表达。

§2.3.1 激励模型

发浊音时，由于声门不断开启和关闭，产生间隙的脉冲。经仪器测试它类似与斜三角形的脉冲。也就是说，这时的激励波是一个以基音周期为周期的斜三角脉冲串。

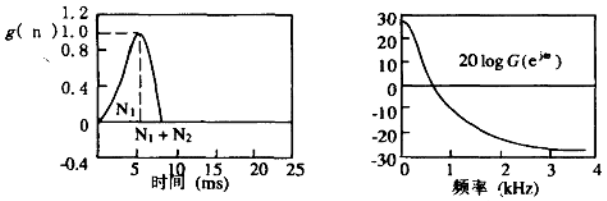


图 2.3.1 斜三角波及其频谱

如图 2.3.1 所示，单个三角形波的数学表达式为

$$g(n)=\begin{cases}\dfrac{1}{2}\left[1-\cos\dfrac{n\pi}{N_1}\right] & 0\leqslant n\leqslant N_1\\ \cos\left[\dfrac{n-N_1}{2N_2}\pi\right] & N_1\leqslant n\leqslant N_1+N_2\\ 0 & \text{其它}\end{cases}$$

式中，N_1 为斜三角波的上升时间，N_2 为其下降时间，由图 2.3.1 可以看出单个斜三角波的频谱 $G(e^{j\omega})$表现出一个低通滤波器的特性。可以把它表示成 z 变换的全极点形式：

$$G(z)=\frac{1}{\left(1-e^{-cT}\cdot Z^{-1}\right)^2}$$

这里 c 是一个常数，显然上式表示斜三角波形可以描述为一个二极点模型。因此，作为激励的斜三角波串可以用一串加了权的单位脉冲序列去激励上述单位斜三角波模型实现。这个单位脉冲串和幅值因子可以表示成下面的 Z 变换形式

$$E(z)=\frac{A_v}{1-z^{-1}}$$

所以整个激励模型可表示为

$$U(z)=\frac{A_v}{1-Z^{-1}}\cdot\frac{1}{\left(1-e^{-cT}z^{-1}\right)^2}$$

在发清音的场合，声道被阻碍形成湍流。所以可以模拟成随机白噪声。因此，可将激励模型表示成图 2.3.2 的结构

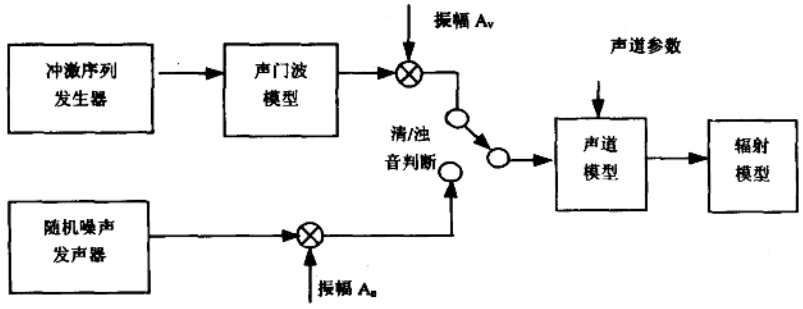


图 2.3.2　二元激励的语音发声模型

图 2.3.2 所示的二元激励模型，在早期语言信号处理研究中使用了许多年。尽管人们认识到二元激励过于简化，对于鼻音和擦音，模型还应考虑零点。对于浊擦音需要浊音和清音两种激励,并且两种激励不是简单的叠加关系。但是直到八十年代中期开始，新的激励模型才开始取代二元激励模型。

八十年代中期，人们开始在一个基音周期内采用多个脉冲来构造激励模型。新的激励方法本质上可以归结为存储器模型。就是说将可能的各种激励预先放在存储器内，通过某种判据决定哪一种激励是当前信号的最佳激励，并把这个最佳激励的存储地址作为激励的表征。例如码激励模型或矢量激励模型等。存储器内容随时间变化的部分称为自适应码书。自适应码书的搜索等价于基音检测。

§2.3.2　辐射模型

从声道模型输出的是速度波 $u_l(n)$，而语音信号是声压波 $P_l(n)$。二者倒比称为辐射阻抗 Z_l 它表征口唇的辐射效应。如果认为口唇张开的面积远远小于头部的表面积，利用单板开槽辐射的处理方法，可以得到辐射阻抗

$$Z_l(\Omega)=\frac{j\Omega L_r R_r}{R_r+j\Omega Lr}=R_0\left(1-z^{-1}\right)$$

式中：

$$R_r=\frac{128}{9\pi^2},\qquad L_r=\frac{8a}{3\pi c}$$

这里 a 是口唇张开的半径，c 是声波传播速度。由辐射引起的能量损耗正比于辐射阻抗的实部，其频响曲线表现出一阶高通滤波器的特性。在实际信号分析时，常用所谓预加重技术，