SPEECH CODING

Speech Coding

语音编码

王炳锡 编著

Speech Coding

语 音 编 码

王炳锡　编著

西安电子科技大学出版社
2002

内 容 简 介

本书从人的发音机理和听觉机理出发，以信号模型为基础，在介绍现代信号分析和压缩编码的基本理论之后，沿着激励和声道两条线逐步深入研究，以标准算法为范例，系统地阐述了语音信号压缩编码的基本原理。读者可以沿着这两条线看到各个阶段有代表性的语音信号压缩编码(声码器)的算法分析和典型设备的构成。由于书中收集了大量的实际参数和图表，因此具有较强的可操作性和参考价值。

全书共分 12 章。前 5 章是语音信号、信号模型和语音信号数字处理的基础知识，后 7 章是语音信号编码和各种典型的(公布的各种标准)压缩编码算法分析。每章后附有复习思考题或习题，供复习巩固之用。另附有参考文献供深入研究考证。书末附有英汉名词对照，供阅读外文资料时参考。

本书可作为高等学校理工科通信和信息处理及相关专业的高年级本科生和(硕士、博士)研究生的教材和参考书，也可供相关专业的工程技术人员和科研人员参考。

图书在版编目(CIP)数据

语音编码/王炳锡编著．—西安：西安电子科技大学出版社，2002.7

ISBN 7-5606-1121-4

Ⅰ.语… Ⅱ.王… Ⅲ.语音数据处理-编码 Ⅳ.TN912.3

中国版本图书馆 CIP 数据核字(2002)第 036363 号

策　　划　臧延新　陈宇光

责任编辑　龙　晖

出版发行　西安电子科技大学出版社(西安市太白南路 2 号)

电　　话　(029)8227828　　邮　编　710071

http://www.xduph.com　　E-mail：xdupfxb@pub.xaonline.com

经　　销　新华书店

印　　刷　陕西画报社印刷厂

版　　次　2002 年 6 月第 1 版　2003 年 4 月第 2 次印刷

开　　本　787 毫米×1092 毫米　1/16　印张 19.75

字　　数　465 千字

印　　数　4 001～8 000 册

定　　价　24.00 元

ISBN 7-5606-1121-4/TN·0201

XDUP 1392001-2

前　言

语音信号压缩编码是语音信号处理的一个方面，它和通信领域联系最为密切。而语音识别、语音合成、语音增强等方面在理论和方法上与语音信号压缩编码有很多相通之处。因此，系统、全面地掌握当今语音信号压缩的原理和方法，对语音信号处理领域工作的开展具有重要意义。

在多年教学与科研实践中，编者试图将语音信号压缩编码的研究和发展系统化，从中理出规律性的东西，给人以启示。

本书从人的发音机理和听觉机理出发，以信号模型为基础，在介绍现代信号分析和压缩编码的基本理论之后，沿着激励和声道两条线逐步深入研究，以标准算法为范例，构成了语音信号压缩编码的发展史。读者可以沿着这两条线看到各个阶段有代表性的语音信号压缩编码(声码器)的算法分析、典型设备的构成。由于书中收集了大量实际参数和图表，编者的同事们对每个系统认真地做了计算机模拟，在理论和实践的结合上做了初步的尝试，因此本书具有较强的可操作性和参考价值，对从事语音信号处理的本科生、研究生及广大工程技术人员会有所帮助。

本书内容安排如下：第 1 章为语言、语音及语音信号，第 2 章为语音信号的模型，第 3 章为语音信号时域处理方法，第 4 章为语音信号的线性预测分析，第 5 章为语音信号的矢量量化，第 6 章为语音信号编码，第 7 章为线性预测声码器，第 8 章为多脉冲激励及规则脉冲激励线性预测声码器，第 9 章为码激励线性预测声码器，第 10 章为对结构代数码激励线性预测语音压缩编码，第 11 章为多带激励声码器，第 12 章为混合激励线性预测声码器。每章后面都附有复习思考题或习题，供复习巩固之用，参考文献供深入研究之用。书末附有英汉名词对照表，供阅读外文资料时参考。本书对本科高年级学生讲授需 80 学时，对硕士研究生讲授需 60 学时。

本书的编写得到了解放军信息工程大学各级领导的关心和支持，得到了国内广大学者的支持和帮助。尤其是课题组同志和研究生们的研究成果，充实、完善了本书的内容，增强了本书的可读性和可操作性。西安电子科技大学易克初教授对本书提出了宝贵的修改意见，在此表示衷心的感谢。书中引用了大量的文献资料，在此向原作者表示深深的谢意。

虽经本人再三努力，但因水平有限，错漏之处在所难免，望读者不吝赐教，在此表示衷心的感谢。本书算法程序及一些语音处理工具已制成光盘，需要的读者请与作者联系，联系地址为河南省郑州市 1001 信箱 837 号(邮编：450002)。

王炳锡

2001 年 12 月

于解放军信息工程大学

目　　录

第 0 章　绪　　论

第 1 章　语言、语音及语音信号

第 2 章　语音信号的模型

第 3 章　语音信号时域处理方法

第 4 章 语音信号的线性预测分析

第 5 章 语音信号的矢量量化

第 6 章 语音信号编码

第 7 章 线性预测声码器

第 8 章 多脉冲激励及规则脉冲激励线性预测声码器

第 9 章　码激励线性预测声码器

第 10 章　对结构代数码激励线性预测语音压缩编码

第 11 章　多带激励声码器

第 12 章　混合激励线性预测声码器

第0章 绪 论[①]

0.1 概 述

21世纪的通信应在人与人之间、人与机器之间提供高质量的无缝的信息交换手段。无论何时、何地，以任何方式通信，语音通信将是最基本、最重要的方式之一。多媒体信息交换包括电话、电视电话会议、可视电话、语音信箱、电子邮件、图像传真、数据等等。无缝通信是指用户可方便地综合使用这些手段，而不影响通信质量，并能随意地把一种通信手段转换为另一种通信手段；高质量是指通信质量不随用户环境及传输媒介的变化而降低，用户使用起来方便快捷[1]。这取决于信息高速公路的建设和计算机、微电子、材料、网络、通信等诸多关键科学领域的发展，而语音压缩编码将是最基本、最重要的技术。这是因为最终产生信息、获取信息的是人，而人是以语音作为主要通信手段的。话带语音压缩编码领域的研究已有几十年的历史。近10余年来，人们对这一领域的研究兴趣大大地增长，已有大量的技术应用于远程通信和存储。一些国家和国际标准化组织相继制定了语音压缩编码的标准，直接推动了语音压缩编码的发展。家用和专业数字音响取得了商业成功。在市场牵动下，高保真音频压缩在近几年发展也很快。在通信系统中为了节省带宽，以及在语音存储系统中节省存储空间，音频信号的压缩编码技术有大幅度的发展[2]，音频带宽也从3.2 kHz(200 Hz～3.4 kHz)的话带发展到7 kHz会议电视宽带的语音压缩和20 kHz音乐宽带音频信号压缩，尤其是DHDTV研究开发的AC－3方案，因其多声道、立体声等高保真特点，已被美国联邦通信委员会(FCC)采纳。因此，语音压缩编码的发展在频带方面可归纳为三部分：3.2 kHz话带、7 kHz电视话带和20 kHz音乐话带。从应用角度看，语音压缩编码已在有线/无线电话、会议电视和HDTV以及高保真音乐等领域有广泛的应用。

0.2 回顾与总结

(1) 语音压缩编码的发展，一直是在用尽可能低的数码率获得尽可能好的合成语音质量的矛盾中发展的。数码率实质上反映的是频带宽度，降低数码率实质上是压缩频带宽度。当然随着数码率的降低，相应的算法延迟时间和计算复杂度也要增加。

在半个多世纪的研究中，各国学者做出大量的努力，从人类发音机理和听觉机理出发，对语音的基本元素的声学特性、频谱特征和语意表达等做了大量研究，建立了发音模

① 本文三个部分是作者在1997年11月“第八届全国语音、图像、通信信号处理学术会议”上的主题报告，并作了补充。

型、听觉模型，在不同程度上逼近真正的语言过程，并取得了长足的进展，逐步形成了通信和信息处理学科的重要研究方向，所以系统、科学地对语音压缩编码回顾和总结是十分必要的。在语音压缩编码的发展过程中，在众多的理论和技术中，以各种语音压缩编码标准为基准，研究其历史沿革有事半功倍之效果。它作为技术标准，至少代表了当时的技术最高水平，是技术成熟完善的标志，同时经过标准的制定，对技术又是一个很好的指导和激励。下面就我们熟悉的情况作简要回顾与总结。

自从1939年美国的Homer Dudley发明声码器以来，语音处理开始了参数编码或模型编码的研究，它是以滤波器为主构造的通道声码器。20世纪60年代以前，研究出实用的共振峰声码器。Sato，Itakura(1966)和Atal，Schroeder(1967)[3]最早把"线性预测(LPC)"技术应用到语音分析和合成。他们以线性组合模型均方误差最小意义下逼近原始波形的方法提取参数，研究出自相关法、协方差法、格型法等实用快速算法。1966年，J. L. Flanagan提出了以瞬时频率为基础的相位声码器[4]。1969年，A. V. Oppenheim提出了以倒谱为基础的同态声码器[5]。在众多声码器中，LPC声码器终因其成熟的算法和参数的精确估计成为研究的主流，并逐步走向实用。1982年，美国国家安全局(NSA)公布了2.4 kb/s的LPC－10声码器标准(FS－1015)；1984年，美国国防部制定了STU－Ⅲ计划，采用2.4 kb/s的LPC－10e增强型，1986年正式投入使用，这可以说是50年的研究总结。

(2) 从1985年B. S. Atal和M. R. Schroeder提出CELP算法以来，闭环分析算法(LPABS)成为主流。美国国防部公布了4.8 kb/s CELP联邦标准(FS－1016)。欧洲电讯管理局(GSM)于1988年公布了13 kb/s RPE－LTP线性预测语音编码方案。1989年，北美蜂窝电话工业组织(CTUA)公布了IS－54，8 kb/s矢量和激励线性预测(VSELP)语音编码方案(日本：6.7 kb/s)。1992年，CCITT公布了G.728 16 kb/s短时延码激励线性预测语音编码(LD－CELP)方案，1995年公布了G.723 5.3/6.3 kb/s ACELP/MLQ双速率多媒体语音编码标准，1996年公布了G.729 8 kb/s CS－ACELP对结构代数码激励的语音编码标准。

在这10年中就产生了3个国际标准、2个地区性标准和2个国家标准，可见语音压缩编码的研究发展之快。这些算法的共同特点是采用闭环LPABS算法、知觉加权技术、复合窗技术、LSP(LSF)技术、后置滤波技术、增益自适应技术、分数基音内插技术等。另外，多带激励(MBE)(1988年)[6]、自适应变换编码(ATC)(1977年)[7]和子带编码等语音压缩编码的实用方案在最近也有报道，它们都属于正弦编码。国际海事卫星组织(Inmarsat)于1990年公布了4.15 kb/s改进型多带激励(IMBE)语音编码标准。因此，在这10年中，CELP算法是语音压缩编码的主流。

(3) 近年来随着第三代移动通信的发展，变速率语音压缩编码技术相应得到发展。为了充分利用CDMA技术，Qualcomm于1993年提出了可变速率的CELP，通常称为QCELP[8]。它有4个可供选择的传输速率(1，2，4，8 kb/s)，通过计算输入能量，并与三个阈值能量比较来选择传输速率。这种技术已成为北美数字蜂窝通信标准(CTIA－is95)。1999年公布的第三代伙伴计划[9](3rd Generation Partnership Project)把自适应多速率(AMR)语音编解码作为主要技术。该技术有8种速率(12.2，10.2，7.95，7.40，6.70，5.90，5.15，4.75 kb/s)供选择，并采用语音激活技术(VAD)、舒适背景噪声(CNA)、源控速率(SCR)、重帧及误码消除(ECU)、抗稀疏处理等先进技术。它能根据信道质量选择

不同的编码速率，通信质量接近或达到长途电话质量。变速率语音压缩编码理论上仍属于CELP，但在“变”上有新的研究，引入了相关的先进技术。随着因特网的发展，语音IP[10](VoIP)对语音压缩编码的需求十分迫切。在H.323系列建议中规定了音频编/解码算法符合ITU标准，如G.711(A律或μ律)、G.722、G.723.1、G.728、G.729A等。但问题在于数据包在网上传送延迟时间有时太长(ITU-T G.192建议环路延迟应保持在300 ms以下[11])，会影响VoIP作为电话的使用。更低码率的声码器语音识别声码器可做到600 b/s，这点已有报道，并有待更深入研究。主要利用相邻帧间的语音频谱特征的变化程度作为语音相似程度的衡量依据，通过逐帧跟踪共振峰的变化来确定帧参数的发送，此算法码率做到600 b/s[12]，但有些音已不可懂，若采用帧间插值的算法会更精确。

(4) 高保真音频信号压缩编码，即带宽15～20 kHz的家用、专业高保真音响，包括动画和HDTV音频以及多媒体系统。有时音频编码这一术语也指宽带语音编码，即带宽7 kHz的语音信号、电视以及ISDN上的语音通信。

20世纪70年代早期，英国广播公司(BBC)研制了一种NICAM(Nearly Instaneous Companding Audio Multiplex)技术，它用于数字音频传输。最初由英国BBC应用于MAC制卫星直播电视的声音通道，后以多工复用的方式用到模拟电视广播中，并被EBU(欧广联)和ITU(国际电联)推荐为电视(地面广播和CATV)附加数字声的标准。1986年，Princen和Bradley提出了TDAC(Time Domain Aliascancellation)变换编码技术，它的特点是采用一个全重叠式的窗口结合子带和变换编码，变换后的信号成分一般要进行标量量化和熵编码，并根据听觉掩蔽模型自适应地为各频谱成分分配比特数。Davidson等人研制的ASPEC(Adaptive Spectrum Perceptual Entropy Coding)用于对高质量的音乐信号进行变换编码。Dolby实验室的Godd等研究的AC-3方案，被美国联邦通信委员会(FCC)采纳为高清晰度电视(HDTV)中多通道的音频部分的标准。它采用TDAC滤波和感觉掩蔽模型、变频率分辨率及混合前/后向自适应动态比特分配。它是第一个专门为编码多声道数字音频信号而设计的感觉编码系统。它可以满足单声道到5.1声道数字音频的编码要求，可恢复出具有非并行空间现实的声音效果，并在电影工业中得到了成功的应用。Theile等人研制出的一种称为自适应掩蔽模型的子带编码与复接(MASCAM)技术和欧洲数字音频广播技术标准(MUSICAM)都是以子带编码为基础的。

目前国际上音频压缩算法主要集中在ISO-MPEG音频编码标准。MPEG工作组成立于1988年，1992年11月完成了它的第一个标准MPEG-1。采样率为32，44.1，48 kHz，支持单声道模式、立体声模式、双声道立体声模式、联合立体声模式等4种模式，音频编码组合了MUSICAM和ASPEC的特点，提供3个编码层，单路32 kb/s，立体声第一层为384～448 kb/s，第二层为192～384 kb/s，第三层为128 kb/s。1994年11月完成了它的第二个标准MPEG-2，在与MPEG-1兼容的基础上实现了低码率和多声道扩展。1997年4月完成的MPEG-2 AAC(Advanced Audio Coding)对低至64 kb/s/ch的多声道编码，它都能提供相当高的声音质量。经测试，AAC标准以320 kb/s的数码率传送五声道全频带的音频信号，比MPEG-2以640 kb/s的数码率传送的音质还略好一些，达到ITU-R音质。MPEG-4是ISO/IEC于1998年11月完成的，1999年1月成为国际标准。MPEG-4的制定是基于数字电视、交互式图像以及万维网(World Wide Web)领域的成就而进行的。其音频部分将音频的合成编码与自然编码相结合。合成部分的组成工具可以实现对音

乐和语音按符号进行定义，它包括 MIDI 系统和文本—语音转换系统。另外，它还包括对声音的三维空间定位工具，可以利用人工声源或自然声源人为地制造出声音环境。MPEG－4 的音频部分对比特率在 2～64 kb/s 范围内的自然音频进行了标准化。为了在充分利用比特率的条件下获得最好的音频质量，标准定义了三种类型的编/解码器：用于低比特率的参数编/解码器，用于中比特率的 CELP 编/解码器，以及用于高比特率的时频(TF)编/解码器，包括 AAC 和基于矢量量化器的编/解码器。MPEG－4 的新颖之处是利用音素信息对唇的同步控制，在语音编码和图像编码的有机结合上迈出了可喜的一步，为声、图联合编码数据融合给出了一个范例。

家用高保真声频产品 DCC(音频压缩磁带)和 MiniDisc 都用到了感知掩蔽方法，DCC 用的是精度自适应子带编码，MiniDisc 系统用的是自适应变换声学编码(ATRAC)。

由上所述，子带编码是宽带语音压缩的主要技术。因为对传输时延有较高的要求，因此 ISDN 和会议电视大部分都采用 CELP 作一些感知加权和延时约束。

0.3 面 临 的 问 题

编者认为，语音编码面临的问题有四个。一是极低数码率，二是低速率语音编码合成语音音质要有更好的自然度，三是声码器在高背景噪声环境下的使用，四是经多次音频转接仍能正常使用。根据信息论的观点，语音压缩编码的码率可以做到 150～60 b/s。也就是说，语音压缩编码的工作空间还很大。从语音信号分析看，信息量最大的部分是在辅音和元音衔接的过渡部分，辅音和元音相比，辅音所携带的信息更多一些。而这些部分信息的提取仍需作出更大的努力。编者认为只有新的理论方法的应用，在这些方面才能取得突破，语音编码才会有一个飞跃。其他问题的研究必须和第一个问题综合考虑，是语音编码技术实用化必须解决的课题。

另外，对语音编解码器(声码器)的性能评价方法研究也是一个重要的研究课题。其本身不是语音编码问题，但和语音编码密切相关。评价声码器的性能好坏，需要进行多种指标的测试和评估，目前没有统一的国际标准，但普遍认为至少应包括编码速率、合成语音质量、顽健性、编解码延时、误码容限、计算复杂度和算法可扩展性等 7 个方面。其中对合成语音质量的测试和评价是最难的。合成语音的质量主观评价方法经各国学者多年研究，有判断韵字测试(Diagnostic Rhyme Test，DRT)，用来衡量声码器的可懂度，用百分数表示；有平均意见分(Mean Opinion Score，MOS)，用来对声码器的通话满意度(即自然度)和可辨识说话人能力给予整体综合评价，用 5 分制表示。中科院声学所在这方面做过大量的工作，在汉语语音的测试方面做出了重要贡献。但主观评价对测试环境、听音人群的条件要求很高。做一次标准测试，费时、费力，费用很高，往往受听音人的生理、心理、情绪和文化水平诸因素的影响，测试的重复性较差，严重影响了语音压缩编解码器的研究和发展。为了推动声码器研究，吸引更多的学者参与，研究客观评价方法，找到一种主、客观评价对应关系，国内外已有人着手研究并有相关报道，但很不成熟。这的确是一个很基础、很重要的研究方向，应该给予足够的重视。

参 考 文 献

[1] 王仁华. 面向2000年通信的语音处理技术. 中兴新通讯, 1996, 2(1): 40～43

[2] Allen Gersho. Advances in Speech and Audio Compression. Proc. of the IEEE, 1994, 82(6): 900～918

[3] J·D·马卡尔, A·H·格雷·乔. 语音信号线性预测. 娄乃英等译. 北京: 中国铁道出版社, 1987

[4] J. L. Flanagan, R. M. Golden. Phase Vocoder. Bell Syst. Tech., 1966, 45(5): 1493～1509

[5] A. V. Oppenheim. A Speech Analysis - Synthesis System Based on Homomorphic Filtering. J. Acoust. Soc. Am., 1969, 45(2): 458～465

[6] Griffin. D. W, Lim. J. S. Multi - Band Excitation Vocoder. IEEE Trans. ASSP, 1988, 36(8): 1223～1335

[7] R. Zelinski, P. Noll. Adaptive Transform Coding of Speech Signals. IEEE Trans. ASSP, 1977, 25(4): 299～309

[8] Jerry D. Gibson 等. 多媒体数字压缩原理与标准. 李煜辉等译. 北京: 电子工业出版社, 2000

[9] 3G TS 26.090～094

[10] 舒华芙等. IP电话技术及其应用. 北京: 人民邮电出版社, 1999

[11] Richard V. C., Candace A. K.. Speech and Language Processing for Next - Millennium Communication Services. Proc. of IEEE. 2000, 88(8): 1314～1337

[12] 邹绘华, 李双田. 基于频率斜率约束的变速率语音编码算法研究. 信号处理, 2001, 第17卷增刊: 193～198

第1章 语言、语音及语音信号

1.1 概 述

语言是从千百万人的言语中历史地概括总结出来的规律性的符号系统，是人们用以进行思维、交际的形式。语音是声音和意义的结合体，声音是语言的物质形式，语音是语言的物质外壳、信息的载体。但是声音和意义之间没有必然的联系，也就是声音表达什么意义不是天然生成的，而是社会约定俗成的结果。

语音是由一连串音所组成的，这些音在相互间的过渡就是代表信息的符号，这些音(符号)的排列由语音规则所控制。对这些规则及其在人类通信中含义的研究属于语言学的范畴，而对语音中音的分类的研究称为语音学。

研究表明，自从人类从能劳动起，就开始有了语言。通常认为，语言至少有几十万年，甚至一百万年的历史。文字的发明可能与农业产生有关，而农业的产生差不多是一万年前的事，也就是说，文字至多只有一万年的历史。如果说人类选择用语音形式的符号系统来表达语义内容——思想感情，那么，记录语音的文字就是第二次符号——符号的符号。这是由于人类发觉语言有其局限性——暂时性，这个局限性使语言在时间上和空间上都受到了限制。所以随着社会的进化，尤其是农业的产生，而出现了文字。

世界上有多少种语言呢？据东德《语言学及语言交际工具问题手册》中的介绍有 5651 种，但有些正在衰亡。据语言学家预言，21 世纪每年将有 12 种语言消失。这些面临死亡的语言不只是第三世界国家的少数民族语言，欧洲也有 150 种语言受到“生死存亡”的考验。

携带语言信息的语音声波就是语音信号。如果经过声电转换就形成语声的电信号，如果经过声光转换就形成语声的光信号。在现代技术条件下，主要是语声电信号。

为了学习语音数字处理，必须了解发音的生理结构与过程，以及汉语语音的特性。从语音产生的声学理论可以得到语音信号的无损声管模型，并在此基础上可以建立语音信号的数字模型。同时必须了解听觉的生理结构与过程，以及听觉的心理特征，从说和听两个方面研究关于口与耳的科学。

在这一章中，我们首先讨论发音的生理机构与过程以及汉语语音的基本特性，然后将语音信号看成是线性时变系统在随机噪声和准周期脉冲序列激励下的输出，应用这一方法就得到了语音信号的数字模型。最后讨论听觉生理器官与听觉心理，从而了解人类听觉的基本特性。

1.2　发音的生理机构与过程

人的发音生理机构如图 1.1 所示。发音时由肺部收缩送出一股直流空气，经气管流至喉头声门处(声门即声带开口处)。在发声之初，声门处的声带肌肉收缩，声带并拢(间隙小于 1 mm)，这股直流空气冲过很小的缝隙，使声带得到横向和纵向的速度，此时，声带向两边运动，缝隙增大(成年男性开到最大时，截面积约为 20 mm^2)，声门处压力下降，弹性恢复力将声带拉回平衡位置并继续趋向闭合，即声带产生振动，而且具有一定的振动周期。

一般把声门以上，经咽喉、口腔(舌、唇、腭、小舌)的这一管道称为主声道。成年男子的主声道长度约 17 cm。而经小舌和鼻腔的这一管道称为鼻道。此外，经肺、支气管和气管的管道称为次声门系统。由声带振动激发声道中空气发生振动，并从口和鼻两处向外辐射产生声音。声道的口、鼻两个管道中，从鼻咽部到鼻孔的分支称为鼻道分支，只有在发鼻音时才打开，从声门到唇是主声道，它被舌面隆起点隔开，近似可分为咽腔(后腔)、小管、口腔(前腔)等几部分。当发出一语音时，声道肌肉(包括舌面)运动到一个特定的部分，构成一定声道的位形，形成该语音的特定音色。

语音按其激励形式的不同大致可以分成三类。当气流通过声门时，如果声带的张力刚好使声带产生张弛振荡式振动，产生一股准周期脉冲气流，这一气流激励声道就产生浊音(Voiced Speech)或称有声语音。如果声带不振动，而在某处收缩，迫使气流以高速通过这一收缩部分而产生湍流就产生清音(Unvoiced Speech)或摩擦音，或称无声语音。如果声道在完全闭合的情况下突然释放就产生爆破音(Plosive Speech)。

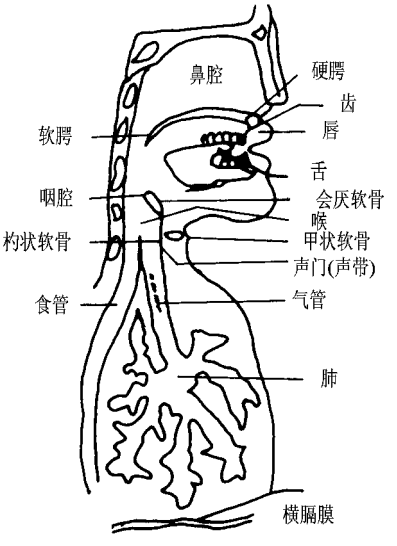

图 1.1　发音器官的生理解剖

图 1.2 给出了发 125 Hz 基频声带开启面积与时间的关系曲线，发音时声带(声门)逐渐开启，约占基音周期 50％的时间，达到开启面积最大。然后逐渐关闭，到完全闭合，约占基音周期 35％的时间。约 15％的时间内声门是完全闭合的，这样声门开启到下一次声门开启形成一个基音周期。基音周期的倒数是基频，它具有时变性和准周期性。基音周期与个人的声带长短、厚薄、韧性、劲度和发音习惯有关，在很大程度上反映个人的特征。

人的声道和鼻道都是非均匀的声道管，声道管的谐振频率称为共振峰频率，简称为共振峰。它与发音器官的确切位置有很大的关系，即共振峰和声道的形状与大小有关。表1.1 给出了汉语普通话 7 个韵母的共振峰频率。从表中可以看到，各韵母音色上的差异可用前三个共振峰(f_1、f_2、f_3)来表示。f_1 主要分布在 290～1000 Hz 范围内，f_2 分布在 500～2500 Hz 范围内，而 f_3 分布在 2.5～4 kHz 范围内。

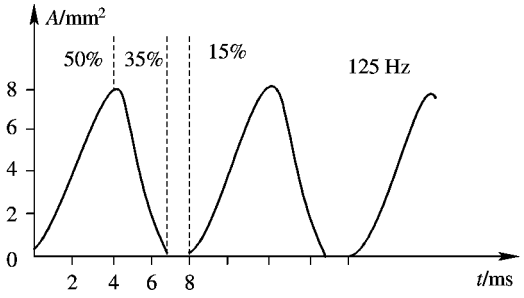

图 1.2 声带开启的面积与时间的关系曲线

表 1.1 汉语拼音七个韵母的共振峰频率(Hz)

共振峰 \ 韵母	性别	i 衣	u 乌	ü 迂	a 啊	o 喔	e 鹅	er 儿
f_1	男	290	380	290	1000	530	540	540
	女	320	420	320	1230	720	750	730
	童	390	560	400	1190	850	880	750
f_2	男	2360	440	2160	1160	670	1040	1600
	女	2800	650	2580	1350	930	1220	1730
	童	3240	810	2730	1290	1020	1040	1780
f_3	男	3570	3660	3460	3120	3310	3170	3270
	女	3780	3120	3700	2830	2930	3030	3400
	童	4260	4340	4250	3650	3580	4100	4030

语音信号随时间而变化的谱特性可以利用语图仪(Spectrograph)图形显示。此图有时也称为语谱图，是一种三维图形，纵轴对应于频率，横轴对应于时间，图像的黑白度正比于语音信号的能量。图 1.3 所示为普通话“北京”语音的语谱图。

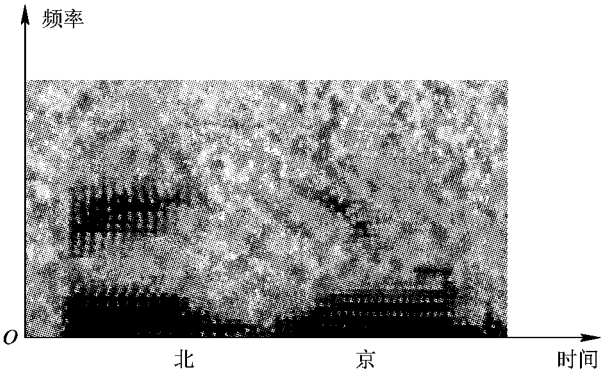

图 1.3 语谱图