普通高等教育“十一五”国家级规划教材
普通高等教育电子信息类规划教材

数字语音编码

赵晓群　编著

机械工业出版社
CHINA MACHINE PRESS

普通高等教育“十一五”国家级规划教材

数字语音编码

赵晓群　编著
朱　杰　万旺根　主审

机械工业出版社

本书作为普通高等教育“十一五”国家级规划教材，全面、系统地阐述了现代数字语音编码的原理、技术和应用。本书是在汲取了国内外相关教材、专著的优点，结合数字语音编码的基本理论与工程应用以及作者的教学经验的基础上编写的。全书内容深入浅出，既保持理论的完整性、系统性，又概念清楚、易读好懂，同时注重数字语音分析和编码的新发展。

本书主要介绍了数字语音处理技术的发展历史，语音的基本特征和语音信号模型，基本的语音分析方法，数字语音的波形编码、参数编码和混合编码方法，语音编码器的性能评价和语音增强的基本方法等内容。

全书共14章，分为：绪论、数字语音处理基础、语音信号的模型、语音信号的时域分析、语音信号的频域分析、语音信号的同态分析、语音信号的线性预测分析、语音信号的矢量量化、线性预测声码器、合成-分析线性预测声码器、多带激励声码器、语音波形编码、语音编码器的质量评价、语音增强。

本书适合作为高等院校电子信息类专业的高年级本科生和研究生教材，对于从事信息科学和技术领域工作和研究的人员也极具参考价值。

图书在版编目（CIP）数据

数字语音编码/赵晓群编著. —北京：机械工业出版社，2007.5
普通高等教育“十一五”国家级规划教材
ISBN 978-7-111-21458-8

Ⅰ.数… Ⅱ.赵… Ⅲ.语音数据处理-编码-高等学校-教材
Ⅳ.TN912.3

中国版本图书馆CIP数据核字（2007）第067166号

机械工业出版社（北京市百万庄大街22号 邮政编码100037）
责任编辑：闫晓宇 版式设计：张世琴 责任校对 刘志文
封面设计：张 静 责任印制：杨 曦
三河市宏达印刷有限公司印刷
2007年8月第1版第1次印刷
184mm×260mm · 21.5印张 · 534千字
标准书号：ISBN 978-7-111-21458-8
定价：35.00元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
销售服务热线电话：（010）68326294
购书热线电话：（010）88379639 88379641 88379643
编辑热线电话：（010）88379727
封面无防伪标均为盗版

前　言

语音是人类交流信息最直接、最方便、最有效的工具，是人类获取信息的重要来源和利用信息的重要手段，也是现代通信系统中最常见的数据内容。随着通信技术，特别是移动通信技术和互联网技术的快速发展，语音通信技术也在不断地更新、发展。

语音数字化通信是当前信息产业中发展最快、普及面最广的业务。在有线、无线通信以及数字存储中，语音信号的数字化及其压缩技术在军事、民用领域中都有其重要的作用和意义。

无线信道的频率资源是有限的，为了满足用户不断增长的需求量，通信的有效性要求显得非常突出。提高有效性主要通过改进信源（语音）编码技术来完成。尤其是第三代移动通信的发展，对语音压缩编码算法提出了更高的要求，不但要求编码码率较低以增加系统容量，而且要求合成音质较高以保证通话质量。

IP 电话采用语音分组技术、语音压缩编码和统计复用，可以提高语音通信的带宽利用率，使语音通信的成本大幅度降低，这也从一个侧面推动了现代语音编码技术的发展。

1982 年首枚数字信号处理器（DSP 芯片）问世以及随后的飞速发展，为语音编码器的实现和应用奠定了基础。语音编码技术的日趋成熟和新通信技术的不断发展，对语音编码不断提出新的要求，迫切需要语音编码标准化。8kbit/s 以上的语音编码技术已经标准化，并进入实用阶段。低码率的编码标准仍处于发展之中，并且研究方向逐步转向更低的码率。

近十几年来，经过科技工作者的不懈努力，在理论发展和市场需要的双重推动下，已经实现了多个标准化的语音编码技术，并在工程中获得了巨大成功。但是，在语音编码理论与技术的教学方面，可供选择的教科书不多。目前我国很多院校通信类专业开设的语音方面的课程多以“数字语音信号处理”为主，教学内容除讲述数字语音信号和编码的相关知识外，还要讲述语音合成、语音识别、说话人与语种识别、语音增强等内容，致使语音编码领域的大量成果（包括现代通信技术中广泛使用的先进的数字语音编码技术的理论与应用成果）未能进入教材，使学生不能掌握现代语音通信技术的基本知识和技能。这对我国通信信息产业的发展和通信工程学科的教育是不利的。因此，编写一部以语音通信，特别是移动语音通信为主，兼顾宽带语音编码（适于网络语音通信）的数字语音编码方面的教材，有着十分重要的意义。

本书从人的发音和听觉机理出发，分析语音的产生过程并建立语音模型，以此为基础，全面系统地介绍了现代语音信号分析、语音编码的基本理论和多种语音编码标准，以及语音编码器的质量评价等。

本书绪论后的主干内容共分为四部分。第一部分是数字语音处理的基础知识，由第 2、3 章组成，主要介绍了语音的发音和听觉机理、生理与心理，语音的声学特性以及语音产生的多种数学模型。第二部分是基本的语音分析方法，由第 3～8 章组成，主要介绍了语音信号的时域分析、频域分析、同态分析、线性预测分析和语音信号的矢量量化技术。这一部分的内容是数字语音编码的基础，也是语音合成、语音识别、说话人与语种识别、语音增强等

各种语音处理技术的基础。第三部分是数字语音编码，由第9～12章组成，主要介绍了线性预测声码器、合成-分析线性预测声码器、多带激励声码器、语音波形编码器的理论与方法，以及相应的多种编码标准。这一部分的编写特色是理论与工程相结合，在介绍基本编码理论的同时，结合已标准化的语音编码技术规范进行展开的。第四部分是语音编码器的质量评价和语音增强的基本方法简介，由第13、14章组成。

本书的内容较多，在教学过程中，可根据具体情况进行取舍。对于少学时教学，建议在以下各方面进行取舍：适当减少第3章中声管模型的理论推导内容，略去3.7节“语音信号的非线性模型”；适当减少第5章的教学学时，可略讲5.6节“语音的短时合成技术”、5.7节“基于FFT的短时Fourier分析”，略去5.8节“频域基音检测”、5.9节“语音信号的时频表示”；略去第6章“语音信号的同态分析”；适当减少第7章的教学学时，可略去7.5节“LPC的频域特性”、7.7节“LPC的几种推演参数”；由于第9～12章涉及较多的语音编码方法与标准，可根据需要在每种编码方法中选取有代表性的编码方法与标准进行教学；第13、14章作为本课程的补充知识，建议学生自学。

本书在编写过程中，得到了同济大学的领导和老师们的大力支持。宫云梅博士认真阅读了全书，提出了一些宝贵的修改建议，孙凤宇硕士为本书计算了大量的实验数据。在此，对为本书的写作提供帮助的同事和同学们表示深切谢意。

最后要感谢我的妻子段晓英女士和我的家人们，他们的关心和支持是我写作的最好动力和最佳保障！

本书得到了同济大学教材出版基金的资助。

鉴于作者水平有限，书中难免有错误和不妥之处，敬请专家和读者不吝赐教。

作　者

2007年5月

目　录

第1章 绪 论

1.1 概述

语言是人类进行交流的重要手段，通信系统中最常见的数据形式就是语音数据。语音通信是人类通信最基本、最重要的方式之一。随着移动通信与互联网的飞速发展，语音通信技术也在不断地进行更新并与之相融合。数字化的语音信号进行传输和存储时，在可靠性、抗干扰能力、快速交换、安全性等方面远胜于模拟化的语音信号，且灵活方便、价格低廉。所以，从20世纪50年代以来数字化语音在通信系统中所占的比重越来越大。语音编码是数字语音通信中的一项关键技术。为了压缩数字语音信号的传输码率，以使同样的信道带宽能传输更多路的语音，节省存储空间，语音压缩编码理论与技术得到了极大的发展，并在有线、无线电话的话带语音信号、会议电视的宽带语音信号、数字高清电视和高保真音乐等的音频信号等领域有广泛的应用。

通信系统是围绕着通信传输数据的数量和质量这两个类型的三种指标（有效性、可靠性和安全性）进行不断优化的。有效性是指占用尽可能少的信道资源（如频段、时隙和功率）传送尽可能多的信息，它是通信的数量指标；可靠性主要是指在传输中抵抗各类客观自然干扰的能力，但是在军事通信中它也包含电子对抗；安全性则是指在传输中的安全保密性能，即收端防窃听、发端防伪造和篡改的能力等。移动通信中的各类新技术，都是以解决移动通信中的有效性、可靠性和安全性为目标而设计的。

移动通信属于无线通信。在无线通信中有效性的要求显得非常突出，这是由于无线信道的频率资源有限。提高移动通信的有效性主要通过改进信源（语音）编码技术来完成。近年来，通信系统发展迅速，随着移动通信的发展，尤其是第三代移动通信的发展，对语音压缩编码算法提出了更高的要求，不但要求编码码率较低以增加系统容量，而且要求合成音质较高以保证通话质量。用传统的编码方式很难同时满足这两个要求。为此，提出了变速率语音压缩编码的方法。在移动通信系统中采用变速率语音压缩编码，可以根据需要动态调整编码速率[⊖]，在合成语音质量和系统容量中取得灵活的折衷，最大限度地发挥系统的效能。

在当前应用广泛、前景广阔的码分多址（CDMA）移动通信系统中，采用的变速率语音编码算法对于系统的容量和通话质量有非常重要的影响。由于移动通信市场竞争异常激烈，因此对变速率语音编码的研究成为一个热点。近几年来，变速率语音编码技术发展得很迅速，并不断有新的标准公布。随着技术的成熟，它的应用领域也越来越广阔，不仅限于移动通信系统，在IP电话、互联网等方面也有很好的应用前景。

⊖ “编码速率”后文中简称为“码率”。

1.2 语音信号处理的发展及应用

语音信号处理作为一个重要的研究领域，已有很长的研究历史。它的核心内容是认识和描述人类语音和语言的基本特征，即语音分析，并应用于语音编码、语音合成、语音识别、说话人识别、语种识别、语音增强和语音理解等众多分支领域。

语音信号处理是研究用数字技术处理语音信号的一门学科。语音信号处理的理论和研究包括紧密结合的两个方面：一是从语音的产生和感知来研究，这一研究与语音和语言学、认知科学、心理和生理学等学科密不可分；二是将语音视为一种信号来进行处理。

语音信号处理的研究工作最早可以追溯到1876年Bell发明电话时，该装置首次用声电、电声转换技术实现了远距离的语音传输。但是它的快速发展是从1940年前后开始的。1939年Dudley研制成功第一个声码器，从此奠定了语音产生模型的基础，这一成就在语音信号处理领域具有划时代的伟大意义。1947年Bell实验室Potter等人发明了语谱图仪，将语音信号的时变频谱用图形表述，为语音分析提供了一个有力的工具。

1952年Bell实验室的Davis等人首次研制成功能识别10个英语数字的实验装置。1956年Olson等人采用8个带通滤波器组提取频谱参数作为语音的特征，研制成功一台简单的语音打字机。20世纪60年代初由于Fant和Stevens的努力，奠定了语音生成理论的基础，在此基础上语音合成的研究得到了扎实的进展。60年代中期形成的一系列数字信号处理方法和技术，如数字滤波器、快速傅里叶（Fourier）变换等，成为语音信号数字处理的理论和技术基础。随着计算机和微处理器技术的发展，以硬件为中心的研究也逐渐转化为以软件为主的处理研究。

1971年，以美国ARPA（American Research Projects Agency）为主导的“语音理解系统”的研究计划开始起步。这个研究计划不仅在美国国内，而且在世界各国都产生了很大的影响，它促进了连续语音识别研究的兴起。历时5年的庞大的ARPA研究计划，虽然在语音理解、语言统计模型等方面的研究积累了一些经验，取得了许多重要进展，但没能达到巨大投资应得的成果，在1976年停了下来，进入了深刻的反省阶段。但是，在整个20世纪70年代还是有几项研究成果对语音信号处理技术的进步和发展产生了重大的影响。70年代初日本学者板仓（Itakura）等提出的动态时间规整技术，使语音识别研究在匹配算法方面开辟了新思路。70年代中期线性预测技术用于了语音信号处理，此后隐马尔可夫模型法也获得初步成功，该技术后来在语音信号处理的多个方面获得巨大成功。70年代末，Linda、Buzo、Gray和Markel等人首次解决了矢量量化码书生成的方法，并首先将矢量量化技术用于语音编码获得成功。从此矢量量化技术不仅在语音识别、语音编码和说话人识别等方面发挥了重要作用，而且很快推广到其他许多领域。因此，80年代开始出现的语音信号处理技术产品化的热潮，与上述语音信号处理新技术的推动作用是分不开的。

20世纪80年代，由于矢量量化、隐马尔可夫模型和人工神经网络等相继应用并得到不断改进与完善，语音信号处理技术有了突破性的进展。其中，隐马尔可夫模型作为一种统计模型，在语音信号处理的各个领域中得到广泛应用。其理论基础是1970年前后由Baum等人建立起来的，随后由美国卡内基梅隆大学的Baker和美国IBM公司的Jelinek等人将其应用到语音识别中。隐马尔可夫模型已经成为目前语音识别的主流研究途径。

20世纪90年代以来，语音信号处理在实用化方面取得了许多实质性的研究进展，语音识别已走向实用化。随着声学语音学统计模型研究的深入，鲁棒的语音识别、基于语音段的建模方法及隐马尔可夫模型与人工神经网络的结合成为研究的热点。在语音合成方面，有限词汇的语音合成已在自动报时、报警、报站、电话查询服务、发音玩具等方面得到广泛应用。文-语转换系统的研究已在90年代初达到了商品化程度，其语音质量为大众所接受。

语音中的情感信息也是一种很重要的信息资源，它是人们感知事物的必不可少的一部分信息。而传统的语音信号处理技术把这部分信息作为模式的变动和差异，视为一种噪声给处理掉了。实际上，人们是同时接受各种形式的信息的，如何有效利用各种信息以达到最佳的信息融合、传递和交流效果，是今后信息处理研究的发展方向。所以分析包含在语音信号中的情感信息特征，并用于判断和模拟说话人的喜怒哀乐等是一个意义重大的研究课题，也是20世纪90年代以来兴起的一个新的语音信号处理研究领域。

抗噪声技术的研究以及实际环境下的语音信号处理系统的开发，在国内外作为语音信号处理的非常重要的研究课题，已经作了大量的研究工作，取得了丰富的研究成果。目前国内外的研究成果大体分为三类解决方法：一类是采用语音增强等算法；第二类方法是寻找稳健的语音特征；第三类方法是基于模型参数适应化的噪声补偿算法。然而，解决噪声问题的根本方法是实现噪声和语音的自动分离，尽管人们很早就有这种愿望，但由于技术的难度，这方面的研究进展很小。近年来，随着声场景分析技术和盲分离技术的深入发展，利用在这些领域的研究成果进行语音和噪声分离的研究也取得一些重要进展。

说话人识别和语种辨识是语音识别的两种特殊形式。它们和语音识别一样，都是通过提取语音信号的特征和建立相应的模型进行分类判断的。说话人识别力求找出包含在语音信号中的说话人的个性因素，强调不同人之间的特征差异；而语种辨别则要从一个语音片段中判别它是哪一个语种，所以就要尽可能找出不同语种的差别特征。

1.3 语音编码算法综述

语音信号经A/D转换后直接编码，将产生大量的数据，这不利于传输或存储。因此为提高效率，必须对语音信号进行压缩处理。各种编码技术就是为减小传输码率或存储量，以提高传输或存储的效率。传输码率是指每传输一秒钟语音信号所需要的比特数，也称为数码率。采用编码技术后，同样的信道容量就能传输更多路的信号，如果是用于存储则只需要较小容量的存储器，因而这类编码又称为压缩编码。实际上，压缩编码需要在语音的可懂度和音质、降低传输码率、降低编码过程的计算代价3方面进行折衷。近20年来固定电话和移动通信的高速发展，信道使用效率成为一项关键因素，这促使了语音编码技术的快速发展。即使在今天，光纤的使用使得有线通信的带宽变得更廉价，但是在有线通信以及移动通信、卫星通信和掌上电脑的语音传送应用中，语音编码依旧扮演着十分重要的角色。

信息从“这里”传输到“那里”，即通信；或者信息从“现在”传输到“将来”，即存储，这两个物理过程均可用图1-1的数字传输系统模型来概括。图中的信源编码和信源解码即为本书的研究内容，统称为信源编码（点画线以左部分）；信道编码和信道解码统称为信道编码。信源编码和信道编码都是信息科学的重要分支。其中，信源编码主要解决有效性问题。通过对信源的压缩、扰乱、加密等一系列处理，力求用最小的传输码率传递最大的信息

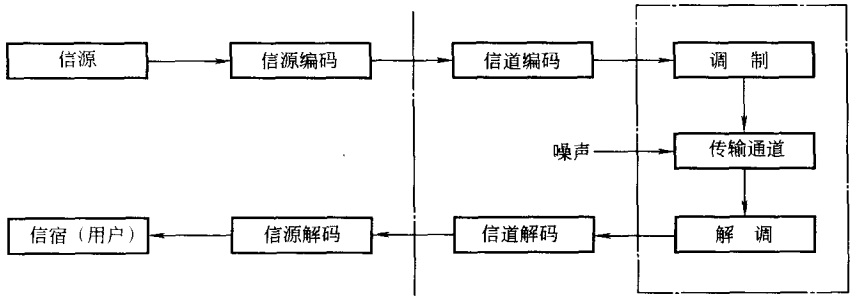

图 1-1 数字传输系统模型

量，使信号更适宜传输和存储。信道编码主要解决可靠性问题，即尽量使处理过的信号在传输的过程中不出错或者少出错，并能自动检错和尽量纠错。信源编码的主要目的是解决数据压缩的问题，以减少容纳给定信息集合或数据样本集合的信号空间。在今天，“数据压缩”与“信源编码”已是两个具有相同含义的术语了。语音编码属于信源编码的范畴。

语音编码通常分为波形编码、参数编码和混合编码三类。波形编码与参数编码的主要区别在于重建的语音时域信号是否在波形上尽量与原始信号一致。波形编码力图使重建后的语音时域的波形与原语音信号波形保持一致，它具有适应能力强、话音质量好等优点，但码率较高。自适应差分脉冲编码调制是这类编码的主要代表。

参数编码通常称为“声码器技术”。它根据对声音形成机理的分析，在使重建的语音信号具有足够的可懂性的原则上，通过建立语音信号的产生模型，提取代表语音信号特征的参数来编码，并不要求在波形上与原始信号匹配。频域上这一模型对应于具有一定零极点分布的数字滤波器，编码器只需要发送滤波器参数和相关的语音特征值。由于语音在短时间内可认为其模型特征近似不变（短时平稳性），所以模型特征参数更新的频度较低，这就有效地降低了码率。参数编码的优点是码率低，甚至可以工作在 2.4kbit/s[⊖]以下。其主要问题是合成语音质量差，特别是自然度较低；另外对环境噪声较敏感，需要较安静的环境才能给出较高的可懂度。共振峰声码器和线性预测声码器都是典型的参数声码器。

自 20 世纪 30 年代末提出脉冲编码调制原理以及声码器的概念后，语音编码曾一直沿着这两个方向发展。表 1-1 为两类编码方法的特点比较。在此基础上发展起来的混合编码是上述两类编码方法的有机结合，由于突破了波形编码和参数编码的界限，性能有很大的提高，

表 1-1 波形编码与参数编码的比较

编码方案 / 比较项目	波形编码	参数编码
编码信息	波形	模型参数
码率/kbit/s	9.6～64	2.4～9.6
语音质量评价方法	信噪比	谱失真和主观听音
缺点	随着量化粗糙语音质量下降	合成语音质量较低，处理复杂度高

⊖ 传输速率和数字系统内部的其他数字速率（例如编码速率）的标准单位是 bit/s，在其他技术文献中也有写做非标准形式 bps 的，请读者注意它们的含意是相同的。

故得到了更广泛的应用。混合编码基于语音产生模型，并采用合成-分析技术，因此它同时利用语音模型和时间波形信息，增强了重建语音的自然度，明显地提高了合成语音质量。其代价是码率相应上升，一般在 16 ~ 2.4kbit/s 之间。多脉冲激励线性预测编码、规则脉冲激励线性预测编码、码激励线性预测编码等都属于混合编码。

语音编码的研究起于 70 年前，主要是适应窄带电话线语音信号传输系统的发展需要。早期的声码器基于对语音信号基音周期和频谱的分析，通过周期脉冲或随机噪声激励 10 个带通滤波器（表示声道模型）合成语音信号。主要有通道声码器、共振峰声码器和模式匹配声码器。20 世纪 50 年代后期语音编码研究着重于线性语音源-系统的生成模型。这种模型包括一个线性慢时变系统（声道模型）和周期脉冲激励序列（浊音信号）以及随机激励（清音信号）。源-系统是一个自回归时序模型，声道是全极点滤波器，参数通过线性预测分析得到。除了线性预测模型之外，同态分析也可分离出卷积的信号。同态语音分析最大的优点是可以从倒谱中得到基音信息。70 年代前后，由于 VLSI（超大规模集成电路）技术的出现和数字信号处理理论的发展，为语音编码提供了新的解决方案。语音的分析-合成采用了短时 Fourier 变换、变换编码和子带编码。并且基于线性预测的语音编码技术得到进一步的发展。目前在通信应用中提出了鲁棒性、低码率和高质量的语音编码要求。这些新的编码技术包括余弦分析合成技术、多带激励声码器、多脉冲和矢量激励以及矢量量化。对线性预测参数而言，矢量量化是一种非常有效的编码方式。

自从 1937 年 A. H. Reeves 提出脉冲编码调制开创了语音数字化通信的历程以来，在 70 多年的时间里，语音编码已取得了迅速的发展。特别是从 1980 年至今，语音编码领域已经取得了很多重要的进展。这些进展的取得主要有下列原因：

1）对语音产生机理和语音信号结构的更深入理解。

2）对人的听觉系统的深入理解，利用人耳的掩蔽效应提出了易于实现的听觉加权滤波器方案。

3）提出了更好的量化技术，特别是合成 - 分析的技术，使得重构语音的质量有显著的提高。

4）数字信号处理芯片的广泛使用，为语音编码器的商品化打下了良好基础。

这些发展趋势似乎还在继续，至少目前的情况是这样。但是已经看到，ITU-T 的语音编码专家组研究焦点有移动的倾向。大约 1992 年以前，在语音编码上的主要进展大都是基于线性预测，在合成-分析法的基础上获得的。很多年以来，对于 4.8 ~ 16kbit/s 之间的码率，这种方法几乎占了统治地位。但是，现在有情况表明，如果码率降低到 4.8kbit/s 以下，基于线性预测的合成-分析方法，超过其他方法的优点逐渐减少。所以，这种方法未能进入 2.4kbit/s 语音编码器的范围。在保证语音质量的前提下进一步降低码率，仍然是语音编码研究的主要焦点。将会有更多的参数编码器进入应用领域，例如多带激励编码器、正弦变换编码器、波形内插编码器等。

近年来，在语音编码领域中出现了许多新的应用。例如蜂窝电话和应答机，它们并不需要固定的码率。因此，工作的重心已转向可变速率编码，随着这些编码器应用的增加，可变速率的语音编码研究也会明显加强。

语音编码所需要的最低信息速率是一个很复杂的问题。它受多种因素的限制，例如语音信号所包含的信息内容，而信息内容又依赖于测量的方法。但是，作为一个底限，临界信息

速率应该是人理解信号所需要的速率。这是一个还需要继续深入研究的问题，因为有关语音信号的某些信息，人能够感觉到有变化，而编码器却找不到对应的特征参量。反之，有时语音的波形和特征参量变化很大，而人同样可以理解。例如，一个发音人，他将一段文章读两次，产生了两段非常不同的波形，但是，这些差别并不影响收听者的理解。因此，要说语音编码器具有多少码率才是最终的结果，目前还是很困难的。要达到语音表示的底限速率，必须对人脑感知信号的过程有更深入的研究，这恐怕也是长期和艰巨的工作。

1.4 语音编码标准的发展

众所周知，任何工业产品的标准化都是非常重要的。尤其对于公用的通信工具而言，更迫切地需要标准化。制造商、服务商和用户都认识到一个公用的标准对于任何一方都是有利的。对于一种新的应用，不同的用户可能会建议很多种标准。但是，现行的标准应该建立在一个专利产品的基础上，当然，产品之间应该相互兼容。如果对于一种新的应用，市场还没有一个专利产品是明显的优胜者，则应该把意见提交到标准化的组织者，然后作一真正的技术讨论，以便形成一个新的标准。

标准化组织的一些实体负责制定新的标准。国际电信联盟（ITU）是联合国经济、科学、文化组织的一部分，由他们负责制定全球通信标准。最初，ITU 是由国际电话电报咨询委员会（CCITT）和国际无线电咨询委员会（CCIR）组成。CCITT 负责建立电信标准，包括语音编码标准，而 CCIR 负责建立无线电标准。1993 年 ITU 重新组织，CCITT 变成了 ITU 电信标准部门（ITU-T）的一部分。在 ITU-T 中，15 研究组（SG15）负责制定语音编码标准，12 研究组（SG12）负责测试这些标准在网上的性能，在试验语音编码器时，是和 SG15 一起工作的。在 ITU-T 中的另外一些研究组，还有 ITU 的无线电标准化部门（ITU-R），对于一些满足他们应用的语音编码标准，也可以提出要求。

数字蜂窝电话标准是由一些地方标准化组织建立的。他们负责规定全部蜂窝系统。其中，语音编码器是最有活力，但也是很小的一部分。在欧洲，欧洲电信标准研究所（ETSI）负责制定数字蜂窝标准。在北美，这项工作是由电信工业组（TIA）负责执行。在日本，由无线系统开发和研究中心（RCR）组织这些标准化的工作。总之，不同的地区和国家都有相应的组织，从事有关标准化技术方面的工作。

另外一些组织，也建立了一些专门的应用标准。国际海事卫星组织（INMARSAT）管理地球上的同步通信卫星，已经制定了一系列的卫星电话应用标准。此外，各国政府能够制定自己国家的标准，例如，美国政府有保密电话语音编码标准，北大西洋公约组织也有自己的保密电话标准。

因为语音编码器是一种标准，这就意味着对于任何一个使用者，都是开放的，这一点非常重要。同时由于开发语音编码器，并制定相应的语音编码标准，都是由个体者的努力而完成的，应该对他们的工作给予补偿，收取专利税就是一种有效的方法。产品供应商愿意出售已经标准化的产品，但是，供应商得到的许可证必须包括他们所使用的标准，因此事先应该和具有语音编码标准专利权的组织协商有关知识产权问题。

第2章　数字语音处理基础

数字语音处理是研究用数字信号处理技术对数字语音进行处理的一门学科。其目的一是通过处理得到反映语音信号重要特征的语音参数，以便高效地传输或储存语音信息；二是通过某种处理运算以达到某种用途的要求，例如人工合成出语音、增强语音中某些成分、辨识出讲话者、识别出讲话的内容等。因此，在研究数字语音处理技术及其各种应用之前，有必要了解一些有关语音信号重要特性的知识，在此基础上才可以建立既实用又便于分析的语音信号产生模型和语音信号感知模型，它们是贯穿整个数字语音处理的基础。

2.1　发音的生理器官与过程

人类的语音是由人体发音器官在大脑控制下的生理运动产生的。人体的发音器官由肺和气管、喉（包括声带）、声道（咽腔、鼻腔和口腔）3部分组成。肺和气管是整个语音系统的能源提供者，喉是主要的发音生成机构，声道则对生成的声音进行调制。

肺是胸腔内的一团有弹性的海绵状组织，它可以存储空气。肺的主要生理功能是进行血液和空气之间的气体交换，也就是将空气中的氧气吸收到血液中，将血液中的二氧化碳排出到空气中，这就是人体的呼吸功能。肺的另一个重要功能是将压缩空气供给发音器官。人是在正常呼吸的情况下说话的，在不说话时人的呼吸通常是规则的、平稳的、节律性的；在说话时为保持语言的连续性，就不得不有短暂的停顿，特点是吸气短、呼气长，且呼吸受到句子结构的控制，并没有一个固定的规则。空气由肺部排入喉部，经过声带进入声道，最后由嘴（或鼻，或嘴和鼻）辐射出声波，形成了语音。因此，肺提供了语音产生的能量。气管连接肺和喉，是肺与声道联系的通道。

图2-1是喉构造图。喉是由软骨和肌肉组成的复杂系统，其中包含重要的发音器官——声带。声带是一个阀门，又是一个振动部件。声带紧绷在喉头的前后壁上，有折叠，声带的长度约10～14mm。声带的前端由甲状软骨支撑，后端由杓状软骨支撑。杓状软骨与环状软骨的上部相连。这些软骨由附在环状软骨上的一组肌肉控制，可以移动声带的末端使之开启或闭合。当声带末端分离开启时，就是正常呼吸状态。两片声带之间的空间叫做声门。当声带末端闭合时，肺部被密封成一个密闭的小室。声带有生物学和声学双重功能。它的生物学功能是封闭气管以保护肺道（例如，在吞咽食物时防止食物进入肺道），或在胸腔和腹腔建立一定的气压（例如，

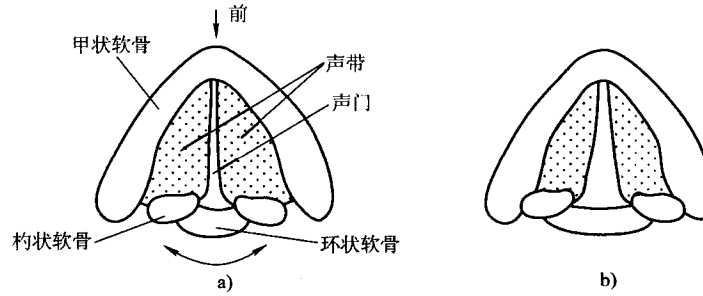

图2-1　喉的构造

a）发音状态　b）呼吸状态

帮助排泄和分娩）。声带的声学功能是为产生语音提供主要的激励源。

人在说话时两片声带在杓状软骨的控制下相互靠近但不完全封闭，使声门变成一条窄缝。当肺部的气流通过这个窄缝隙时，缝隙间的压力减小，两片声带完全合拢使气流不能通过。气流阻断后，压力恢复正常，声带间的空隙再次形成，气流再次通过。这一过程周而复始地进行，形成进入声道的一串周期脉冲气流（声门脉冲串）。声带每开启/闭合一次的时间（即声带的振动周期）就是语音的基音周期，其倒数称为基音频率。图 2-2 是发 125Hz 基音频率的语音时，声带开启的面积与时间的曲线。可见，声带开启后大约 4ms，声带的面积约达到最高峰 8mm^2 左右；随后大约 3ms 闭合；然后受气管的气流冲击 1ms 又重新开放。基音频率与声带的大小、厚薄、松紧程度以及声门上下之间的气压差等有关，通常约为 50～450Hz。基音频率范围随发音人的性别、年龄而不同，老年男性偏低，小孩和青年女性偏高。基音频率决定了声音频率的高低，频率高则音调高，频率低则音调低。成年男性的基音频率一般为 50～250Hz，女性的基音频率一般为 200～450Hz。

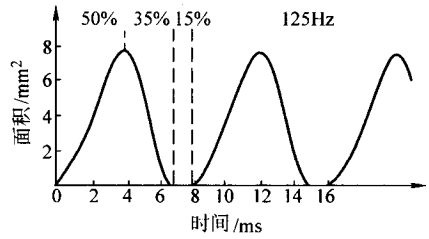

图 2-2 声带开启的面积与时间的关系

从声门至口唇的所有发音器官称为声道，其纵剖面图如图 2-3 所示。声道包括咽腔、口腔和鼻腔 3 个空气腔体组成，它是一根从声门延伸至口唇的非均匀截面的声管，其外形变化是时间的函数。口腔包括上下唇、上下齿、上下齿龈、上下腭、舌和小舌等部分。上腭分为硬腭和软腭两部分。舌分为舌尖、舌面和舌根 3 部分。鼻腔在口腔上面，靠软腭和小舌将其与口腔隔开。当小舌下垂时，鼻腔与口腔便耦合起来；当小舌上抬时，口腔与鼻腔是不相通的。发音时，口腔和鼻腔都起共鸣作用。口腔中各器官能够协同动作，使空气流通过时形成各种不同情况的阻碍并产生振颤，从而发出不同的声音。一般成年男子的声道长度约 17cm 左右，最大截面积可达 20cm^2 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，咽腔的形状是变化的，如图 2-4 所示。咽腔与口腔一起使声道的形状变化增多，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 10cm 长。发鼻化语音时软腭下垂。口腔是声道最重要的部分，其大小和形状可以通过调整舌、唇、齿和腭来改变。在调整发音时，舌是最活跃的部分。

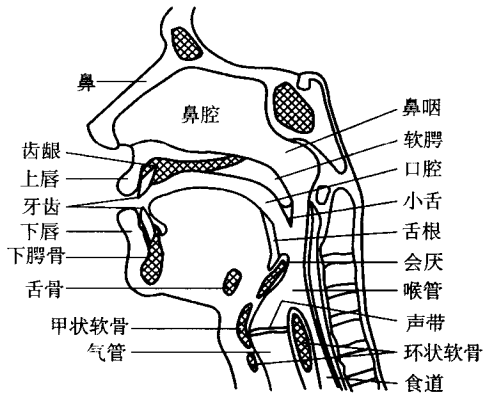

图 2-3 声道纵剖面图

上述声音产生机制的原理如图 2-5 所示。

在发音过程中，肺部与相连的肌肉相当于声道系统的激励源。当声带处于收紧状态时，流经的气流使声带振动，这时产生的声音称为浊音；而不伴有声带振动的音称为清音。当声带处于放松状态时，有两种方式能发出声音。一种方式是通过舌，在声道的某一部分形成狭窄部位，也称为收紧点，当气流经过这个收紧点时会产生湍流，形成噪声型的声音；对应的收紧点的位置不同及声道形状的不同，形成不同的摩擦音。另一种方式是声带处于松懈状态，