21世纪高等院校信息与通信工程规划教材
21st Century University Planned Textbooks of Information and Communication Engineering

语音信号处理实用教程

吴进 编著

Speech Signal Processing Practical Course

人民邮电出版社
POSTS & TELECOM PRESS

高校系列

21世纪高等院校信息与通信工程规划教材

21st Century University Planned Textbooks of Information and Communication Engineering

语音信号处理实用教程

吴进 编著

Speech Signal Processing Practical Course

人民邮电出版社

北京

高校系列

图书在版编目（CIP）数据

语音信号处理实用教程 / 吴进编著. -- 北京 : 人民邮电出版社, 2015.2
21世纪高等院校信息与通信工程规划教材
ISBN 978-7-115-38087-6

Ⅰ. ①语… Ⅱ. ①吴… Ⅲ. ①语声信号处理－高等学校－教材 Ⅳ. ①TN912.3

中国版本图书馆CIP数据核字(2015)第014733号

内 容 提 要

本书系统地介绍了语音信号处理的相关知识，系统地论述了语音信号处理的基础、概念、原理、方法与应用，以及该学科领域取得的一些新成果、新进展及新技术。全书分 3 篇共 15 章，其中第 1 篇语音信号处理基础篇，包括第 1 章绪论，第 2 章语音信号处理的基础知识；第 2 篇语音信号分析篇，包括第 3 章至第 9 章，介绍语音信号的各种分析方法和技术，包括传统方法，如时域、频域处理等，还包括新方法和新技术，如同态处理、线性预测分析、矢量量化、隐马尔可夫模型技术等；第 3 篇语音信号处理技术与应用篇，包括第 10 章至第 15 章，分别介绍语音编码、语音合成、语音识别、说话人识别、语音增强及语音处理的实时实现。

本书在编写上既重视基础知识，又跟踪前沿技术；既具有学术深度，又具有教材的系统性和可读性。全书层次分明，条理清晰，结构严谨，并注意各部分内容的有机结合；既有较强的理论系统性，又体现一定应用的观点。

本书可作为高等院校电子信息工程、通信工程、模式识别与人工智能等专业的高年级本科生、硕士研究生教材，也可供该领域的科研及工程技术人员参考。

◆ 编　　著　吴　进
责任编辑　张孟玮
执行编辑　李　召
责任印制　沈　蓉　彭志环

◆ 人民邮电出版社出版发行　　北京市丰台区成寿寺路 11 号
邮编　100164　　电子邮件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
三河市潮河印业有限公司印刷

◆ 开本：787×1092　1/16
印张：21.75　　2015 年 2 月第 1 版
字数：530 千字　　2015 年 2 月河北第 1 次印刷

定价：56.00 元

读者服务热线：(010)81055256　印装质量热线：(010)81055316
反盗版热线：(010)81055315

目　录

第 3 篇 语音信号处理应用篇

第1篇

语音信号处理基础篇

第 1 章 绪论

1.1 语音信号处理概述

通过语言相互传递信息是人类最重要的基本功能之一。语言是人类特有的功能，是人类最重要的交流工具，它自然方便、准确高效。虽然人们可以通过多种手段获得外界信息，但是最重要、最精细的信息源只有语言、图像和文字三种。与用于声音传递信息相比，显然用视觉和文字相互传递信息，其效果要差得多。这是因为语音中除包含实际发音内容的语言信息外，还包括发音者是谁及喜怒哀乐等各种信息。所以，语言是人类最重要、最有效、最常用和最方便的信息交换形式。另一方面，语言和语音与人的智力活动密切相关，与文化和社会的进步紧密相连，它具有最大的信息容量和最高的智能水平。

语音信号处理是研究用数字信号处理技术对语音信号进行处理的一门学科，处理的目的是用于得到某些参数以便高效传输或存储；或者是用于某种应用，如人工合成出语音、辨识出讲话者、识别出讲话内容、进行语音增强等。

语音信号处理是一门新兴的学科，同时又是综合性的学科，是语音学与数字信号处理技术相结合的交叉学科。虽然从事这一领域研究的人员主要来自信息处理及计算机等学科，但是它与语音学、语言学、声学、认知科学、生理学、心理学、模式识别和人工智能等许多学科也有非常密切的联系。语音信号处理技术的发展依赖于这些学科的发展，而语音信号处理技术的进步也会促进这些学科的进步。

语音信号处理是许多信息领域应用的核心技术之一，是目前发展最为迅速的信息科学研究领域中的一个。语音信号处理是目前极为活跃和热门的研究领域，其研究涉及一系列前沿科研课题，且处于迅速发展之中；其研究成果具有重要的学术及应用价值。随着电子计算机和人工智能机器的广泛应用，人们发现，人和机器之间最好的通信方式是语言通信。而语音是语言的声学表现形式，要使机器能够听懂人的语言并能使用人类的语言进行表达，需要做很多工作，这就是研究了几十年的语音识别和语音合成技术。而随着移动通信的迅猛发展，人们可以随时随地通过电话进行交流，其中语音压缩编码技术发挥着重要的作用。上述这些应用领域构成了语音信号处理的主要研究内容。

20 世纪 60 年代中期形成的一系列数字信号处理方法和算法，如数字滤波器、快速傅里叶变换（FFT）等是语音信号处理的理论和技术基础。进入 20 世纪 70 年代之后，语音技术

取得了许多实质性的进展：用于语音信号的信息压缩和特征提取的线性预测技术（LPC）已成为语音信号处理最强有力的工具，广泛应用于语音信号分析、合成及各个应用领域；出现了用于输入语音与参考样本之间时间匹配的动态规划方法。20 世纪 80 年代初，一种新的基于聚类分析的高效数据压缩技术——矢量量化（VQ）应用于语音信号处理中；而隐式马尔可夫模型（HMM）的研究取得了迅速发展，语音信号处理的各项课题是促使其发展的重要动力之一；同时，它的许多成果也体现在有关语音的各项应用之中，尤其语音识别是神经网络的一个重要应用领域。

从技术角度讲，语音信号处理是信息高速公路、多媒体技术、办公自动化、现代通信及智能系统等新兴领域应用的核心技术之一。在高速发达的信息社会，用数字化的方法进行语音的传递、存储、识别、合成、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。同时，由于语言是人类相互间进行沟通的最自然和最方便的形式，所以它是一种理想的人机通信方式，因而可为计算机、自动化系统等建立良好的人机交互环境，进一步推动计算机和其他智能机器的应用，提高社会信息化和自动化的程度。

语音处理技术的应用极其广泛，包括工业、军事、交通、医学、民用等各个领域。目前，语音处理技术正处于蓬勃发展时期，已有大量产品投放市场，并且不断有新产品被开发研制，具有极其广阔的市场需求和应用前景。

目前对语音信号均采用数字处理的方式，这是因为数字处理与模拟处理相比，具有许多优点，具体表现为：

（1）数字技术能完成许多很复杂的信号处理工作；

（2）语音可以看成是音素的组合，具有离散的性质，特别适合于数字处理；

（3）数字系统具有高可靠性、价廉、紧凑、快速等特点，很容易完成实时处理任务；

（4）数字语音适于在强干扰信道中传输，易于和数据一起在通信网中传输，也易于进行加密传输。

1.2 语音信号处理的发展

语音信号处理的研究工作最早可以追溯到 1876 年贝尔发明的电话，它首次完成了用声电—电声转换来实现远距离传输语音的技术。电话的理论基础是尽可能不失真地传送语音波形，这种“波形不变”原则几乎统治了电话通信一个世纪之久。

1939 年达德利（Dudley）研制成功了第一个声码器，打破了语音信号的内部结构，使之解体，提取其参数加以传输，在接收端重新合成语音。根据载波电话原理，将声带中产生的音源类比载波信号，口腔运动看成是对载波的调制，将 3000Hz 带宽的语音信号压缩到 300Hz 以内，打破了垄断一个多世纪的“波形不变”原则，导致语音参数模型的出现。声码器技术奠定了语音产生模型的基础，在语音信号处理领域具有划时代的意义。

1947 年贝尔实验室发明了语谱图仪，将语音信号的时变频谱用图形表示出来，为语音信号的分析提供了一个有力的工具。语谱图仪的研制成功对声学语音学的发展曾经起过很大的推动作用。1948 年美国 Haskins 实验室研制成功了“语图回放机”，它把手工绘制在薄膜片上的语谱图自动转换为语音，可以进行语音合成。共振峰合成方法就是源于这一思想。1952 年，贝尔实验室的戴维斯（Davis）等人首次研制成功了特定说话人孤立数字识别系统，该系统利

用每个数字元音部分的频谱特征进行识别。1956 年达德利（Dudley）等人又将语音分割成元音、辅音等，改进了这一装置。同年，RCA 实验室的奥尔森（Olson）等人也独立地研制出 10 个单音节词的识别系统，该系统采用从带通滤波器组活动的频谱参量作为语音的特征。

进入 20 世纪 60 年代，语音信号处理的研究工作取得了新的进展，其主要标志是 1960 年瑞典科学家范特（Fant）开创性论文“语音产生的声学理论”的发表，它为建立语音信号数字模型奠定了基础。20 世纪 60 年代中期，数字信号处理的技术和方法取得了突破性进展，其主要标志是快速傅里叶变换（FFT）算法的成功应用。这样，出现了第一台以数字计算机为基础的孤立词语音识别器，继而又研制出第一台有限连续语音识别器。

20 世纪 70 年代之前，语音识别的研究特点是以孤立词的识别为主。进入 20 世纪 70 年代，语音识别的研究在多方面取得了诸多成就，在孤立词识别方面，日本学者 Sakoe 给出了动态规划方法进行语音设别的途径——DTW 算法，该算法是把时间归整和距离测度计算结合起来的一种非线性归整技术，是语音识别中一种非常成功的匹配算法。当时在小词汇量的研究中获得了成功，从而掀起了语音识别的研究热潮。

语音编码技术是伴随着语音的数字化而产生的，在语音编码方面，如何在中低速率上获得高质量的语音，一直是其研究的主要目标。20 世纪 70 年代中期，特别是 20 世纪 80 年代以来，语音编码技术有了突破性进展，提出了众多新型编码算法，产生了新一代的声码器，在 16 kbit/s 以下速率上能够得到高质量的语音。

20 世纪 80 年代初，林德（Linde）、布佐（Buzo）、格雷（Gray）等提出了矢量量化码本生成的方法，并将矢量量化技术成功地应用到语音编码中，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的作用，而且很快推广到其他领域。20 世纪 80 年代开始，语音识别研究的一个重要进展，就是识别算法从模式匹配技术转向基于统计模型的技术，隐马尔科夫模型（Hidden Markov Model，HMM）技术就是其中的一个典型。由于该模型能很好地描述语音信号的时变性和非平稳性，因此，从 20 世纪 80 年代起，它被广泛地应用到语音识别研究中。直到目前为止，HMM 方法仍然是语音识别研究中的主流方法。

近年来，计算机和集成电路技术的发展推动了语音信号处理的实用化。目前有很多专用语音处理芯片，这些芯片与微处理器或微型计算机相结合可以组成各种复杂的语音处理系统。其中语音合成在技术上比较成熟，在语音处理中影响也最大，上市产品多为有限词汇量的语音合成器。

然而，目前各种合成系统输出的语音音质跟自然语言的语音音质相差甚远，它并未真正解决机器说话的问题，因为其本质上只是一个不完满的声音还原过程。日前的合成只是停留在声道系统的发声过程上，其结果只是将书面语言转换成口头语言。而实现真正意义上的合成涉及到大脑的高级神经活动，而目前对这方面知道的还很少。同时，单音字节的声学语音学表现与音节在词语中有什么不同，尚未找到普遍的原则。由于涉及语言学、心理学和人脑的神经活动等问题，真正的语音合成问题尚处于研究阶段，这有待于信号处理、计算机、生理学、语言学、人工智能等领域研究人员的共同努力。

在语音识别方面，很多专业人员对其理论和应用进行了广泛的研究，关于这方面的文献浩瀚如海。目前，国内外有关论文每年达数千篇之多，但语音识别的研究比语音合成困难得多，其起步也较晚。

语音识别具有极其广泛的应用领域，但它毕竟是一项、难度很大的综合性高科技项目，

从话语中提取满意信息的过程是一项艰巨复杂的任务。虽然语音识别的研究已取得了很大进展，但还有很多困难甚至是原理性的问题有待解决。目前，语音识别领域的应用多是小词汇量特定人孤立词语音识别，是针对单个讲话者的，能够得到较高的识别率。

在语音识别中，必然涉及到人是怎样从声音中提取信息和理解含意的问题。只有弄清人在收听声音时的生理过程并研究出模仿这些过程的模型，语音识别才可能得到一个飞跃的发展。如何充分借鉴和利用人在完成语音识别和理解时所利用的方法和原理就是一大课题，因而语音识别与人工智能之间有密切的联系。而目前只能从语音信号出发，用“隐过程”（如隐马尔可夫模型）来进行神经系统和听觉过程的模拟，是无法达到理想的识别和理解的效果的。

语音信号处理的理论和研究包括紧密结合的两个方面：一方面是从语音的产生和感知来对其进行研究，这一研究与语音学、语言学、认知学、心理学和神经心理学等密不可分；另一方面是将语音作为一种信号进行处理，包括传统的数字信号处理技术以及前面提到的一些新的应用于语音信号的处理方法及技术。

1.3 语音信号处理的应用

语音信号处理技术是计算机智能接口与人机交互的重要手段之一。就语音识别技术而言，其基本任务是将输入语音转化为相应的文本或命令。语音识别的市场前景广泛，在一些应用领域正迅速成为一个关键的具有竞争力的技术。如在声控应用中，计算机识别输入的语音内容，并根据内容来执行相应的动作。这些应用包括声控电话转换、声控语音拨号系统、声控智能玩具、信息网络查询、家庭服务、宾馆服务、旅行社服务系统、医疗服务、银行服务、股票查询服务、工业控制等。语音识别也可用于将文字以口授的方式输入的计算机中，即广泛开展的听写机研究，如声控打字机等。语音识别技术还可以用于自动口语翻译，通过将口语识别技术、机器翻译技术、语音合成技术等相结合，可将一种语言输入的语音翻译为另一种语言的语音输出，实现跨语言的交流，如目前美国、日本、欧洲，包括中科院自动化所参加的 CSTAR 计划，正在重点开展多语种口语自动翻译研究。随着无处不在的计算技术的发展，各种移动计算设备、可穿戴计算设备日益增多，这些设备尺寸越来越小，并且要求在行走或驾驶时进行信息的输入，传统的键盘输入方式已不能满足其方便、自然、在行进中有效地输入信息的需要，采用语音识别技术可以解放用户的双手和眼睛，有效地改变人机交互手段。如目前在一些手持计算机、手机等嵌入式电子产品上已经使用语音识别技术来进行控制。

对说话人识别技术，近年来已经在安全加密、银行信息电话查询服务等方面得到了很好的应用，此外，在公安机关破案和法庭取证等方面也发挥着重要的作用。

就语音合成而言，它已经在诸多方面得到了实际应用，发挥了很好的社会效益，如公共交通中的自动报站、各种场合的自动报时、自动告警、电话自动查询服务、文本校对中的语音提示等。在电信声讯服务领域的智能电话查询系统中，采用语音合成技术可以解决以往通过电话只能进行静态查询的不足，满足海量数据和动态查询的需求，可查询一些动态的信息，如股票、成绩、节目、热点问题、机场、车站、购物、市场、售后服务等信息；也可用于基于个人计算机的办公、教学、娱乐等智能多媒体软件，如文稿校对、语音学习（帮助外国人、残疾人、儿童等学习语言）、语音秘书、语音书籍、教学软件、语音玩具等。通过与互联网的结合，可以获取有声的 E-mail、进行网上信息的有声获取及进行网上语音聊天。将语音合成

技术与机器翻译技术相结合，可以实现语音翻译；与图像技术相结合，可以输出视觉语音（Visual Speech）。

就语音编码技术而言，它的根本作用是使语音通信数字化，目前已广泛应用于数字通信系统、移动无线通信、保密语音通信等方面。语音编码技术也可应用于呼叫服务，如数字录音电话、语音信箱、电子留言薄等。与模拟语音通信系统相比，数字语音通信系统具有抗干扰性强、保密性好、易于集成化等优点。在当前正在蓬勃兴起的移动通信中，语音编码技术是其中非常重要的支撑技术。

随着信息技术的不断发展，尤其是网络技术的日益普及和完善，语音信号处理技术正发挥着越来越重要的作用，并且出现了一些新的研究方向。

基于语音的信息检索是随着网络技术及面向数字图书馆技术的发展而出现的新的应用技术。传统的信息检索技术大多是基于文本信息的，诸如雅虎、谷歌等各种搜索引擎，就是这方面的典型应用。随着语音识别技术的不断发展和完善，基于语音识别的信息检索技术正成为当今的研究热点。

随着 Internet 网络技术的迅速发展，出现了 Internet 电话技术，它是一种用 VoIP（Voice over Internet Protocol）技术实现的通过 TCP/IP 网络而不是传统的电话网络来传输语音的新的通信方式，通常称为 IP 电话技术。对这种经过数据压缩，并经过网络以数据包形式传输后的语音进行识别，与传统的语音识别技术有着很大的不同，这提出了一个新的研究课题，即网络环境下的语音识别问题，它在电子商务和国防军事应用领域有着广阔的应用前景。而随着手持计算机、手机等电子设备的迅猛发展，研制开发这些设备上嵌入式的语音识别算法越来越引起人们的重视，目前已经出现了一些可用语音识别进行声音拨号，以及口述关键词进行信息查询的手机，这类技术的不断完善对移动技术的发展有着重要的意义。

语音训练与校正技术也是近年来的一个重要研究方向。当今社会越来越多的人希望学习和掌握其他的非母语语言，以利于更方便地进行交流。因此，语言学习已成为当今教育领域的一个热点。实践证明，采用传统的课堂教学对于学习一门非母语语言来说是远远不够的。自学是一种有效的途径，它具有不受时间地点限制、灵活方便等特点。随着计算机技术的迅速发展，一种称之为计算机辅助语言学习（Computer-Aided Language Learning，CALL）的技术应运而生，伴随着语音识别技术的进步，人们开始研究进行辅助发音学习的 CALL 技术。在发音学习中，有效地反馈是必不可少的一个重要环节。在课堂教学中，教师是一个有效的反馈源，而传统的发音自学中，要么是没有任何反馈，要么就是反馈最终还得依赖于学习者自身的判断能力，如利用复读机学习发音时，学习者只能依靠自己的感知能力去比较其发音与标准发音的差别，从而进行发音的修正。如果利用辅助发音的 CALL 系统，学习者就可以随时获得有效的反馈，包括分值或等级等简洁直观的形式，图谱或口形等具体形象的形式，以及直接的指导性建议。

语种识别（Language Identification）也是近年来出现的研究方向，它是通过分析处理一个语音片段以判断其所属语言的种类，本质上也是语音识别的一个方面。由于世界上的不同语种间有着多种区别性特征，如音素集合、音位序列、音节结构、韵律特征、词汇分类、语法及语义网络等，所以在自动语种识别中有多种可以利用的特征。对于一个语种识别系统，它和语音识别系统与说话人识别系统有着很多相似之处，如都要经过数字化、特征提取、模式匹配等过程。语种识别可以应用于多语言语音识别的前端处理，在信息检索、军事领域和

国家安全事务中有着重要的作用。

基于语音的情感处理研究是当今一个重要的研究方向。在人与人交流中，除了言语信息外，非言语信息也起着非常重要的作用。随着计算机技术的迅速发展，人机交流变得越来越普遍，计算机正成为日常生活工作中的得力助手。为使人机交流更自然、更人性化，十分有必要进行人机非言语交流方式的研究。尽管人们早已认识到非言语交流的重要性，但时至今日，大多数研究还仅仅是基于视觉信息的工作，如面部表情识别、手势识别等。语音作为语言的表现形式，是人类交流信息最自然、最有效、最方便的手段。人类的语音中不仅包含了语言学信息，同时也包含了人们的感情和情绪等非言语信息。例如，同样一句话，往往由于说话人的情感不同，其意思和给听者的感觉就会不同。传统的语音处理系统仅仅着眼于语音词汇传达的准确性，而完全忽视了包含在语音信号中的情感因素，所以它只是反映了信息的一个方面。直到近年人们才发现，由于情感和态度所引起的变化对语音合成、语音识别、说话人确认的影响较大，并逐步重视起来。目前许多研究者都在致力于研究情感对语音的影响，以及情感状态下语音信号处理的有效方法。

1.4 本书内容

本书系统介绍了语音信号处理的原理、方法与应用，以及新方法和新技术。全书共分15章，其中第2章介绍了语音处理需要的一些基础知识，包括语言和语音的基本特点；语音生成、语音感知等语音学、生理学和心理学基础。为了突出重点和节省篇幅，这一章只介绍与本书其余内容有直接关系的最基本的部分，如需进一步了解可参阅书中列出的参考文献。从第3章开始介绍了语音的各种分析和处理技术，包括经典方法，如时域分析、频域分析等各种新技术；同态处理、线性预测分析、矢量量化及隐马尔可夫模型技术等；还介绍了语音信号处理的各种应用，包括基因提取与共振峰估值、波形编码、声码器、语音合成、语音识别、说话人识别及语音增强等。

思考与练习

1.1 语音信号处理主要研究哪几方面的内容？

1.2 列举工农业生产、人民生活中的5种语音信号处理应用技术或产品。简述其工作原理。

第2章 语音信号处理的基础知识

在研究和分析各种语音信号处理技术及其应用之前，必须了解有关语音信号的一些基本特性。为了对语音信号进行数字处理，需要建立一个能够精确描述语音产生过程和语音全部特征的数学模型，即根据语音的产生过程建立一个既实用又便于分析的语音信号模型。为了处理和实现上的方便，这个模型应尽可能简单。

本章首先对语音的产生过程进行分析，这属于发音语音学的内容；接着介绍了语音的声学特性，这属于声学语音学的内容；然后给出了传统的线性语音产生的数字模型，这是以后各章讨论的基础；最后对人耳的听觉过程进行分析，即语音感知，这属于听觉语音学的内容。

这些都是从事语音信号处理研究的基础知识，对于语音信号处理的任何一个研究领域都是必需的，其中贯穿全书的是语音信号产生的数字模型。

2.1 语音和语言

构成人类语音的是声音，然而这是一种特殊的声音，是由人讲话所发出的声音。声音是一种波，能被人耳听到，它的振动频率在20～20 000Hz之间。语音是由一连串的音所组成，它是组成语言的声音。语音具有成为声学特征的物理性质。语音中各个音的排列由一些规则所控制，对这些规则及其含义的研究属于语言学的范畴，而对语音中音的分类和研究称为语音学。

形成文章的基础是单词，单词简称词，是有意义的语言的最小单元。各单词由音节组成，音节又由音素组成，所谓音素是语言的元素，即语言的最小基本单位，是发出各不相同音的最小单位。也就是说，音素都有其独立的各不相同的发音方法和发音部位，它是让听者能区别一个单词和另一个单词的声音的基础。音素分为两类：元音和辅音。在已知的语言中元音有少至两个而多至12个，辅音从10多个至70多个。英语中有43个音素。而音节的定义不一定明确，但是一个音节可以是1个元音和1～2个辅音组合。实际上，各个音素组合而构成语言时的连接方法有几种限制，并不是所有的组合都存在。因此，一种语言中所用的音节数远少于音素的组合数。

重音、语调和声调也是构成语言学的一部分，它们或者用来表示一句话中重要的单词，或者用来表示疑问句，或者用来表示说话人的感情。重音和语调是一种附加的信息，其中词的重音是外方语言（如英语）的一个重要特点，而语调实际是讲话声音的调节，它决定于诸

多因素，如语气、环境、讨论的话题等。语音中还有一个问题是同音异义词，它是指有相同的语音但是有两个或更多的不同意义。如汉语中的“语”、“与”、“雨”，英语中的“site”、“sight”“cite”等就是同音异义词。语音除了上述一些特点外，还存在所谓超语言学特点，如低语表示秘密，高声说话表示愤怒等。

对于我们所使用的汉语，有其特殊的、不同于英语的特点。汉语里也有元音和辅音的不同，其中不同的元音是由不同的口腔形状造成的，而不同的辅音是由发音部位和发音方法不同造成的。但是，汉语语音分析中总是把一个汉语音节分为声母和韵母两部分：声母就是一个汉语音节开头的辅音，而韵母是汉字音节除了开头的声母以外的部分。在汉语中，有 21 个声母和 39 个韵母。

汉语的特点为汉语的自然单位是音节，每一个字都是单音节字，即汉语的一个音节就是汉语的一个字的音，这里字是独立的发音单位。再由音节字构成词（其中主要是两音节字构成的词），最后再由词构成句子。而每一个音节字又都是由声母和韵母拼音而成；在音节中，声母比较简单，它们只是一个音素；而韵母则比较复杂。

汉语语音的另一个重要特点是它具有声调（即音调在发一个音节中的变化），这使得它在使用语声时较其他语言更为经济。我国公布的汉语拼音方案中采用声调这个词。声调是一种音节在念法上的高低升降的变化。汉语有 4 种声调，即阴平（-）、阳平（ˊ）、上声（ˇ）、去声（ˋ），上面括号内表示的是该声调的符号。由于有声调之分，所以参与拼音的韵母又有若干种（包括轻声在内至多有 5 种）声调。

汉语的特点是音素少、音节少。它大约有 64 个音素，但只有 400 个左右音节，即 400 个基本的发音。如考虑每个音节有 5 个声调，也只不过有 1 200 多个有调音节，即不同的发音。

2.2 语音产生的过程及其声学特性

人类生成语音过程的第一阶段是决定想传给对方的内容是什么，然后将内容转换成语言的形式。选择表现其内容的适当语句，将其按文法规则排列，便能构成语言的形式。由大脑对发音器官发出运动神经指令，发音器官各种肌肉运动振动空气而形成语音波。这个过程可分为神经和肌肉的生理学阶段和产生语音波、传递语音波的物理阶段。

人类能以语言沟通，进而累积知识，形成文化，其中一个主要的原因，就是人类具有较其他生物优越的发音器官。人类的发音器官能够产生多样性的声音，构成丰富的词汇。

2.2.1 语音的发音器官

人的发音器官包括肺、气管、喉（包括声带）、咽、鼻和口等，如图 2.1 所示。这些器官共同形成一条形状复杂的管道，其中喉以上的部分称为声道，随着发出声音的不同其形状是变化的；而喉的部分称为声门。在发音器官中，肺是语音产生的能源所在，喉是主要的声音生成机构，而声道则对生成的声音进行调制。

产生语音的能量，来源于正常呼吸时肺部呼出的稳定气流，喉部的声带既是阀门，又是振动部件。在说话的时候，声门处气流冲击声带产生振动，然后通过声道响应变成语音。由于发不同的音时，声道的形状不同，所以听到不同的声音。以上就是发音器官发出声音时的大致情况。