高等院校通信与信息专业规划教材

语音信号处理

第2版

SPEECH SIGNAL PROCESSING

赵力　编著

机械工业出版社
CHINA MACHINE PRESS

高等院校通信与信息专业规划教材

语音信号处理

第2版

赵　力　编著

机械工业出版社

本书介绍了语音信号处理的基础、原理、方法和应用，以及该学科领域近年来取得的一些新的研究成果和技术。全书共分14章，包括绪论、语音信号处理基础知识、语音信号分析、矢量量化技术、隐马尔可夫模型、神经网络在语音信号处理中的应用、语音编码、语音合成、语音识别、说话人识别与语种辨识、语音转换与语音隐藏、语音信号中的情感信息处理、耳语音信号处理、语音增强等内容。

本书可作为高等院校教材或教学参考用书，也可供从事语音信号处理等领域的工程技术人员参考。

图书在版编目（CIP）数据

语音信号处理/赵力编著. —2版. —北京：机械工业出版社，2009.5
（高等院校通信与信息专业规划教材）
ISBN 978-7-111-27190-1

Ⅰ.语… Ⅱ.赵… Ⅲ.语言信号处理-高等学校-教材 Ⅳ.TN912.3

中国版本图书馆CIP数据核字（2009）第077942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：李馨馨
责任印制：洪汉军

北京市朝阳展望印刷厂印刷

2009年6月第2版·第1次印刷
184mm×260mm · 21.5印张 · 529千字
0001—3000册

标准书号：ISBN 978-7-111-27190-1
定价：36.00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 68993821
购书热线电话：（010）88379639 88379641 88379643
编辑热线电话：（010）88379753 88379739
封面无防伪标均为盗版

高等院校通信与信息专业规划教材
编委会名单

（按姓氏笔画排序）

出 版 说 明

为了培养21世纪国家和社会急需的通信与信息领域的高级科技人才，为了配合高等院校通信与信息专业的教学改革和教材建设，机械工业出版社会同全国在通信与信息领域具有雄厚师资和技术力量的高等院校，组成阵容强大的编委会，组织长期从事教学的骨干教师编写了这套面向普通高等院校的通信与信息专业规划教材，并且将陆续出版。

这套教材将力求做到：专业基础课教材概念清晰、理论准确、深度合理，并注意与专业课教学的衔接；专业课教材覆盖面广、深度适中，不仅体现相关领域的最新进展，而且注重理论联系实际。

这套教材的选题是开放式的。随着现代通信与信息技术日新月异的发展，我们将不断更新和补充选题，使这套教材及时反映通信与信息领域的新发展和新技术。我们也欢迎在教学第一线有丰富教学经验的教师及通信与信息领域的科技人员积极参与这项工作。

由于通信与信息技术发展迅速而且涉及领域非常宽，这套教材的选题和编审难免有缺点和不足之处，诚恳希望各位老师和同学提出宝贵意见，以利于今后不断改进。

机械工业出版社
高等院校通信与信息专业规划教材编委会

前　言

本书是根据机械工业出版社高等院校通信与信息专业规划教材编审出版规划，由通信与信息专业规划教材编审委员会编审、推荐出版的。自从2003年3月第1版出版以来，时间已过去了近6个年头。几年来，随着我国高等教育的发展和教学要求的提高，特别是本学科领域技术的进步，以及新的应用需求的不断提高，相应地对本教材内容的更新提出了紧迫的要求和更高的标准。正是在这样的背景下，编者在保持教材总体格局不出现大变化的前提下，对第1版教材进行了修订、补充和部分更新。

新版教材力求系统地反映语音信号处理的基本原理和方法，以及近年来该领域的新进展和新技术；突出基本概念、原理、方法、应用、研究现状及学科发展趋势，而不是去过多追求数学推导和证明的严谨性。在篇幅上，按照基础—分析—处理与应用的顺序组织材料；在选材上，使其既能满足教学需要，又反映出本学科领域近年来发展的新成果。

第2版教材除了增减了部分章节以外，基本保持了原作风貌。总体结构同第1版基本相同，认真修订了第1版的部分错误和疏漏。在内容的增删与更新方面，根据作者多年来给本科生和硕士研究生讲授“语音信号处理”课程的体会，除了对部分较烦杂的内容进行了删减以外，还增加了一些现在较流行的内容，如基于小波的语音参数分析技术、语音转换和语音隐藏技术、耳语音信号处理技术等。

本书主要面向信号与信息处理、电路与系统、通信与电子工程、模式识别与人工智能、计算机信息处理等学科有关专业的高年级学生和研究生，也可以作为从事语音信号处理这一领域科研工作的技术人员的参考书。

本书的参考学时为本科生32学时、研究生40学时，可以根据不同的教学要求对其内容进行适当取舍，灵活安排讲课学时数。

语音信号处理是一门理论性强、实用面广、内容新、难度大的交叉学科，同时这门学科又处于快速发展之中，尽管作者在编写过程中始终注重理论紧密联系实际，力求以尽可能简明、通俗的语言，深入浅出地将这门学科介绍给读者，但因作者水平有限，缺点错误在所难免，敬请广大读者批评指正。

为了配合本书教学，作者为本书提供了配套的电子教案，读者可在机械工业出版社教材服务网（http://www.cmpedu.com）下载。

作　者

2009年2月

目　录

第1章 绪 论

通过语音传递信息是人类最重要、最有效、最常用和最方便的交换信息的形式。语言是人类特有的功能，声音是人类常用的工具，是相互传递信息的最主要的手段。因此，语音信号是人们构成思想疏通和感情交流的最主要的途径。并且，由于语言和语音与人的智力活动密切相关，与社会文化和进步紧密相连，所以它具有最大的信息容量和最高的智能水平。现在，人类已开始进入了信息化时代，用现代手段研究语音处理技术，使人们能更加有效地产生、传输、存储、获取和应用语音信息，这对于促进社会的发展具有十分重要的意义。

让计算机能听懂人类的语言，是人类自计算机诞生以来梦寐以求的想法。随着计算机越来越向便携化方向发展，以及计算环境的日趋复杂化，人们越来越迫切要求摆脱键盘的束缚而代之以语音输入这样便于使用的、自然的、人性化的输入方式。尤其是汉语，它的汉字输入一直是计算机应用普及的障碍，因此利用汉语语音进行人机交互是一个极其重要的研究课题。作为高科技应用领域的研究热点，语音信号处理技术从理论的研究到产品的开发已经走过了几十个春秋并且取得了长足的进步。它正在直接与办公、交通、金融、公安、商业、旅游等行业的语音咨询与管理，工业生产部门的语声控制，电话·电信系统的自动拨号、辅助控制与查询以及医疗卫生和福利事业的生活支援系统等各种实际应用领域相接轨，并且有望成为下一代操作系统和应用程序的用户界面。可见，语音信号处理技术的研究将是一项极具市场价值和挑战性的工作。我们今天进行这一领域的研究与开拓就是要让语音信号处理技术走入人们的日常生活当中，并不断朝向更高目标而努力。

语音信号处理这门学科之所以能够长期地、深深地吸引广大科学工作者不断地对其进行研究和探讨，除了它的实用性之外，另一个重要原因是，它始终与当时信息科学中最活跃的前沿学科保持密切的联系，并且一起发展。语音信号处理是以语音语言学和数字信号处理为基础而形成的一门涉及面很广的综合性的学科，与心理·生理学、计算机科学、通信与信息科学以及模式识别和人工智能等学科都有着非常密切的关系。对语音信号处理的研究一直是数字信号处理技术发展的重要推动力量。因为许多处理的新方法的提出，首先是在语音处理中获得成功，然后再推广到其他领域的。例如，许多高速信号处理器的诞生和发展是与语音信号处理的研究发展分不开的，语音信号处理算法的复杂性和实时处理的要求，促使人们去设计许多先进的高速信号处理器。这种产品问世之后，又首先在语音信号处理应用中得到最有效的推广应用。语音信号处理产品的商品化对这样的处理器有着巨大的需求，因此它反过来又进一步推动了微电子技术的发展。

语音信号处理作为一个重要的研究领域，有很长的研究历史。但是它的快速发展可以说是从1940年前后Dudley的声码器(Vocoder)和Potter等人的可见语音(Visible Speech)开始的。1952年贝尔(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和Belar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年代前期，由于Faut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、快速傅里叶变换（FFT）等成为语音信号数字处理的理论和技术基础。在方法上，随着电子计算机的发展，以往的以硬件为中心的研究逐渐转化为以软件为主的处理研究。然而，在语音识别领域内，初期有几种语音打字机的研究也很活跃，但后来已全部停了下来，这说明了当时人们对语音识别难度的认识得到了加深。所以 1969 年美国贝尔研究所的 Pierce 感叹地说："语音识别向何处去？"

到了 1970 年，好似反驳 Pierce 的批评，单词识别装置开始了实用化阶段，其后实用化的进程进一步高涨，实用机的生产销售也上了轨道。此外社会上所宣传的声纹（Voice Print）识别，即说话人识别的研究也扎扎实实地开展起来，并很快达到了实用化的阶段。到了 1971 年，以美国 ARPA（American Research Projects Agency）为主导的"语音理解系统"的研究计划也开始起步。这个研究计划不仅在美国国内，而且对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，取得了许多成果，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 70 年代期间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。这就是 70 年代初由板仓（Itakura）提出的动态时间规整（DTW）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（LPC）被用于语音信号处理，此后隐马尔可夫模型法（HMM）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（VQ）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 年代期间，由于矢量量化、隐马尔可夫模型和人工神经网络（ANN）等相继被应用于语音信号处理，并经过不断改进与完善，使得语音信号处理技术产生了突破性的进展。其中，隐马尔可夫模型作为语音信号的一种统计模型，在语音信号处理的各个领域中获得了广泛的应用。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（CMU）的 Baker 和美国 IBM 公司的 Jelinek 等人将其应用到语音识别中。由于美国贝尔实验室的 Rabiner 等人在 80 年代中期，对隐马尔可夫模型深入浅出的介绍，才使其被世界各国从事语音信号处理的研究人员所了解和熟悉，进而成为一个公认的研究热点，也是目前语音识别等的主流研究途径。

进入 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。其中，语音识别逐渐由实验室走向实用化。一方面，对声学语音学统计模型的研究逐渐深入，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。另一方面，为了语音识别实用化的需要，讲者自适应、听觉模型、快速搜索识别算法以及进一步的语言模型的研究等课题倍受关注。

在语音合成方面，有限词汇的语音合成群已在自动报时、报警、报站、电话查询服务、发音玩具等方面得到了广泛的应用。关于文本—语音自动转换系统（TTS）的研究，许多国家、多个语种都已在 20 世纪 90 年代初达到了商品化程度，其语音质量能被广大公众接受。从研究技术上可分为发音器官参数合成、声道模型参数合成和波形编辑合成；从合成策略上可分为频谱逼近合成和波形逼近合成。其中采用波形拼接来合成语音的方法，越来越被广泛地应用。其

中最具代表性的是基音同步叠加法(PSOLA),这种方法既能保持所发语音的主要音段特征,又能在拼接时灵活调整其基频、时长和强度等超音段特征,在语音合成中影响较大。

在50多年的时间里,语音编码已取得了迅速的发展。最早的标准化语音编码系统是速率为64kbit/s的PCM波形编码器;到90年代中期,速率为4~8kbit/s的波形与参数混合编码器,在语音质量上已接近前者的水平,且已达到实用化阶段。当前的研究主要集中在4kbit/s码率以下的高音质、低延迟的声码器,提高在噪声信道中低码率编码器的性能,并能传输多种信号,包括音频信号。为此在寻找更为有效的参数量化技术、非线性预测技术(Non-Linear Predction)、多分辨率时频分析技术(如Wavelets)和高阶统计量的使用、对人耳感知特性的进一步研究和探索等方面有较多的研究工作。

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其是哪个语种,所以就要尽可能找出不同语种的差别特征。目前,这方面的研究重点转向对各种声学参数的线性或非线性处理以及新的模式匹配方法上,如DTW、主分量(成分)分析(PCA)、隐马尔可夫模型与人工神经网络组合等技术上。

语音转换就是保持语义信息不变,仅改变一个说话人的语音个性特征(称为源说话人),使其听起来像是另一个说话人(称为目标说话人)的语音个性特征。语音转换作为语音信号处理领域的一个新兴的分支,具有重要的研究价值和应用前景。语音信息伪装和语音数字水印技术同属语音信息隐藏。目前国际学术界对信息隐藏研究的侧重点在于商用的数字水印系统,对于信息伪装的研究还比较少,而专门针对语音的伪装方法的研究则更少。因此,对语音信息伪装的研究及其对隐藏容量的分析,可为保密通信的研究与设计提供一个新的发展方向。

包含在语音信号中的情感信息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部分信息。例如,同样的一句话,由于说话人表现的情感不同,在给听者的感知上就可能会有较大的差别。所谓"听话听音"就是这个道理。然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,通过规则化噪声处理给去掉了。实际上,人们是同时接受各种形式的信息的,怎样有效地利用各种形式的信息以达到最佳的信息传递和交流效果,是今后信息处理研究的发展方向。所以包含在语音信号中的情感信息的计算机处理研究,分析和处理语音信号中的情感特征、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。

耳语音是人们常见的语言交流方式之一,在会场、音乐厅、图书馆等禁止大声喧哗的场所被广泛使用。因此,耳语音的研究具有广泛的应用前景。早期由于技术条件的限制,关于耳语音的研究主要停留在基础语音学和医学工作领域。随着科学技术的发展,近年来对耳语音的研究逐渐走向多领域和实际应用,如耳语音转换为正常音、耳语音的语音识别和说话人识别、耳语音的语音增强等。

有关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,已经做了大量的研究工作,取得了丰硕的研究成果。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算法等;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研究发展,利用这些领域的研究成果进行语音和噪声分离的研究取得了一些进展。

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。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·语言学、认知科学、心理·生理学等学科密不可分。另一方面是将语音作为一种信号来进行处理,包括传统的数字信号处理技术以及一些新的应用于语音信号的处理方法和技术。

本书将系统介绍语音信号处理的基础、原理、方法和应用。全书共分14章,其中第2章介绍了语音信号处理的基础知识,如语音·语言学、汉语语音学、发音与听觉器官、语音信号的数学模型、语音信号的统计特性分析等;第3章介绍了语音信号特征分析和处理技术,包括时域分析、频域分析、同态分析、线性预测分析、音调检测和共振峰检测方法等。为了突出重点和节省篇幅,书中对语音信号处理的基础知识部分和语音信号特征分析和处理技术部分等进行了压缩,目的是将主要篇幅放在语音信号处理应用的原理与方法的阐述上,力求提高读者实际应用语音信号处理技术的能力。从第7章开始介绍了语音信号处理的各种应用,包括语音编码、语音合成、语音识别、说话人识别和语种辨识、语音转换和语音隐藏、语音信号中的情感信息处理、耳语音信号处理以及语音增强等。为了帮助读者理解和掌握语音信号处理的各种应用知识,便于学习和教学,本书在第7章开始介绍语音信号处理应用的原理与方法之前,专门安排3个章节,介绍了当前语音信号处理应用的3个主流技术,即在第4章介绍了矢量量化技术;在第5章介绍了隐马尔可夫模型技术;在第6章介绍了人工神经网络在语音信号处理中的应用技术等。

语音信号处理是目前发展最为迅速的信息科学技术之一,其研究涉及一系列前沿课题,且处于迅速发展之中。因此本书的宗旨是在系统地介绍语音信号处理的基础、原理、方法和应用的同时,向读者介绍该学科领域近年来取得的一些新成果、新方法及新技术。数字语音信号处理属于应用科学,要学好这门课程,关键在于理论必须联系实际应用,才能很好地掌握数字语音处理的理论和技术方法。因此,本书在每一章后面都附有课外思考题,并且在全书的最后附有两个语音处理的实用程序。建议学习者仔细选做书中的习题,并进行计算机上机实验以获得实际经验,以尽快掌握所学的知识。

第2章　语音信号处理基础知识

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。它的目的有两个：一个是要通过处理得到一些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号信息；另一个是要通过处理某种运算以达到某种用途的要求，例如人工合成出语音、辨识出讲话者、识别出讲话的内容等。因此，在研究各种语音信号数字处理技术应用之前，首先需要了解语音信号的一些重要特性，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等。它们是贯穿整个语音信号数字处理的基础。

2.1　语音和语言

人们讲话时发出的话语叫语音，它是一种声音，具有称为声学特征的物理特性。然而，它又是一种特殊的声音，是人们进行信息交流的声音，是组成语言的声音。因此，语音(Speech)是声音(Acoustic)和语言(Language)的组合体。可以这样定义语音，语音是由一连串的音组成语言的声音。所以对语音的研究包括两个方面，一个是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称为语言学；另一个是对语音中各个音的物理特征和分类的研究，称为语音学。

语音和语言是研究人类话语的一门科学。所以，研究语音和语言之前首先要了解一下人说话的过程。

人的说话过程如图2-1所示，可以分为5个阶段。

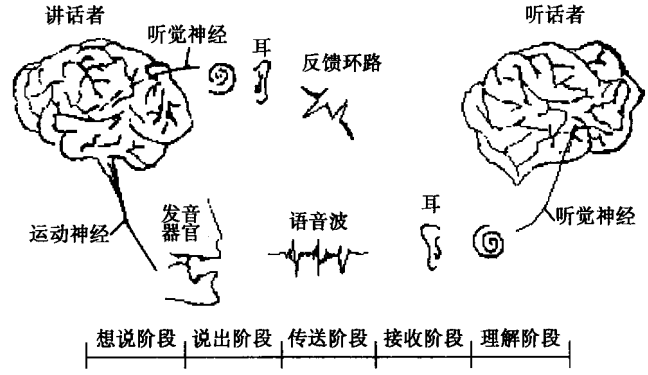

图2-1　人的说话过程

(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。

(2) 说出阶段

由想说阶段大脑中枢的决策,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,来帮助修改语音。这就是:他不但发出语音,而且他自己的听觉系统也在听自己的话语。但是,在这个阶段中,主要是与发音器官的活动有关。

(3) 传送阶段

说出来的话语是一连串声波,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。

(4) 接收阶段

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。

(5) 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。

从这 5 个阶段来看,说话的过程包括着相当复杂的因素,其中有心理的、生理的、物理的以及个人的和社会的因素。这里,个人的因素是指讲话的口音和用词造句的特色以及听话者的听力和理解能力;社会的因素则是指讲话者和听话者对用于进行交际的手段有共同的理解的社会基础。

语言是从人们的话语中概括总结出来的规律性的符号系统。包括构成语言的语素、词、短语和句子等的不同层次的单位,以及词法、句法、文脉等语法和语义内容等。句法的最小单位是单词,词法的最小单位是音节。不同的语言有不同的语言规则。语言学是语音信号处理的基础。例如,可以利用句法和语义信息减少语音识别中搜索匹配范围,提高正确识别率。随着现代科学和计算机技术的发展,除了人与人之间的上述自然语言的通信方式之外,人机对话及智能机器人等领域也开始使用语言了。这些人工语言同样有词汇、语法、句法结构和语义内容等。因此,语言学又称为自然语言处理,它是一门专门的学科。

语音学(Phonetics)是研究言语过程的一门科学。它考虑的是语音产生、语音感知等的过程以及语音中各个音的特征和分类等问题。从某种意义上讲,语音学与语音信号处理这门学科联系的更紧密。正如上面所介绍的一样,人类的说话交流是通过联结说话人和听话人的一连串心理、生理和物理的转换过程实现的,这个过程分为发音、传递、感知三个阶段。因此现代语音学发展成为与此相应的 3 个主要分支:发音语音学、声学语音学、听觉语音学。

发音语音学(Articulatory Phonetics):发音语音学也称生理语音学,主要研究语音产生机理,借助仪器观察发音器官,以确定发者部位和发音方法。这一学科在 19 世纪中期就已经形成,近年来由于新型仪器设备的发明和改进,又有很大发展,目前已相当成熟。

声学语音学(Acoustic Phonetics):声学语音学研究语音传递阶段的声学特性,它与传统语音学和现代语音分析手段相结合,用声学和非平稳信号分析理论来解释各种语音现象,是近几十年中发展非常迅速的一门新学科。