计算语言学研究系列

陈小荷 主编

语音韵律的实验分析与建模

Experimental Analysis and Quantitative Modeling of Speech Prosody

顾文涛 著

世界图书出版公司

计算语言学研究系列　陈小荷主编

语音韵律的实验分析与建模

EXPERIMENTAL ANALYSIS AND QUANTITATIVE MODELING OF SPEECH PROSODY

顾文涛　著

世界图书出版公司
北京·广州·上海·西安

图书在版编目(CIP)数据

语音韵律的实验分析与建模/顾文涛著. —北京：世界图书出版公司北京公司，2013.1

ISBN 978-7-5100-5652-9

Ⅰ.①语… Ⅱ.①顾… Ⅲ.①汉语—韵律(语言)—文集

Ⅳ.①H11-53

中国版本图书馆 CIP 数据核字（2012）第 316597 号

语音韵律的实验分析与建模

著　　者：顾文涛

责任编辑：梁沁宁

出　　版：世界图书出版公司北京公司

出 版 人：张跃明

发　　行：世界图书出版公司北京公司

（地址：北京朝内大街 137 号　邮编：100010　电话：64077922）

销　　售：各地新华书店和外文书店

印　　刷：三河市国英印务有限公司

开　　本：711 mm × 1245 mm　1/24

印　　张：17

字　　数：422 千

版　　次：2013 年 1 月第 1 版　2013 年 1 月第 1 次印刷

ISBN 978-7-5100-5652-9　　定　　价：39.00 元

《语言科技文库》总序

李葆嘉

当代语言学已经进入了一个科学与技术的互补时代，信息处理水平成为衡量国家现代化水平的重要标志之一。知识世界的载体是语符系统，信息处理的根本对象是语言信息处理。与计算机的出现使得语言符号有可能成为数据处理对象相似，神经科学实验仪器设备的应用，使得在大脑神经层面探讨语言机制成为可能。这些无疑都引导语言研究走向科技化，“语言科技新思维”（李葆嘉 2001）应运而生。所谓“语言科学”包括理论语言学、描写语言学、历史语言学、应用语言学等分支学科，所谓“语言技术”指语言研究的现代技术手段，包括语言信息处理、语音实验分析，以及语言的神经、心理和行为实验分析的技术手段等。就语言信息处理而言，又可以分为语料库研制技术、知识库研制技术、知识挖掘和抽取技术、句法信息处理技术、词汇信息处理技术、语音信息处理技术、语义信息处理技术、语用信息处理技术等。

2001 年 5 月，南京师范大学文学院创办了史无前例的“语言科学及技术系”，率先迈出了从传统文科教育范型向现代科技教育范型转变的步伐。“十五”期间，南京师大“211 工程”重点学科建设项目“语言信息处理与分领域语言研究的现代化”（陈小荷教授主持），以基础平台建设、资源建设和理论探索等为主，迈出了语言科技研究的一大步。

“十一五”期间，南京师大文学院、外国语学院和国际文化教育学院联袂申报“211 工程”三期重点学科建设项目。该项目以“语言科技”为引导，以“多学科交叉、跨院系整合、开放型营运”为理念，建设具有前瞻性、原创性、成长性的语言科技高级工作平台。以典型课题的工作原理为核心，进行资源开发和系统研制，拓展语

音科技、二语习得的神经机制研究、言语能力受损儿童的语言能力研究等新方向。同时造就新一代学术领军人物和培养一批高层次复合型人才，以期形成一支高水平的交叉学科团队。该项目设计，体现了工作平台建设、理论创新、应用研究、人才培养、团队建设的学科发展一体化思路。其旨趣在于，加速语言研究从传统文科范型向现代科技范型的转变，以引领21世纪语言科技的新潮流。

作为新兴交叉学科项目，通过教育部组织的专家匿名评审，“语言科技创新及工作平台建设”（2008～2011）获批，总投入1 000万元。总体而言，这一“语言科技创新”团队，分支学科齐全，专业知识互补。涵盖了理论语言学、计算语言学、语义科技、语音科技、实验方言学、历史语言学、神经语言学、二语习得研究、话语行为语言学等领域。这一期间，项目组成员获批的国家级基金项目达20多项。该项目理念之前瞻、实力之雄厚、工程之浩大、经费之保障，为学界瞩目。

2008年秋，本项目以南京师范大学语言科技研究所为实施单位正式启动。主要有三大任务：建设一个领先性的语言信息科技实验室、建立一个独创性的语言科技工作平台、撰著一套有特色的语言科技文库。

从实验室方案设计到设备招标采购，再到实验室用房改造，经过8个月的努力，2009年12月，语言信息科技实验室建成，为语言研究从传统范型向科技范型的转变提供了基本保障。该实验室划分为实验工作区、科研工作区和管理服务区。实验工作区建有语音实验与计算室、神经认知实验与计算室、课堂话语实录室三个专门实验室。科研工作区建有语义科技工作室、语音科技工作室、方言实验工作室、知识工程工作室Ⅰ（先秦词汇）、知识工程工作室Ⅱ（中古词汇）、知识工程工作室Ⅲ（敦煌俗语言文字）、语言习得神经机制工作室、语言习得中介机制工作室，以及参研工作室。管理区服务包括办公室、管理室、编辑室和交流室。出席“语言科技高层论坛暨语言信息科技实验室落成仪式”（2009年12月14日）的专家认为，该实验室体现了语言学跨学科研究的当代性和先进性，具有整体性、科技型、开放型三个特点，处于全国领先地位，是“语言科技新思维”的又一体现。同时认为，该实验室的科研工作涵

盖了四个二级学科、四个博士学位点，有稳定明确的研究方向，有合理的设计规划和很好的科研基础；整体设计合理，功能齐备。以教育部重点实验室建设标准衡量，很多方面超过了指标。

语言科技工作平台是基于工作原理（课题定位—理论方法—技术路线—关键技术—评估方式）而建设的高级平台。一方面，从语言信息、语言知识和语言机制三个层面，围绕典型课题进行设备配置、资源建设和软件开发；一方面，将典型课题研究与工作平台建设融为一体，依据典型课题建设的子平台应具有解决同类课题的功能。

建设语言科技工作平台的目标是要实现语言研究手段的技术化和模型化，总体设计包括三个二级平台和八个子系统。

一、语言信息工作平台　1. 语义科技工作系统（李葆嘉教授主持）：基于词汇语义－句法语义的一体化研究思路，开发“人－机交互语义标注工具”，研制“深度语义标注信息库”；研制“幼儿（2～6）日常话语跟踪语料库”，完成幼儿语义系统和话语行为分析研究。2. 语音科技工作系统（顾文涛教授主持）：研制“多语言、多语境、多语用的语音语料库”，基于声学信号分析、感知实验和数学建模，完善语音韵律理论与相关技术应用。3. 方言实验工作系统（刘俐李教授主持）：完成“网络版汉语方言有声语料库”，拟定系统的可操作性语音、词汇、语法实验模型和研究方法，进一步完善新兴交叉学科“实验方言学”。

二、语言知识工作平台　1. 先秦词汇统计与知识检索系统（陈小荷教授主持）：研制“先秦文献语料库”、“专名知识库”、“汉语词汇档案库”等，开发先秦文献自动分词算法、古籍版本异文自动发现算法、同指专名检索软件工具等，完成“先秦汉语词汇统计与知识检索”。2. 中古词汇统计与知识检索系统（董志翘教授主持）：研制“中古文献语料库”、“专名知识库”、“中古汉语词汇档案库”等，开发中古文献自动分词和标注工具等，完成“中古汉语词汇统计与知识检索”。3. 敦煌俗语言文字统计与检索系统（黄征教授主持）：研制“敦煌文献资料库”、“敦煌文献俗词语档案库”，开发相应工具，完成“敦煌文献资料与知识检索”。

三、语言机制工作平台　1. 二语习得的神经机制研究系统（倪

传斌教授主持)：研制“英语受蚀词汇库”等，基于行为学、脑成像和脑电三维度模型，进行中国人英语习得与磨蚀的神经机制研究，完成“基于神经机制的英语个性化学习分析系统”。2. 二语习得的中介机制研究系统（肖奚强教授主持)：研制“留学生汉语口语中介语语料库”，基于中介语理论、对比分析理论、偏误分析理论以及二语习得影响因素等，完成“留学生汉语习得的中介机制研究”。

这一工作平台，既是科技研究平台，也是人才培养平台，即一个现代化的科学研究和人才培养工作体系。

作为本项目的文本成果，《语言科技文库》包括计算语言学研究、语义语法学研究、汉语方言学研究、古代汉语学研究、语言教学与研究、语言新专题研究六个系列。其总体特征为：领域的开拓性、理论的原创性、选题的新颖性、方法的交叉性、考据的精审性、成果的应用性。在研究过程中，除了数据采集分析、资源建设和软件开发，更重要的还是要有新思路、新理论和新材料。陈小荷提出的先秦文献信息处理新方法，从先秦典籍注疏文献中挖掘出用于自动分词和词义消歧的知识，再注入已开发的古汉语分词和词性标注工具中去，所取得的先秦古籍版本异文自动发现、先秦词汇知识自动挖掘等成果均具开拓性。李葆嘉提出的语义语法学理论和话语行为理论，基于研制专用语料库或语义信息库和技术手段，开拓了语义网络建构、深度语义分析和话语行为研究等新的领域。刘俐李建构的实验方言学理论和方法，为方言学向现代科技方法的转型研究提供了新路，并取得了一系列新成果。黄征多年来从事敦煌文献及其俗词语文字研究，古代汉语学研究系列中的敦煌文献校录整理，以及敦煌写本字词考释、以古佚和疑伪经为中心的敦煌佛典词语和俗字研究、两汉声母系统研究等新见迭出。肖奚强基于汉语中介语语料库的二语习得研究，在对外汉语教学研究界已经产生了影响。钱玉莲的汉语介词与相应英语形式比较研究等专著各有亮色。倪传斌依据语言测试和认知实验等数据，从行为学、生理学和语言学三个层面分析影响中国英语学习者外语磨蚀的相关因素。刘宇红基于隐喻的理论探讨，对各类隐喻形式的结构、特性和解读规律进行了多视角的深入探讨。

《语言科技文库》所收论著，由作者在2008年12月申报选题，

2011 年始逐步完稿。系列主编审读了书稿，主要就其学术价值、章节安排、内容关联、行文表述、图表绘制等方面，提出审阅意见。此后，作者们对书稿又进行了修改和润色。《语言科技文库》的作者，大多数是具有博士学位的年轻教师。对于我们这些 20 世纪 80 年代走进语言学研究领域的而言，出版论著可能已不足为道。然而，对于年轻学者而言，其论著的出版既是几年来研究的结晶，也是对其继续探索的促进。换而言之，“211 工程”重点学科建设的目的之一，就是为年轻教师搭建一个可持续发展的科研和教学平台。学科带头人的主要任务之一就是提携后进。

尽管从根本上来说，科学或学术研究是一种个人的探索行为，然而复杂问题的研究，无疑需要群体协作。“学科建设”或团队合作模式，是 20 世纪 90 年代后期出现的一个新概念。这种模式涉及总体规划、多方协调，是需要付出精力和心血的。2008 年，通过投票方式推举我担任该项目总负责时，就意识到自己成了一个“劳动班委”。2009 年，前往安徽大学拜访黄德宽教授时，曾谈到“学科负责人的任务就是规划设计，争取项目经费和提供科研设备设施”，得到黄教授的赞许。2010 年，申报江苏省高校哲学社会科学重点研究基地时，评审专家柳士镇教授提问的“作为一个交叉学科项目，各学科之间的协调是怎么考虑的，有什么做法”，可谓一语中的。作为后学，深知交叉研究之艰、学科整合之难。相关学科之间的整合协调需要借助行政机制，但凭借行政方式并非就能完成。当时的回答是，目前做到的是建成了一个可以合作研究的场所，至于学科之间的进一步沟通合作应有较长过程。有一点很明确，只有通过交叉项目，相应学科才能渗透，合作者才能逐步磨合。我们只是在一步步探索。

“十一五”期间的“211 工程”建设项目即将完成，但是学科建设的任务并没有结束。2010 年，“语言信息科技研究中心”被评审为江苏省高等学校哲学社会科学重点研究基地，为“语言科技”这一交叉领域注入了新的建设活力。重点研究基地建设，除了“跨院系整合、多学科交叉、开放型运行”理念，需要凸显“合作性攻关”。围绕交叉性项目，实施计算语言学、语音科技、神经语言学、语义科技等力量的联合攻关计划。只有通过全面开放以及和与国内

外同行的合作交流，才有望建成具有影响的语言科技研究、人才培养和学术交流基地。

十年前，我（2001）曾写道："语言科技"的内涵是以理论研究为指导，以描写研究为基础，以应用研究为枢纽，促使语言研究向计算机应用、认知科学和现代教育技术领域等延伸，沟通文理工相关学科以实现语言研究过程及其成果的技术化。"语言科技"的外延为语言工程科技、语言教育科技和语言研究科技。其中，"语言研究科技"是将语言研究活动与资源建设、软件开发相结合，其目标是实现语言学自身的科技化。还应包含语言实验、数据处理这些实验语音学、神经语言学研究的科技手段。

虽然语言学家不可能也不必要都转向语言计算或实验研究，尽管描写、考据和内省始终是最基本的方法，但是具有一定的语言科技意识却非常必要。语言学家只有了解有哪些可供利用的资源、软件或仪器，才能提高其研究深度、精度和效率。语言学家也只有了解到信息处理的语言研究需求，才有可能为之提供可资应用或参考的基础成果。"语言科技"是21世纪语言学研究的潮流。

此为出版缘起。是为总序。

2011年8月谨识于南都

作者简介

顾文涛，江苏扬州人。上海交通大学通信与信息系统工学博士，日本东京大学博士后。现任南京师范大学文学院语言科技系特聘教授、博士生导师。曾在美国贝尔实验室研究总部访学半年，曾任日本东京大学 JSPS 外国人特别研究员、香港中文大学副研究员。主要研究方向为实验语音学与语音信息处理，尤其关注语音韵律分析和建模。在 *Phonetica*，*IEEE Transactions on Audio Speech and Language Processing*，*Speech Communication*，*IEICE Transactions on Information and Systems* 等国际权威期刊以及 INTERSPEECH，ICPhS，ICASSP，ISCSLP，Speech Prosody，ASRU，TAL 等重要国际会议上发表论文 40 余篇。TAL 2012 国际会议主席。现主持国家社会科学基金青年项目、国家社会科学基金重大招标项目子课题、江苏省社会科学基金项目、江苏高校哲学社会科学重点研究基地重大项目各 1 项。

内容简介

语音韵律在言语交际中不仅传递了语法、语义等信息，而且传递了副语言信息以及说话人的特征信息。不仅语音韵律是语言学研究的重要课题，而且韵律信息处理也是语音信息工程中的关键技术。本书节选了作者关于语音韵律研究的部分论文，系统考察了时长、字调、声调协同、句调、韵律结构、焦点重音、情感表达、说话人风格等各个层面的韵律特征，特别突出了定量建模的研究方法。研究对象以普通话和粤语为主，涉及多种汉语方言，同时包括跨语言对比及语言接触的研究。

自　序

语音韵律在言语交际中不仅传递了语法、语义等信息，而且传递了副语言信息以及说话人的特征信息。不仅语音韵律是语言学研究的重要课题，而且韵律信息处理也是语音工程中的关键技术。众所周知，汉语是典型的声调语言，声调与语调之间的复杂互动关系，使得汉语语音韵律的研究具有特别重要的意义，同时也增加了额外的难度。

本书节选了本人关于语音韵律研究的部分论文，系统考察了字调、声调协同、时长与节奏、句调、韵律结构、焦点重音、情绪、态度、说话人风格等各个层面的韵律特征，特别突出了定量分析与建模的研究方法。研究对象以普通话和粤语为主，涉及多种汉语方言，同时包括跨语言对比及语言接触的研究。

本书由上、下两篇组成。上篇（Part I）选自本人 1999 年上海交通大学博士论文的英文稿，主题是汉语文语转换系统中音长模型的说话人适应方法。下篇（Part II）选自本人 2004 年之后以第一作者兼通讯作者身份正式发表的部分英文论文，主要围绕语音基频曲线的定量分析与建模，涉及多语言、多层次韵律特征的研究。为统一起见，已发表的中文论文本书不做收录。

需要说明的是，上篇的研究背景是文语转换（TTS）系统的音长模型。事实上，TTS 在过去十多年中已经取得了显著的进展，例如，基于超大规模语料的单元选择与级联的合成方法、基于隐马尔可夫模型（HMM）参数训练的合成方法，都已成为 TTS 的主流。因为上篇写成于 13 年前（美国贝尔实验室的多语言 TTS 系统是当时最前沿的），文中对于 TTS 的个别细节描述，难免落后于最新的技术发展。但是，合成语音的个性化仍然是 TTS 未来的追求目标，而且随着近年的技术发展，受到了越来越多的重视。个性化的、具有特定风格或表现力的韵律特征的重建仍然是技术上的难题。这一问题的挑战性在于，从

实际应用的角度，往往要求在非常有限的训练数据上，快速有效地建立新说话人或者新的表达风格的韵律特征。因此，当年所做的研究在今天看来仍有借鉴意义。

上篇由前七章组成。以音长模型为研究对象，提出了一种将TTS的音长模型映射到新的说话人的方法。其目标是利用尽可能少的训练语料来获取特定说话人的音长特征，建立反映该说话人特点的音长模型。方法是对各种语言通用的。与此同时，系统地考察了普通话的音长变化规律。

第1章简要介绍了文语转换系统的音长模型，在点明研究目标后给出了基本假设，即音长模型可以分解为语言相关的共性特征以及说话人相关的个性特征两部分，对于新的说话人只需重建说话人相关的特征信息。

第2章详细讨论了最优文本选择的方法。对于方差分析模型，文本选择问题就转化为线性参数估计问题。通过将不同音素类的设计矩阵级联，提出了一种基于多模型合并的贪心算法，使训练集的文本规模最小化。

第3章简要叙述了语音数据采集过程，包括语音录制、切分标注、数据处理。针对语音与文本不完全吻合导致的模型可解性破坏的问题，给出了解决方法。

第4章通过对普通话音长数据的详尽的统计分析，建立了普适的乘法音长模型，详细考察了每个音类中各因子对音长的作用关系，与前人文献的结果做对比分析，并特别讨论了各种音长补偿关系以及音节时长模式。

第5章通过对模型参数向量的分析发现，不同说话人之间既保持了参数次序的一致性，又呈现了不同的尺度关系。基于此，提出了一种尺度可伸缩的乘法模型，用一组权值修正参数提取说话人的音长特点。通过实验比较，证明了该模型的有效性。

第6章特别讨论了语速对音长的作用关系，研究发现语速与不同音长因子之间有截然不同的交互作用模式，表明在不同语速下只做简单的线性伸缩不能有效地反映音长的变化规律。

第7章是对前6章内容的概括，简要总结了TTS音长模型的说话人适应方法。

下篇则由第 8 至 19 章组成，主题是语音基频曲线的定量分析与建模，考察了字调、声调协同、句调、韵律结构、焦点重音、情绪与态度表达等各个层面的韵律特征，其中建模的工作是在指令响应模型（即 Fujisaki 模型）的基础上做的。

第 8 与第 9 章在简要介绍了指令响应模型之后，详细讨论了模型指令参数的自动提取方法，主要面向声调语言提出了声调指令的提取方法。其中第 8 章针对普通话，第 9 章则扩展至更一般的情形，并做了进一步改进。

第 10 至 13 章分别讨论了将指令响应模型运用于粤语、吴语（含上海话、苏州话、吴江话）、客家话的研究。其中第 10 章对粤语的讨论是重点，对建模方法做了详尽的阐述；第 11 至 13 章则进一步测试该模型对于不同汉语方言的适用性。

第 14 章运用建模的方法研究了焦点重音与疑问调对粤语基频曲线的作用规律。粤语焦点重音的实现可以描述为短语指令的增加或增强；粤语疑问调则主要体现为句末音节后半声调指令的替换，这也证实了赵元任先生提出的汉语有两种边界调（同时叠加与后续叠加）的思想。

第 15 章定量地分析了粤语的前后声调协同作用以及焦点重音的韵律实现，并做了跨语言对比研究。对声调协同的研究揭示了在四个不同时间域内存在四种不同的协同作用方式；焦点重音的韵律特征分析则与第 14 章建模方法的研究结果吻合，焦点重音的实现方式在普通话和粤语中恰好相反，表明副语言信息的韵律实现与语言的调系结构密切相关。

第 16 章从韵律产生的角度定量地考察汉语口语的韵律结构，借助模型分析方法，由基频曲线推导出短语指令的分布。研究表明，基于模型解析的短语指令和基于听感标注的各级韵律边界之间只存在一定概率的松散对应关系。

第 17 章定量地分析了粤英语码混用时的韵律变化规律，揭示了在两种语言在语流中直接交接时，声调语言（粤语）的字调与重音语言（英语）的词重音之间的耦合变换关系。

第 18 章运用特征分析与建模这两种方法，考察了四种基本情绪下的普通话韵律特征。不同于副语言信息的韵律实现（往往与语言

的调系结构密切相关），反映说话人特征的非语言信息（如情绪）的韵律表现具有跨语言的共通之处。

第19章对情感语音研究中往往混为一谈的“情绪”和“态度”在概念上做了严格的区分，采用角色扮演的诱导式方法，设计并采集了五个态度类的语音语料。通过听辨实验和声学分析探讨了各类态度表达的语音韵律特征。

重读过去的论文，自然就回想起自己十多年来在语音学特别是语音韵律研究领域探索的历程。在导师张煦教授（中国科学院院士）与郑志航教授的鼓励下，1997年9月至1998年3月，由母校上海交通大学派至美国朗讯科技公司贝尔实验室总部访学，跟随Chilin Shih博士、Jan P. H. van Santen博士、Sunil K. Gupta博士等人，踏上了语音学与语音技术研究之旅，更结识了一批蜚声国际语音学界的著名科学家；虽然因为临时更换研究方向（原为数字图像通信）导致博士论文研究不够深入，却也有幸在一个很高的起点上启程。2001至2003年、2004至2006年，两度赴日本东京大学做博士后研究，跟随语音韵律研究的国际权威学者Hiroya Fujisaki教授（日本工程院院士）、Keikichi Hirose教授，真正迈进了语音韵律研究的殿堂。2006年至2008年，再赴香港中文大学，在P. C. Ching教授与Tan Lee副教授的语音团队中继续研究粤语韵律，更有幸在同一实验室获得国际著名语言学家王士元教授（中央研究院院士）的多方指导。一路走来，有幸与这些国际著名学者共事并获得他们持续的支持与鼓励，正是我能取得微薄成绩的基础。

我来到南京师范大学工作以后，2009年12月，主持建成了南京师范大学语音科技实验室；2011年9月，主持召开了由中国语言学会语音学分会主办的全国语音科技实验室学术交流与发展规划研讨会；2011年12月，在《南京师范大学文学院学报》（现为CSSCI源期刊）新开了“语音科学与技术”专栏（刊首）；2012年5月，主持召开了由国际言语通讯协会ISCA、国际语音协会IPA、日中科学技术交流协会JCSTEA以及5个中国国家级学会联袂主办的The Third International Symposium on Tonal Aspects of Languages（TAL 2012，网址为www. TAL2012. org），并在会后选编优秀论文在国际权

威期刊 Phonetica（SCI 源期刊）与 Journal of Chinese Linguistics 各出一期专辑；再加上这本小书，一并算是我对中国语音学与语音技术研究的微不足道却写满诚意的小小推动吧。

当然，现在回头去看以前写的论文，发现有很多肤浅与稚嫩之处，读来令我汗颜不已。虽然很想从头到尾彻底改写一遍，但是受限于丛书的出版时间要求，加之各类琐事缠身，只能基本保持历史原貌不动，仅做了一些格式上的调整和个别字句的润色。因此，一定存在很多疏漏和缺陷，有待今后修正。借此机会，恳请学界同行不吝赐教。

顾文涛

2012 年 7 月于南师大随园

Contents

PART I
SPEAKER ADAPTATION FOR DURTION MODEL IN MANDARIN TEXT-TO-SPEECH SYNTHESIS