国外信息科学经典教材

语 音 合 成

Progress in Speech Synthesis

Jan P. H. van Santen
Richard W. Sproat
(美) Joseph P. Olive 编
Julia Hirschberg

蔡莲红 杨鸿武 吴志勇 等译

机械工业出版社
CHINA MACHINE PRESS

附光盘

国外信息科学经典教材

语 音 合 成

Jan P. H. van Santen
Richard W. Sproat
Joseph P. Olive
Julia Hirschberg
（美） 编

蔡莲红 杨鸿武 吴志勇 等译

机 械 工 业 出 版 社

本书介绍了语音合成技术近年来取得的进展。全书分 8 部分共 46 章，主要内容包括语音信号处理和声源建模、语言学分析、发音器官合成与可视语音、拼接合成与自动切分、自然语音的韵律分析、韵律合成、评价与感知，并简单介绍了两个系统及其应用。

本书可供计算机工程、智能信息处理、电子工程等专业的研究人员或工程技术人员参考使用，也可作为相关专业研究生和本科生的教材。

本书版权登记号：图字 01-2003-3155

图书在版编目（CIP）数据

语音合成/（美）简·凡桑塔（Jan P. H. van Santen）等编；蔡莲红等译. —北京：机械工业出版社，2005.3
书名原文：Progress in Speech Synthesis
国外信息科学经典教材
ISBN 7-111-15529-7

Ⅰ. 语… Ⅱ. ①简…②蔡… Ⅲ. 语音合成－研究－进展 Ⅳ. H017

中国版本图书馆 CIP 数据核字（2004）第 111902 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：李利健 版式设计：张世琴 责任校对：陈延翔
封面设计：刘吉维 责任印制：石 冉
三河市宏达印刷有限公司印刷·新华书店北京发行所发行
2005 年 3 月第 1 版·第 1 次印刷
787mm×1092mm 1/16·26.5 印张·657 千字
0001—3000 册
定价：55.00 元（含 1CD）

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
本社购书热线电话（010）68326294
封面无防伪标均为盗版

译 者 序

语言处理是一个古老而又崭新的课题，它在不断地进步。从1939年纽约世界博览会展出的早期“声码器”，到Haskins实验室从Frank Cooper制造模式回放机开始所作的许多贡献，到Fant的名著，再到Klatt所作的巨大贡献。从这些研究以及其他一些重要贡献来看，语音合成与语音生成研究是密切相关的。如果我们想让一个机器像人一样讲话，我们最好先了解人类语音产生的机理。同样，为了解人类是如何产生语音的，我们可以这样测试我们的模型：让模型合成语音（或语音的某些特性），再将模型的输出与我们在自然语音中的测量结果进行比较。这就是Mary Beckman呈现给我们的思想。

本书是在第二届ESCA/IEEE/AAAI文语转换研讨会的基础上，由技术委员会成员来遴选候选文章，并邀请了一致认可的文章的作者来编写的。虽然本书介绍的是几年前的技术，但它涉及到的理论、方法、技术仍是需要我们认真学习和效仿的。

近些年，语音合成的研究在国内也成了一个关注的热点，其成果已走出实验室，在信息领域获得了应用。语音应用的需求也促进了技术的发展。然而相应的书籍还较少。

在1997年初，Bell实验室的朋友给我们带来了这本书的英文版。它一直是清华大学计算机系人机语音交互组研究生的必读资料。不敢说我们都读懂了，理解了，但本书确实给我们上了一课。书中如此丰富的内容，如此多等待“进展”的课题，激发了研究生们的兴趣。

本书的翻译出版又给我们一次深入学习的机会，我们很高兴承担此任务。中译本的出版可以让国内更多的同行和学子了解语音合成，通过此书，让我们结交更多的朋友。

在本书的翻译出版过程中，清华大学计算机系人机语音交互组的研究生杨鸿武、吴志勇、蒋丹宁、倪昕、王志明、蔡锐、崔丹丹、郑敏、马磊参加了翻译工作，蔡莲红、杨鸿武、吴志勇审校了全书。由于译者的水平所限，书中难免出现不够准确或错误的地方，敬请读者指出。

译 者

出 版 说 明

在人类迈入信息时代的今天，信息技术的应用无处不在，我国对信息技术的重视和鼓励也达到了空前的程度。信息技术的发展速度很快，可谓日新月异，尤其在一些发达的国家更是如此。我国信息技术起步较晚，但发展速度惊人，这正是改革开放的具体体现。在我国大力发展信息产业的今天，为了能融入国际的潮流和掌握最新的技术，从国外引进先进的知识和技术就显得格外重要。为此，我们决定引进一系列国外信息技术领域有代表性的优秀教材，将它们献给我们的学子、教师和IT业的有志之士，藉此为我国的信息产业贡献一份微薄之力。

随着我国加入WTO，国际间的竞争将越来越激烈，而国际间的竞争实际上就是人才的竞争、教育的竞争。为了加快培养具有国际竞争力的高水平技术人才，加快我国教育改革的步伐，使我国的高等教育尽快与国际接轨，这就需要引进先进的教学思想和教学方法，而引进国外优秀的教材无疑是一种很好的途径。同时，引进国外的优秀教材也有利于提高我国自编教材的水平，让我们的教育工作者从中得到启发。

我们这套丛书遵循“新、优、特”的原则，做到知识新、质量优和内容有特点。这套丛书涵盖了计算机、通信、电子技术等领域，每一本书都是精心挑选，在某个领域或学科内具有很强的代表性和很高的价值，很多在国外也被作为大学的教科书，由国际知名的出版公司出版。在引进过程中，我们邀请有关专家对书稿的整体水平进行了评定；在翻译过程中，我们聘请国内相关领域内的有很高学术水平的专家和学者，以保证书籍的水平和质量，做到对读者负责，为读者着想。

相信这套丛书的出版对正在苦读和即将面临挑战的学子们会有很大的帮助和提高，也能让我们的教学工作者从中得到启发，同时对从事IT行业的工程技术和研究人员而言也是很难得的工具书。

机械工业出版社

前　言

从文本到语音的转换、合成，包含了从输入文本到语音信号的各种计算。要满足这些计算需求，文语转换系统必须具备从对话结构的抽象语言学分析到语音编码的众多功能组件。

这一事实隐含了以下几个含义：首先，文语转换具有跨越多学科的性质。这一点可以从本书众多作者的背景得到证明，他们分别来自工程领域、语言学的多个领域、计算机科学、计算心理学和声学领域；第二，由于面对问题的复杂度差异很大，这些领域的研究进展是不平衡的。比如，为复杂的、多段的输入文本赋予无缝衔接的音高重音极为困难，而在其他音素都计算无误的情况下，生成音段时长并非十分困难；第三，总结 TTS 所有相关领域的研究惟一可行的办法是写一本多著者的书，因为任何个人，甚至研究小组都不可能具有足够广博的知识。

当然，这本书最重要的目标是邀请各相关领域的关键专家对各个领域作全面的评述和回顾。这将给读者展现一幅完整描绘 TTS 技术的图画：这一领域中所面临的挑战和解决思路是什么，研究者们正在向哪些方向前进。

本书的第二个重要目标是使读者能依据这些工作对最终结果做出判断——科技尖端的发展是令人鼓舞的，但是它是否能够真正产生高质量的合成语音？本书试图用两个方式来回答这个问题。第一，只要可能，我们都要求作者在他们所写的章节中收入主观评测结果。此外，书中包含有专门论述感知和评测问题的章节。第二，本书附带了一张包含有几个合成器合成样例的光盘。其中的成功与失败样例都十分有趣——尤其是后者，因为包含能够显示系统主要缺陷的样例通常是不太可能的。我们感谢 Christian Benoit 提出的这个建议。

在此，简短介绍一下本书的历史。1990 年，第一届 ESCA 文语转换研讨会在法国的 Autrans 召开。这次研讨会的组织者 Gerard Bailly 和 Christian Benoit 感到出版一本摘录长篇会议论文的书籍十分必要，因此《说话的机器》一书得以出版。在 1994 年，本书的编者组织了第二届 ESCA/IEEE/AAAI 文语转换研讨会，同样，我们感到有必要出版一本书籍来进一步适时、完整地展示当前的工作。为了保证本书的每一章节都能达到可能的最高质量，我们邀请了此次研讨会的技术委员会成员来遴选候选文章。本书编者也加入了他们的选择行列。而后我们邀请了那些得到一致认可的文章的作者来完成此书。

作为本书的编者，我们要感谢那些对本书的出版做出贡献的人们：技术委员会中参与遴选文章的成员、14 名匿名审稿人、Bernd Moebius 的精彩图示、Alice Greenwood 在编辑上的帮助、Mike Tanenblatt 和 Juergen Schroeter 处理了音频和视频文件、David Yarowsky 制作了

索引、Thomas von Foerster 和 Kenneth Dreyhaupt 在 Springer - Verlag 出版公司的高效工作、Cathy Hopkins 在管理上的帮助，以及 Bell 实验室对这一工作的支持和鼓励。

Jan P. H. van Santen
Richard W. Sproat
Joseph P. Olive
Julia Hirschberg
Murray Hill, New Jersey
1995 年 10 月

目　录

第二部分 语言学分析

第三部分 发音器官合成与可视语音

第四部分 拼接式语音合成与自动切分

第五部分 自然语音的韵律分析

第六部分 韵律合成