人机自然交互

Human-Computer Nature Interaction

张有为　等著

国防工业出版社

http://www.ndip.cn

人机自然交互

Human-Computer Nature Interaction

张有为　等著

国防工业出版社

·北京·

图书在版编目(CIP)数据

人机自然交互/张有为等著. —北京:国防工业出版社,2004.9

ISBN 7-118-03544-0

Ⅰ. 人… Ⅱ. 张… Ⅲ. 人—机系统—研究 Ⅳ. TB18

中国版本图书馆 CIP 数据核字(2004)第 065391 号

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号)

(邮政编码 100044)

北京奥隆印刷厂印刷

新华书店经售

*

开本 710×960 1/16 印张 17¼ 314 千字

2004 年 9 月第 1 版 2004 年 9 月北京第 1 次印刷

印数:1—3000 册 定价:58.00 元

致 读 者

本书由国防科技图书出版基金资助出版。

国防科技图书出版工作是国防科技事业的一个重要方面。优秀的国防科技图书既是国防科技成果的一部分，又是国防科技水平的重要标志。为了促进国防科技和武器装备建设事业的发展，加强社会主义物质文明和精神文明建设，培养优秀科技人才，确保国防科技优秀图书的出版，原国防科工委于 1988 年初决定每年拨出专款，设立国防科技图书出版基金，成立评审委员会，扶持、审定出版国防科技优秀图书。

国防科技图书出版基金资助的对象是：

1. 在国防科学技术领域中，学术水平高，内容有创见，在学科上居领先地位的基础科学理论图书；在工程技术理论方面有突破的应用科学专著。

2. 学术思想新颖，内容具体、实用，对国防科技和武器装备发展具有较大推动作用的专著；密切结合国防现代化和武器装备现代化需要的高新技术内容的专著。

3. 有重要发展前景和有重大开拓使用价值，密切结合国防现代化和武器装备现代化需要的新工艺、新材料内容的专著。

4. 填补目前我国科技领域空白并具有军事应用前景的薄弱学科和边缘学科的科技图书。

国防科技图书出版基金评审委员会在总装备部的领导下开展工作，负责掌握出版基金的使用方向，评审受理的图书选题，决定资助的图书选题和资助金额，以及决定中断或取消资助等。经评审给予资助的图书，由总装备部国防工业出版社列选出版。

国防科技事业已经取得了举世瞩目的成就。国防科技图书承担着记载和弘扬这些成就，积累和传播科技知识的使命。在改革开放的新形势下，原国防科工委率先设立出版基金，扶持出版科技图书，这是一项具有深远意义的创举。此举势必促

使国防科技图书的出版随着国防科技事业的发展更加兴旺。

设立出版基金是一件新生事物，是对出版工作的一项改革。因而，评审工作需要不断地摸索、认真地总结和及时地改进，这样，才能使有限的基金发挥出巨大的效能。评审工作更需要国防科技和武器装备建设战线广大科技工作者、专家、教授，以及社会各界朋友的热情支持。

让我们携起手来，为祖国昌盛、科技腾飞、出版繁荣而共同奋斗！

国防科技图书出版基金
评审委员会

国防科技图书出版基金
第四届评审委员会组成人员

名誉主任委员	陈达植				
顾　　问	黄　宁				
主 任 委 员	刘成海				
副主任委员	王　峰	张涵信	张又栋		
秘　书　长	张又栋				
副 秘 书 长	彭华良	蔡　镭			
委　　员	于景元	王小谟	甘茂治	冯允成	
（按姓名笔画排序）	刘世参	杨星豪	李德毅	吴有生	
	何新贵	佟玉民	宋家树	张立同	
	张鸿元	陈火旺	侯正明	常显奇	
	崔尔杰	韩祖南	舒长胜		

前　言

自然科学中有2个重要的概念，就是能量与信息。能量这个概念在科学技术中所起的支配作用是人们所熟知的，而与能量的存在具有同样悠久历史的信息的概念，在科学上取得应有的地位，却只有半个世纪。400年前、200年前、50年前分别出现了制造业、动力业、信息业。20世纪40年代末期，信息科学的开拓者美国学者香农、维纳和苏联学者柯罗莫格洛夫所发表的经典著作标志着信息科学发展的起点。

信息科学的一个重要研究方向是人机交互，近几年来已趋于活跃。人与人的交流是很自然的，这是一个人人交互过程，是人的思想通过语言和情感表达与理解的高级生命双向交互过程。人是作为最高级生命而存在的，其他生命间的交互即使存在，也是处于低级形态，主要表现在同类同种同属的动物之间。生命之间的交互是通过听觉、视觉、触觉、嗅觉和躯体姿态来完成的。对于最高级生命——人类来说，语言是思想表达的主要方式，其信息是通过听觉与视觉通道来传递的。所谓人机自然交互是指人与机器的自然交互或人与计算机的自然交互。自然交互区别于人工的交互，是通过赋予机器的听觉和视觉自然地完成的。智能化的实现主要通过计算机，或者说智能机器的智能是由它包含的计算机而赋予的。

人类的进步，其重要特征是制造工具，智能化的工具就是智能化的机器。人们渴望机器能受人支配，听从人的指使，只要人动嘴要它做什么，它都能够理解，并按要求去做。同时机器还能把对于人的表述的理解反馈回来，回答人提出的问题，达到交互的目的。人类对机器的要求还不止于此，甚至要使人与机器的对话达到人与人之间交流那种自然的水平。这是除人与人之间以外，其他非高级生命与人类交流无法或难以达到的。这种想法所实现的是生命与非生命之间的对话，涉及非生命对生命思想的理解。虽然非生命机器的思维都是人这种高级生命注入给它，并经过训练和它的自学习而完成的，或许相当于人类接受教育的过程。

自然信源是极其丰富的，遂以人类的进步、生态繁茂和经济繁荣的方式向前发展。然而，人造信息体制与自然信息体制是不同的，不但自然信源尚待发掘，二者之间的翻译与接口界面也尚待开发。人机自然交互的研究是发展人造信息与自然信息的接口，本质上是改造人造信息渐与自然信息体制相兼容。人机自然交互作为信息科学与技术的一个先进研究领域，必然涉及哲学、人类学、生物学、数学、物

理学、电子科学、计算机科学、人工智能、信息论、控制论、认知科学、心理学以及伦理学，是它们互相渗透、互相结合的产物。

人机自然交互正如人与人交互一样，首先应能识别交互对象，即知道在与谁交互对话；其次要理解对话的内容；第 3 要理解对方的情态，即是兴奋、喜悦、无动于衷、烦恼、愤怒等情感；第 4 要能对说话人进行定位与跟踪，好像人与人交流过程中总能注视对方，有多个人说话时总能注视正在说话者或区分说话人。在人机交互中存在一个双向过程，人对于机器的理解相对来说是易于完成的，而难点在于机器对人的理解。所以对于人机自然交互的研究，即研究生命与非生命对话过程中，将主要研究非生命对于生命形象、语言和情感的理解。

人类的语言本质上是双模态的，音频语言与说话人发出的声波有关，视频语言与说话人嘴唇、舌头和面部肌肉的运动有关。音频与视频所对应的正是人的听觉与视觉。人机自然交互的研究就是要赋于机器听觉与视觉的功能，能识别人并能理解人的说话内容和感情，进而进行推理与思维，执行人对它的要求或与人进行交流。

本书共有 6 章，第 1 章是导论；第 2 章、第 3 章将分别从视觉和听觉角度研究对说话人的识别；第 3 章、第 4 章是从听觉和视觉角度研究对说话内容的识别与理解；第 5 章研究听觉和视觉双模态的融合问题；第 6 章介绍对于人机交互研究的最基础的条件：听觉-视觉双模态语言识别数据库。关于情态识别问题、对说话人定位与跟踪问题以及多模态人机交互网络环境问题，本书除在导论中给予介绍外，不再另立章节论述。这是由于情态识别问题在理论和技术上都还不很成熟，还没有达到著书的可能；而对说话人定位与跟踪问题就技术实现上来说有很多书籍都可参用。关于多模态人机交互网络环境问题，虽然我们进行了大量研究，但是由于本书篇幅所限，不再列为一章研究和讨论。本书用大量的篇幅着重介绍了作者在实验室中进行的研究结果和验证。然而人机自然交互是一个广阔的研究领域，由于本书主要是就作者所得的研究结果进行论述，只能说是研究了人机自然交互的若干问题。为了弥补这种不足，本书采用极简洁的文字介绍了国内外的主要理论和方法，同时给出了详尽的参考文献，便于需要深入了解的读者查阅。

信息科学是一个年轻的自然科学分支，而人机自然交互是更为年轻的一个研究方向，它被人们所重视还是近 10 年的事，从严格的科学意义上来说，它离完善还有相当的路程，当前仍然处于初始阶段。我们虽然在本书中从听觉-视觉的角度来研究问题、建立模型，但是我们还没有做到建立一个从识别说话人、跟踪说话人、识别理解说话内容和情感的统一模型，尽管我们相信它是存在的。科学的终极在于提供一个简洁的理论去描述整个人机自然交互过程，看来这是非常困难的。世界一切事物存在一个“不变”的规律，就是“变”，科学研究过程亦是如此。对于科学工作者来说，变就是进步。人类已经可以观测到 100 亿光年以外的宇宙，也可以看到

亿分之一纳米以下的微观结构，不断向认知的极限挺进，人们也一定可以认识人机自然交互的规律与本质。

人机自然交互的实现将会给机器带来革命性的变化，当它应用于电子产品、通信设施、机械设备、交通工具、人工智能、智能仪器、多媒体、情报采集、身份认证、安全防范以及武器现代化时，将会对科学技术、生产领域、国家安全、社会的工作方式和生活方式等方面产生深远的影响。

本书由张有为主撰，甘俊英、何强、蒙山、应自炉和张歆奕对本书中涉及的许多理论和方法进行了实验验证和开拓，并进行了相应章节的撰写工作(他们撰写的章节是：第2章和第4章甘俊英，第3章何强，3.9.3节张歆奕，第5章蒙山，第6章应自炉。他们的排名是依姓氏的汉语拼音第1个字母为序排列的，他们对本书做出了同样重要的努力，应视为并列的第2作者)。应自炉进行了本书的计算机编排工作。

本书在写作过程中参考和引用了国内外许多作者的有关论述和结果，从中得到了启发和教益，也得到了学术界朋友的支持与鼓励。中国科学院院士清华大学李衍达教授、中国科学院院士中国科学院声学研究所侯朝焕教授，在2001年底本书初稿形成后，进行了评审，给予了鼓励并推荐本书出版。中国电子学会信号处理分会主任委员北京交通大学博士生导师袁保宗教授、中国电子学会信号处理分会副主任委员北京航空航天大学博士生导师毛士艺教授、中国科学院声学研究所博士生导师杜利民教授对于我们在本书涉及人机自然交互技术和信号与信息处理研究方法等方面曾多次给予指导。在我们进行人机自然交互领域的研究中，得到了2项“863”计划子课题的支持，和3项广东省自然科学基金(No. 960631，No. 000872，No. 032356)的支持，同时也得到了广东省教育厅和五邑大学的大力支持。在此一并表示衷心的谢意。感谢本书全体作者的母校——北京航空航天大学的培育之恩。感谢国防科技图书出版基金评审委员会对于本书出版的有力支持。

在本书的一些章节我们使用了ORL，CAVSR 1.0，CAVBSR-WUIIS(1.0)3个数据库的人脸资料，用以表示对人脸图像的实验处理结果，为本书增加了所述内容的说服力。在进行人机自然交互的研究过程中，我们参阅了众多的论文和著作，有些列入了参考文献，有些由于篇幅所限尚未列入，我们从中获得了许多启发和教益，在此谨表谢忱。研究生赵向阳、何元烈、王东、戴伟、张莉、汪晓东、黄生、傅筠、周延蕾等都进行了许多有意义的研究和开发工作，他们的研究成果丰富了人机自然交互的某些研究。

由于作者水平所限，本书难免有错误与不当之处，欢迎批评指正。

张有为

目　录

Contents