

O'REILLY®

Broadview®
www.broadview.com.cn



语音用户界面 设计

对话式体验设计原则

Designing Voice User Interfaces:
Principles of Conversational Experiences

[美] Cathy Pearl 著
王一行 译



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

语音用户界面设计

对话式体验设计原则

Designing Voice User Interfaces:
Principles of Conversational Experiences

【美】Cathy Pearl 著

王一行 译

电子工业出版社

Publishing House of Electronics Industry

北京•BEIJING

内 容 简 介

本书是一本由拥有17年从业经验的资深语音用户界面（VUI）系统设计专家Cathy Pearl撰写的语音设计经典，从VUI系统简史开始讲述，包括了语音用户界面设计的基本原则，VUI系统设计中必不可少的人物模型设计，还包括一些语音用户界面设计的进阶知识，如语音识别技术，自然语言理解、情感分析、数据收集和文本语音转换策略等，另外还介绍了语音用户界面用户测试的知识，以及VUI在投入实际使用后需要注意的地方，特别提到了包括家庭助理设备和可穿戴设备等如何设计VUI系统。

本书是由资深设计师翻译的，语言通俗易懂，精准到位，可帮助所有对VUI系统感兴趣的设计师从零开始设计出一个完整的VUI系统，也适用于企业管理者和商务人士了解语音用户界面的设计。

2017 by Cathy Pearl

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Publishing House of Electronics Industry, 2017. Authorized translation of the English edition, 2017 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

本书简体中文版专有出版权由O'Reilly Media, Inc. 授予电子工业出版社。未经许可，不得以任何方式复制或抄袭本书的任何部分。专有出版权受法律保护。

版权贸易合同登记号 图字：01-2017-5970

图书在版编目（CIP）数据

语音用户界面设计：对话式体验设计原则 / （美）凯瑟·彼尔（Cathy Pearl）著；王一行译. —北京：电子工业出版社，2018.1

书名原文：Designing Voice User Interfaces: Principles of Conversational Experiences
ISBN 978-7-121-32555-7

I. ①语… II. ①凯… ②王… III. ①语音识别—人机界面—程序设计 IV. ①TP391.1

中国版本图书馆CIP数据核字(2017)第206225号

责任编辑：孙学瑛

封面设计：Randy Corner 张 健

印 刷：三河市鑫金马印装有限公司

装 订：三河市鑫金马印装有限公司

出版发行：电子工业出版社

北京市海淀区万寿路173信箱 邮编：100036

开 本：720×1000 1/16 印张：22.25 字数：210千字

版 次：2018年1月第1版

印 次：2018年1月第1次印刷

定 价：79.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至zlts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819，faq@phei.com.cn。

O'Reilly Media, Inc.介绍

O'Reilly Media 通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自 1978 年开始，O'Reilly 一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly 的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly 为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了 Make 杂志，从而成为 DIY 革命的主要先锋；一如既往地通过多种形式缔结信息与人的纽带。O'Reilly 的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly 现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版、在线服务或者面授课程，每一项 O'Reilly 的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

O'Reilly Radar 博客有口皆碑。

——Wired

O'Reilly 凭借一系列非凡想法（真希望当初我也想到了）建立了数百万美元的业务。

——Business 2.0

O'Reilly Conference 是聚集关键思想领袖的绝对典范。

——CRN

一本 O'Reilly 的书就代表一个有用、有前途、需要学习的主题。

——Irish Times

Tim (Tim O'Reilly, 编著注) 是位特立独行的商人, 他不光放眼于最长远、最广阔的视野, 并且切实地按照 Yogi Berra 的建议去做了: “如果你在路上遇到岔路口, 走小路 (岔路)。”回顾过去, Tim 似乎每一次都选择了小路, 而且有几次都是一闪即逝的机会, 尽管大路也不错。

——Linux Journal

业界评论

古往今来，“声音”一直是人际交互的核心。而现在，“声音”焕然一新，成为了人机交互的核心。Pearl 完美地将她 17 年的经验，浓缩成一本宝书。她用她宝贵的经验，清晰的思维，以及深刻的洞察能力，为我们阐述了她对语音设计的核心观点，也为我们打开了一扇通往全新交互形式的大门。

——MARK STEPHEN MEADOWS

作家，艺术家，BOTANIC 公司总裁

这本书是一本学习语音用户界面设计基础的优秀教程。这本书将帮助越来越多的设计师去学习如何设计出具有好用的语音交互体验的界面。

——CHRIS MAURY

CONVERSANT 实验室 创始人

Cathy Pearl 基于其丰富的实践经验，为我们带来了这本实用且全面的 VUI 设计书。这本书传授了她多年的实战经验。它将帮助你更好地着手创建你自己的语音界面。

——REBECCA NOWLIN

GREEN NUANCE 通信公司，商业咨询师

Pearl 毫无保留地与我们分享了她毕生对可以倾听、说话的机器的激情，并引领我们进入了 VUI 设计的新纪元。其覆盖面之广，实操性之强令人印象深刻。由于语音设计具有特别的挑战性和意义，即使行业“业内人士”也会不得要领。而这本书对业内人士，以及在这个领域内的商业决策者和开发人员都将具有一定价值。随着多模态应用成为一个固定的文化现象，聊天机器人逐渐兴起，虚拟助理也迎来了它的复兴。（还记得 20 世纪 90 年代的 Wilefire 和 General Magic's Portico 吗？）而此时这本书的出版真是太及时了！

——JAMES GIANGOLA

对话与语音设计方向创意主管，Google

Pearl 的这本书对于如何进行有效的 VUI 设计来说是一次重要革新。这本书具有丰富的专家实操建议，并且引用了大量前沿技术的案例，值得每一位 VUI 设计师拥有。

——JENNIFER BALOGH 博士

《语音人机界面设计》(Voice User Interface Design) 合著者

推荐序一

我在用半语音、半打字的模式写下这段文字时，联想到语音用户界面的发展与互联网技术的发展非常类似。

在互联网时代来临之前，用户界面是以 Windows 为主的 GUI，当用户需要工作、娱乐和游戏的时候，就购买相应的软件，安装后才可以使⽤。当时大部分的设计和⽅法都集中在独立程序上。大概在 1998 年，拨号上网时代结束，Web 浏览器在功能上的完善让大家的眼光开始关注到互联网上，那时 Amazon 刚刚有了百万用户，WebVan 和 Pets.com 正在蓬勃发展（这两款应用目前均已消失）。各种产品开始转向互联网，因为 Web 浏览器可以穿透各个程序，方便地把各种各样的内容和功能都呈现在前端页面上。之后，互联网给整个 IT 行业带来了深刻的变化，独立程序渐渐消失，取而代之的是在浏览器界面中的应用。

到了移动互联网的时代，一轮新的程序开发又开始了。App 像雨后春笋一样涌现出来，满足用户在娱乐、游戏、工作等各个方面

的需求，由于移动屏幕尺寸和网络速度的限制，单单使用一个简单的移动浏览器不能解决所有的问题。

语音和当年的浏览器一样，是一个可以穿透所有程序、解决用户所有问题的方法。当需要叫车服务的时候，我并不在乎是哪个软件在叫车，最好的情况是语音发出后，通过各个应用的比较，找出一个最佳的路线，并且最符合我使用习惯的叫车方案。

语音是人类古老和自然的交互方式，在人工智能发展到可以部分理解语音的时候，语音时代开启了。语音除了在娱乐、游戏、工作上有广泛应用，还会广泛应用在 IoT 上。越来越多的设计师的工作会涉及语音界面设计。这本书的作者 Pearl 把自己 17 年的工作经验总结出来，介绍了语音界面的设计原则、语音识别技术、自然语言理解、语音用户研究方法，以及语音在设备上的应用。理论和实践相结合，非常实用！

这本书是天猫精灵设计团队的参考书之一，在使用过程中发现这本书非常有用，因此我们的设计师一行花时间把整本书翻译成中文分享给大家。

Paul Fu

Sr . Director, User Experience, Alibaba Group

推荐序二

一行嘱我为他翻译的书写序，还没有看过是什么，我便欣然同意了。

一行曾是我的学生。学生时代的一行就成熟得不像他的年龄。在初期时我对他说，“小心技术过于成熟而落入俗套。”事实证明我是错的。一行最大的特点是冷静并特立独行。每次问他问题，他先是沉默，想好后可以很缜密地回答，往往挑不出什么漏洞。基于这样的经验，之后的合作我便选择完全地相信他。所以我想一行精心挑选翻译的书籍，一定不会差。

一行陆续给我寄了原稿和翻译稿，我读后有一些感想：

交互方式正从单纯的视觉、硬件交互转化为多感官综合的交互。每一次技术的变革都带来交互上的巨大挑战。“与机器像和人一样对话”，VUI不是一次升级，而是一个完全不同的新世界。这个世界正在来临，但又还有很多未知；这个全新的体验正呼之欲出，等待人们去定义。

VUI感性层面的需求比视觉交互和硬件交互更为复杂。除了合

理的逻辑之外，场景、对话的上下文关联、角色的设定、语境的感知等都对用户体验起到了重要的作用。VUI 像是给机器赋予一个新的性格。在这个领域也将产生不同的人才需求。最近遇见 Dyson 的设计师，他称自己为工程师，并解释他自己也不知道自己是什么样的定义，只知道伴随了产品整个的研发流程。在 VUI 领域，同时需要被打破的职业界限，需要更多元的职业类型介入去共同定义。他们可能是设计师，是开发者，是语言学家，是心理学家，是剧作家，等等。

最近我买了 Amazon Echo，女儿每天早上向 Echo 问好。Echo 没有屏幕，环转的灯光非常友好地展示了拟人化的倾听。我意识到 VUI 已经来了，它很聪明，但是又还很鲁莽；它让人兴奋，但又时时给人带来与机器对话的挫败感受。哪怕产品很智能，一个不好的 VUI 就像是一个聪明人无法倾听和表达一样无奈。在这个技术与设计并行发展的进程中，VUI 还处在一个很早期的状态，会有很多令人困扰的问题：怎样的设计流程可以支持这个新的领域？VUI 和屏幕交互是什么样的关系？什么样的场景最为适合？机器需要有什么样的性格？如何管理用户的期望值？如何弥补对话中带来的挫败体验？交互过程中的情感感受将会怎样？……

VUI 领域还没有太多的经验可以参考，设计教育界也还没有做好足够的应对，而这是一本来得如此及时的、难得的、实用的工具书。一行告诉我，书的内容只是一个开始，希望读者通过此书能开启一个全新的思考。

刘力丹

designaffairs 合伙人、小猴科技创始人兼 CEO

推荐序三

从软件工程转到工业设计专业，现在又在从事交互设计的工作，王一行一直是一个在自己喜欢的方向上执着钻研、追求完美的学生和设计师。当以往的学霸和红点得主来找我为他的译著写序时，我很是欣喜。非常钦佩一行哥（同学中的昵称）在阿里繁忙的工作之余还能挤出时间来学习新知识，并有心布道语音交互界面设计这一交互设计的新成员。

作为一种自然人机交互方式，语音交互由来以久，近年来更是被运用到了越来越多的产品之中，如从手机到智能产品、智能家居，从机器人到车载 HMI。然而对语音交互的设计方法、流程，以及对其特有的设计问题的研究却非常匮乏。15 年前在 IBM T.J. Watson 参与基于对话的语音交互系统的研究与应用实践，到现在进行多模态车载 HMI 的设计中，我大都是把语音交互作为交互系统中的一个重要元素加以运用。本书将语音交互作为一个与视觉交互界面（GUI）并列的界面类型，对其设计进行系统地分析、

研究与实践指导，视角新颖，为对语音交互设计的深入研究打开了一片天地。相信一行精准到位的翻译可以将大家完美地带入这一交互设计的新领地。

孙效华

博士、教授、博士研究生导师

同济大学设计创意学院副院长，交互设计方向负责人

数字创意中心负责人

推荐序四

当很多人正在热议 AI 究竟是会帮助人类还是毁灭人类的时候，人工智能的浪潮正在悄然颠覆着我们生活的一点一滴。而基于 CUI（Conversational User Interface，对话式用户界面）的智能产品，就好像是冲在 AI 浪潮最前沿的浪花，将隐藏其身下的算法和数据幻化成最能“直接拍打在用户皮肤上”的体验。作为一个新生的领域，VUI（Voice User Interface，语音用户交互界面）对于用户体验设计提出了更多语言学、情感塑造、逻辑搭建等方面的新要求，于是这样一本专业、全面的 VUI 设计指南的出现，就更显得弥足珍贵。

我在第一时间就和我的团队一起精读了英文原版，受益匪浅，相信中文译本的出版一定会帮助更多的共同前行的伙伴，受邀作序，备感荣幸。

就在我敲下这些文字的时候，大多数行业从业者都激进地把人工智能在人类历史上的重要性堪比、甚至远超过工业革命，很多人

正担忧人工智能即将陆续取代一些职业并改造我们的社会化分工。我认为有一点是毫无疑问的，这是一个不可避免的智能时代，作为新生产工具的 AI，除了在限定技术边界的垂直行业应用领域，各行各业都有机会被 AI 力收割“low hanging fruits”来迅速提升各类生产效率，将对人类文明和产业变革产生重大影响。

回顾过去，机械设备的出现为人类带来了提升效能的工具，互联网的出现，将效能再次提升，而伴随着 AI 的出现，效能将进一步大规模提升。不可避免地说，在智能时代，大量重复性工作的人力工种肯定不复存在，还有很多具有结构化知识的经验性工作也会被机器智能所取代，甚至未来是创造性工作，但是更多的是通过为人赋能的方式，减轻社会劳动者和服务业的压力，将会重新进行一轮人类社会化分工。

在互联网时代，“连接”是核心关键词，如果 C 不能满足 A 和 B 的完整需求，那互联网有机会让 A 通过连接，跨过 B 和 C 直接找到 D，但如果有一些场景和行业内 B/C/D 是稀缺性资源，那么这个情况在互联网时代就很难得到解决，AI 的出现将有机会扭转这种供需关系。举例来说，以前用户看病花在预约和挂号的时间基本占到 95%，真正用于医患有效沟通和问诊时间仅为 5%。人工智能将代替以往所谓“互联网医疗”盲目且低效的连接，如 VUI/CUI 初诊问答机器人将患者进行分流，缩短医师护士等在预约和问询过程低效的重复时间，改善供需方的不对等关系。

大家都知道过去 60 年里人工智能经历了两次潮起潮落。现在我把这次 AI 的兴起，分为新 AI 时代的头十年和后十年。前十年是 AI

解决行业问题，比如前面讲的医疗、教育、信贷和金融风控等，通过大数据和机器学习迅速提升效率和准确率；新 AI 时代的后十年，机器智能将全面升级消费产品和用户生活场景，人工智能必将革新交互方式，进入计算无处不在的时代。下一代人机交互革命就在眼前，这将是一个激动人心的时代，就好比百年前蒸汽机、汽车、飞机等伟大产品的发明一样，如今将会出现大量被重新定义的智能产品，无人驾驶车、智能家居、家用机器人、下一个时代的个人助理等。当马车被汽车替代的时候，人们是措手不及的，同样的，AI 时代的创新家们已经蓄势待发准备重新发明和改造今天陈旧的生活方式了。

我觉得赋能是人工智能对人类最重要的事情，因为这是人类进化需求的本质。人都渴望被赋能，这是天性。人发明蒸汽机、汽车、飞机、手机、计算机，从运动能力和生产效率等方面突破自身局限。大部分人沉迷于游戏，是因为在游戏中感受到自己被赋予了更大的权利和能力。这些都是顺应人性的事情，我觉得 VUI 最有价值的场景应用就是在人机交互过程中产生的便利性赋能。

无论是从云端到边缘在下一代硬件架构的变化，还是深度学习算法在图像识别和语音识别的单项上逐渐超过人类水平，AI 的技术成熟度还在不断提升和进化过程，但是今天语音对话交互技术的发展尚还没有成熟到可以替代鼠标键盘和触摸屏。不过现阶段已经可以应用在一些特定场景下，比如用户双手不便下达快速指令、智能电视或者智能音箱上的跳步检索某一首歌或者某一个电影关键词，又或是实时机器翻译。但可以肯定的是，未来无处不在的

AI 技术，必将带来交互的革命，颠覆的机会，也将属于那些极少数懂得技术边界、极具场景化思维、勇于革新的创新家们。

2016 年，我和志同道合的小伙伴们一起创立了“物灵”。在工具化大工业产品思维主导的今天，我们希望去开辟另外一条路径，去创造人机共生的灵性世界。我们认为，人和产品的关系，在这个万物智能的时代，正在发生着巨大革新。我们经历了电器设备旋钮时代、键盘鼠标时代、触屏智能机时代，而现在万物智能时代的到来，很快就会将我们从鼠标键盘和低头点手机平板的触摸屏上解放出来，机器与人的交互会从不平等的单向被动模式转变为逐渐平等双向的多模态主动模式。而语音交互正是这个过程的第一步。我们认为这种互动模式的本质就是“关系”。我们把这种“基于关系的交互方式”称作 Relationship UI。物灵想要做的“智能 / 关系式交互”产品，不单单是计算机视觉、自然语言理解、深度学习、大数据等这些正在不断突破的 AI 技术的集合体，更加是一个让用户能体验到生命感，产生依赖，寄托情感的“灵物”。这本书里提到的 Jibo 机器人，正是 Relationship UI 的一个很好的例子。我作为 Jibo 的董事长，物灵作为 Jibo 在中国的独家合作伙伴，我们和 MIT 的 Cynthia Breazeal 的团队也正在一起努力把这个小可爱的小“灵物”带给中国的消费者。

当语音交互、智能匹配和个性化推送成为新一代用户的主流交互方式的时候，苹果构筑的那个以一个个各自为阵的 App 为核心的生态才能再一次瓦解——从移动互联到智能物联的转变，我们将经历 Contextual UI 取代 Selective UI 的产品形态。