基于智慧工厂的语音交互设计研究

桂宇晖[1]，刘婧[1]，刘军[1]，宋刚[2]
（1.中国地质大学，武汉 430074；2.武汉工程大学，武汉 430205）

摘要：目的 在人工智能高速发展，自然语言处理技术得到显著提高的背景下，探讨语音识别技术对智慧工厂类系统中交互设计范式的影响。**方法** 从形式层面、内容层面、操作层面对语音交互的要素进行剖析，结合公理化设计理论，得出语音交互研究框架，基于“中国石油智慧油库设备管理系统”的相关研发设计的实际验证，通过对四十名熟练操作人员的 GUI 与 VUI 输入交互方式进行对比测试分析，评估语音识别技术在智慧工厂交互设计中的应用可行性。**结论** 研究结果表明，选择多通道交互技术并行、随场景转换交互方式、感官语义的有效引导和构建虚拟人物角色的语音交互设计原则，能够降低用户的学习成本和使用智慧工厂系统的操作时间，提升操作效率，为语音交互在智慧工厂中的应用提供参考依据。

关键词：交互设计；语音交互；智慧工厂；人工智能
中图分类号：TB472 **文献标识码：**A **文章编号：**1001-3563(2020)06-0026-06
DOI：10.19554/j.cnki.1001-3563.2020.06.005

Voice Interaction Design Based on Smart Factory

GUI Yu-hui[1], *LIU Jing*[1], *LIU Jun*[1], *SONG Gang*[2]
(1.China University of Geosciences, Wuhan 430074, China; 2.Wuhan Institute of Technology, Wuhan 430205, China)

ABSTRACT: The work aims to discuss the influence of voice recognition technology on interactive design paradigm in smart factory system under the background of rapid development of artificial intelligence and remarkable improvement of natural language processing technology. The elements of voice interaction were analyzed from the aspects of form, content and operation. Combined with the Axiomatic Design (AD), the framework of voice interaction research was proposed. Based on the actual verification of the research and development design of “Petro China Intelligent Oil Depot Equipment Management System”, the feasibility of applying voice recognition technology in the interactive design of smart factories was evaluated by comparing the GUI and GUI input interaction modes of 40 skilled operators. The results show that the voice interaction design principle with multi-channel interaction technology in parallel, changing interaction mode according to scenes, effectively guiding sensory semantics and building Agent role can reduce user's learning cost and time of operating smart factory system, improve operation efficiency, and provide reference for the application of voice interaction in smart factory.

KEY WORDS: interaction design; voice interaction; smart factory; artificial intelligence

“智慧工厂”是智能制造最核心最集中的母体居 所，在某种程度上是 IBM 提出的“智慧城市”的先行

收稿日期：2020-01-09
基金项目：科技部国家重点研发计划重大科学仪器设备开发重点专项“大视场机载高光谱成像器”（国科高技发 2016-26 号）；教育部科技发展中心产学研创新基金-新一代信息技术创新重点项目（2018A04002）；湖北省教育厅：《基于虚拟现实技术的楚建筑复现研究》（18Q071）
作者简介：桂宇晖（1974—），男，江西人，博士，中国地质大学教授，主要研究方向为交互设计及理论、新媒体游戏、设计历史及理论。
通信作者：刘婧（1988—），女，湖北人，中国地质大学硕士生，主攻交互设计及理论。

范本和小区域测试场，起到了率先应用新兴技术，集中高密度智能，推动工业智能化发展的先锋作用[1]。智慧工厂关注机器智能与人类智能的耦合、机器特性和人类自然习性在生产过程中的配合，旨在使机器具备高级别分析、推理、判断、构思和决策等能力，因此需要进一步对智慧工厂的设计、管理及运行进行深入研究[2]。

“中国石油智慧油库设备管理系统”项目是基于人工智能技术，依托工业生产大数据，利用工业物联网的全国试点项目，涉及全国三十一个省油库设备管理系统的提升改造。其交互设计除了要应对基础的生产需求和两千多个子系统的集成，更需要应对人工智能、大数据和万物互联对设计对象与设计思维方法的改变，以应对创新与用户体验方面遇到的各类挑战[3]。如何提升用户的交互自然性，提高系统的可靠性是这一大类产品交互设计成败的关键。

人机交互系统中交互意图的准确传递与理解，对用户意图的准确判断至关重要[4]。语音交互较于传统单一路径操作方式的图形交互更为方便和自然，人的命令能直接传达到机器并获得反馈，简化了人机交互的反馈层级，使人机交互方式从传统的单一通道转向多通道融合成为现实，实现了从固定式输出输入交互阶段转向人机自然交互阶段[5]。语音识别技术、语音合成技术、自然语言处理等人工智能技术的推陈出新，使基于语音的交互设计成为了行业趋势。在国家战略层面，各国也进行了语音交互的布局，美国 2019 年的《国家人工智能研究与发展战略规划》中明确提出了“开发有效的人机协作方法，开发可视化和人机界面技术”，其中语音交互设计是提升用户体验的关键。如苹果的 Siri，微软的 Cortana，亚马逊的 Alexa，IFLYTEK 的讯飞翻译机，阿里巴巴的天猫精灵等。除此之外，语音交互在老人与儿童的陪伴机器人、在应急救援管理系统[6]中，都有相应的研究与应用。国内外对语音交互的研究主要集中在硬件和技术方面，在交互设计方面对语音用户界面的拟人化反馈、反馈时间、反馈语速、反馈人称代词等方面也有涉及[7-9]。本文旨在对“中国石油智慧油库设备管理系统”中的语音交互设计的价值进行分析，提出智慧工厂系统建设中的语音交互设计原则。

1　智慧工厂系统中的语音交互

中国石油智慧油库设备管理系统见图 1，是基于智能手机、平板电脑、MR 头戴显示设备开发的。利用三维建模、数字孪生、虚拟现实、混合现实等技术，构建一个仿真的油库生产环境，通过人工智能、物联网、云计算、数据分析与数据挖掘等技术，模拟整个厂区内的生产流程，并对所有数据进行可视化处理，将实时生产数据同步抓取到管理系统中，实现智能设备的可视化控制和管理。

图 1　中国石油智慧油库设备管理系统
Fig.1　Petro China intelligent oil depot equipment management system

语音交互具有无界面限制特点，因此在用户输入与系统反馈速度上与 GUI 相比更具优势，通过传感器实时将系统中的核心数据、信息和知识，以更智能、更敏捷、更高效的方式传递给用户，同时能够以更加自然的情感化交流提升用户体验。

1.1　自然情感寄托

如今，交互产品设计的重点和卖点在于情感化交互设计[10]，语音交互是近几年发展迅猛的新兴交互模式，在机器的情感表达及用户的情感激发上深受用户认同。通过真实的语音、语调甚至语气模拟，日益智能的“拟人应答”，辅以足够丰富的语料设计，语音交互分别从形式层面与内容层面体现出情感化交互的优势。

1）形式层面。一是创建虚拟人物，即接触系统的用户能够将其想象成一个虚拟的真实人物，而并非只将系统当成一个操作机器，通过接近拟人的智能回答，缓解用户对操作系统的抵触。二是选择音色、音调、响度。用户对声音的感知和声音的音色，影响着用户对声音隐喻意义的认知，以及用户的主观感性体验，因此在设计声音反馈时，应选择符合用户心理预期的声音为基础进行优化设计[11]，从以人为本的角度提升感官体验。如智慧工厂系统根据用户反馈的需求，将虚拟人物设定为一名 40 岁左右音色沉稳、睿智的男性高级工程师，旨在让第一次听到此声音的用户对他产生信任感，产生想要进一步与他沟通的想法，引起用户的兴趣。

2）内容层面。通过语义相似度的文本聚类方法设计多样语料[12]。维特根斯坦在《哲学研究》中提出的“语言游戏说”表明，语言具有多样性和差异性，对词语意义的把握应放在日常生活中去理解。由于每个用户的说话方式不尽相同，即使同一种命令操作，也会产生多种问答形式，因此将语料设计得越全面，用户操作的成功率就越大。例如，视察汽油泵棚设备这一操作，可以将语料设计成“启动汽油泵棚设备”、“查看 TD101 设备”、“调出汽油泵棚数据”、“核实汽油泵棚信息”等，智慧油库设备管理系统基于 VUI 交互设

备展示界面，见图 2。

1.2 接触体验转变

相较于 GUI 中手持设备在有限界面范围内的近距离操作交互体验，VUI 中声音的空间方位性满足了自然交互需求，能在空间虚拟环境内 360°的任何方位传递交互信息，打破了传统二维界面的限制。

1）释放双手。例如，用户在巡检时通常是对厂区内的设施和周围的环境进行环视检查，如果只用手触交互的方式，用户需要重复多次的“低头（看屏幕）—抬头（检查）”动作，长时间操作会增加用户头部的不适感，破坏用户的体验感受，采取语义交互的方式可以替代部分低头动作和减少手势操作的流程。

2）输入速度提高。语音输入比传统的打字输入、鼠标输入要快很多。例如，查找具体型号设备之类的复杂动作，只是输入设备型号这一交互过程就很费时费力，比起语音输入的方式体验感差很多。通过语音交互的方式，缩短了交互步骤，减少了用户通过眼睛凝视和手势输入的次数，可以大幅度提高用户的黏性。

3）无二维界面限制。语音交互帮助用户完成非重要步骤的选择与决策，只把最重要的决定交给用户[13]，让用户直接找到交互目标。GUI 输入受限于交互界面的面积，为了更好地解决空间上的限制，只能通过信息分层的方式安排操作功能，这样做的劣势是将用户操作固定在了统一的结构中，用户只能通过这种方式达到交互目的，而 VUI 输入交互不受信息纵深度约束，突破层级限制，在短时间内获取与处理大量信息。以在同一系统中查询相同设备（A 设备）信息的耗时测试为例，将随机选择的四十个熟练操作人员分为两组，每组二十人，同时进行测试，得出 GUI 交互方式获取信息需要耗时 13~21 s，VUI 输入交互方式获取信息需要耗时 3~6 s。

图 2 智慧油库设备管理系统基于 VUI 交互设备展示界面

Fig.2 Display interface of intelligent oil depot equipment management system based on vui interactive device

2 智慧工厂系统中的语音交互设计框架

交互设计的发展与交互技术紧密相关，人工智能是其中重要的交互技术，扩展了交互设计的视野与维度[14]。语音识别技术作为人工智能中的关键技术，在智能家居、智能车载、智能服务机器人等领域飞速发展，为交互设计提供了新的、以语音和手势为主的自然交互范式。因此，语音交互在智慧工厂中的应用研究能为其关键研究课题之一的软件平台的设计与开发，提供更自然丰富的情感化人机交互体验。

2.1 语音交互研究框架

人机语音交互流程的核心是理解用户的意图，因此语音交互研究的重点是识别用户意图以及处理用户意图。结合美国麻省理工学院 Nam P Suh 教授在“The Principles of Design”中提出的公理化设计理论，通过对受众群体语音交互需求的采集、评估及提取重点语音需求数据，创建相关的语音对话情境，分析所有可能出现的语音对话（即用户意图），将表达同一意思不同句法的语句记录下来，运用 ASR（自动语言识别）、NLU（自然语言理解）、NLG（自然语言生成）、TTS（文字转语音）等重点语音交互技术，开展语音交互设计，对用户的意图进行合理的反馈，语音交互研究框架见图 3。

2.2 语音交互的功能定义

智慧工厂系统的语音交互功能根据用户的需求，分析整理后得出三大功能类型，分别为操作类、演示类和导航类。语音交互功能定义见图 4。

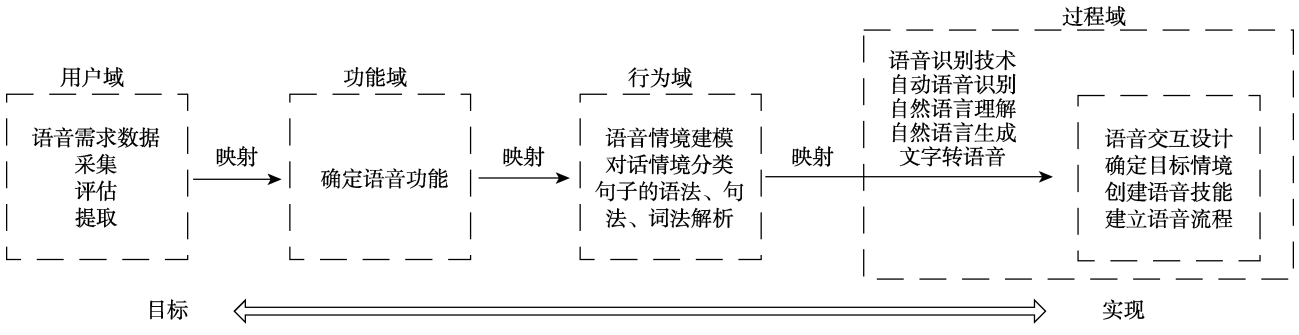

图 3 语音交互研究框架

Fig.3 Research framework of voice interaction

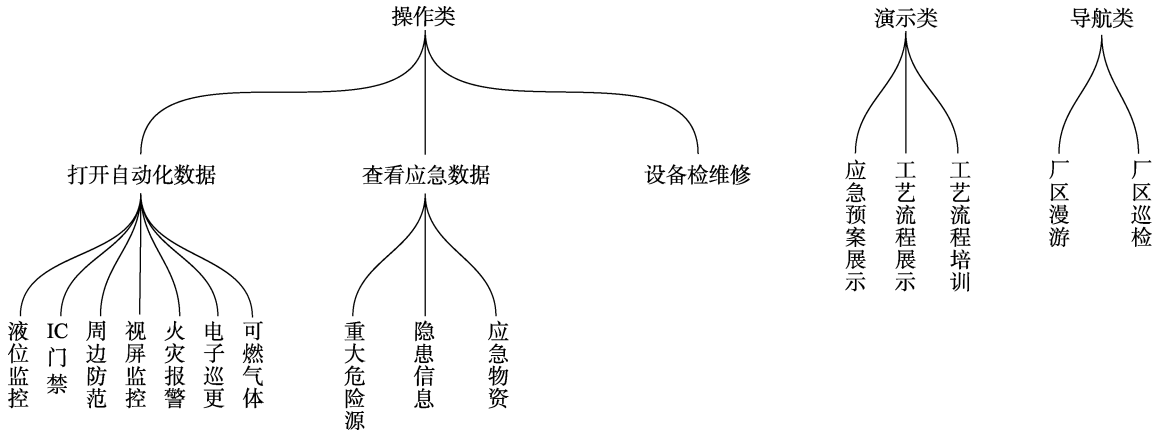

图 4　语音交互功能定义

Fig.4　Definition of voice interaction function

3　语音交互在智慧工厂系统中的设计原则

通过对中国石油智慧油库设备管理系统的 GUI 与 VUI 系统设计比较研究，发现语音交互的应用对提升智慧工厂系统的情感化交互体验，有较为显著的提升作用。语音交互设计模型见图 5。

3.1　多通道交互并行

因为符合人的交互模式，多通道交互（Multi-Modal Human-Computer Interaction，MMHCI）被认为是更为自然的人机交互方式[15]。多通道并行协作的工作模式，降低了单通道的工作负担，以此触发的多感官交互可以产生通感，从而提升用户对产品的体验质量[11]。人与人的交互模式主要通过语言沟通和非语言沟通来传达。语言沟通可以分为口头沟通和笔试沟通两种，非语言沟通是指人的举止神态，如肢体动作、声音的语气语调、眼神交流、微笑等。而最有效的沟通是将语言、身体与表情、以及手势等语言沟通和非语言沟通方式一起应用，才能达到沟通的最佳途径。根据上述理论及智慧工厂的系统设计研究，从人机交互友好性和自然性方面看，多通道交互并行减少用户的认知负担，比传统的单一通道交互，如鼠标键盘输入或笔触、手触的图形输入交互更为合适与高效。多通道交互并行方式与人的沟通方式比较见图 6。

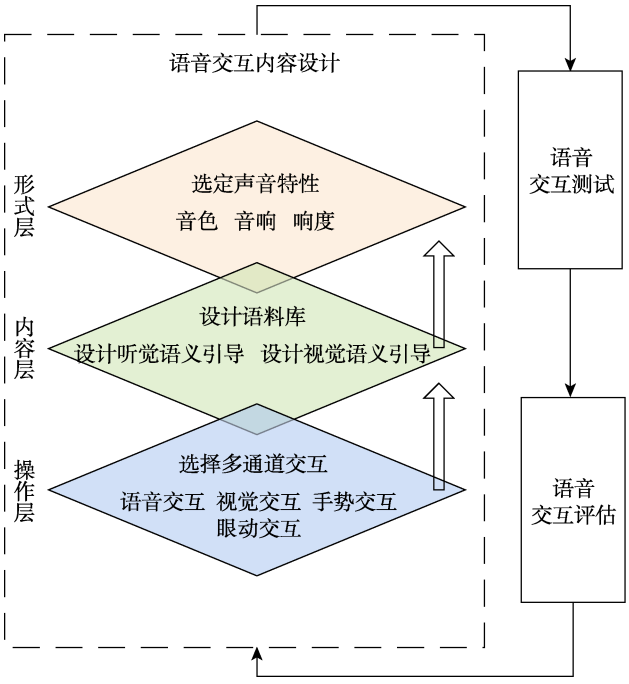

图 5　语音交互设计模型

Fig.5　Voice interaction design model

3.2　随场景转换交互方式

在智慧工厂系统的使用过程中，由于受使用场景位置和环境的影响，机器噪音、人声干扰、物理环境嘈杂等因素对语音识别有不同程度的干扰，因此在进行语音场景设定时，要充分考虑一些特定的场景（如机房、加油站附近噪音比较大的地点），对这些特殊场景提供可选择替换的交互方式，例如弹出提示语，建议用户以图形用户界面操作为主，语音用户界面操作为辅，以保证用户的需求能顺利有效地传达给智能系统。

3.3　感官语义的有效引导

语音交互主要通过说和听的方式完成机器与用户的沟通，用户听到的内容和机器识别的内容都是短暂即逝的，若在听的过程中疏忽掉某些信息，就意味着可能失去理解的机会。针对语音交互这一时效性特点，做好视觉语义引导和听觉语音引导是必不可少的。

1）视觉语义引导。对于有显示屏幕的智能机器，通过对用户触屏行为的分析，在非屏幕中心位置推送相关语义提示，告知用户可以用语音搜索的方法更便捷地达成查找意图。例如，当用户在系统里查找某一设备时，可以提示用户尝试说出“帮我找到上次出问题的设备”；或是用户在左右滑屏时，可能是在查找软件图标，这时可以提醒“可以说出要查找的内容，可能更快哦”。

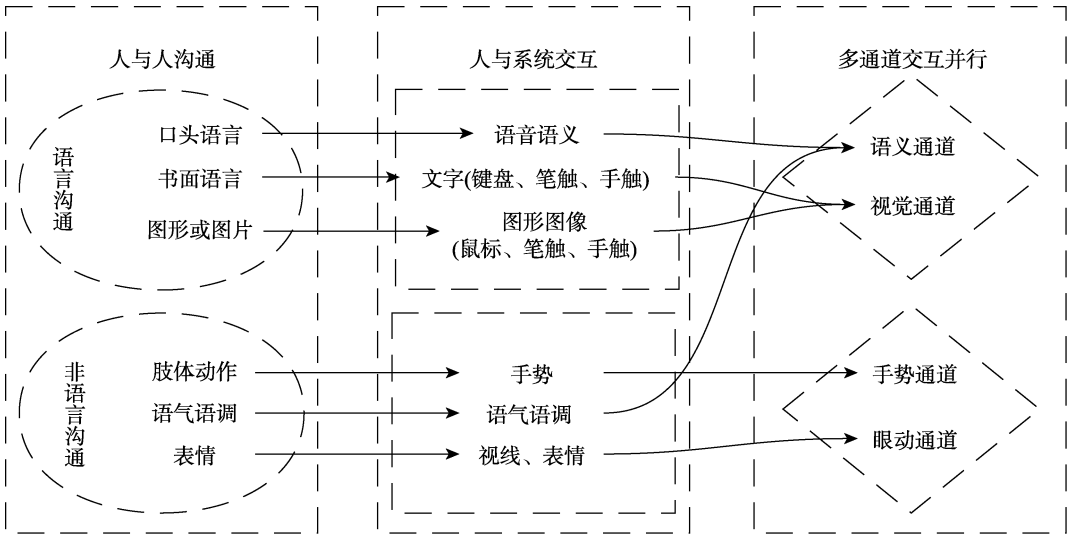

图 6 多通道交互并行方式与人的沟通方式比较

Fig.6 Comparison between multi-channel interaction in parallel and human communication

2）听觉语音引导。除了视觉引导外，听觉引导主要用于首次使用、提问出现迟疑或是识别出错时对异常情况的处理，通过对问题的实时反馈，提高语义识别的容错率。比如，当系统没有听清用户的指令时，不要说“我不明白您说的什么”，而是用“我不太清楚您的意思，您是说‘查看 TD102 号油罐信息吗？’”来代替；当用户出现迟疑时，可以说“您可以这样问我‘今天厂区设备运行情况如何？’”。

3.4 构建虚拟人物角色

语音交互是计算机通过语义识别技术模仿人的语音、语调、情感与用户直接进行沟通。为了提高语音交互的真实性，通过产品定位创造出虚拟的人物角色，让用户感觉到是在跟一个“人”而不是一个“机器”进行对话是必要的。即使没有进行人设角色设定，用户依然会通过对系统对话特征的理解在脑海中设想一个虚拟形象，由于每个用户认知程度的不同，这个虚拟形象是截然不同的，所以使产品定位产生了不可控性。

人物角色的设定可以根据用户画像进行分析，判断符合目标用户个性喜好的人物声音特征与沟通基调，找到真实人物原型，创建与原型一致的角色来完成。比如，智慧工厂系统可以采用“助手”、“专家”或“工程师”等人物形象，以更加自然的交互方式实现情感化交流。

4 结语

目前，语音交互研究方法还缺乏全面、深入的研究，本文所提出的语音交互研究框架只是初步探讨。未来，在人工智能技术的快速发展与推动下，智慧工厂语音交互应用方法研究将持续更新。基于语音与视觉、手势、眼动等多通道结合的自然语言交互方式，是交互设计的发展方向，无论智能技术如何突破、创新，人机交互都会向着以人为本、情感化方向不断进步。

参考文献：

[1] 张志强, 胡山鹰, 胡雪瑶, 等. 智慧工厂综合管理信息系统开发及应用[J]. 化工进展, 2016, 35(4): 1000-1006.
ZHANG Zhi-qiang, HU Shan-ying, HU Xue-yao, et al. Development and Application of Integrated Management System of Smart Factory[J]. Chemical Industry and Engineering Progress, 2016, 35(4): 1000-1006.

[2] 张益, 冯毅萍, 荣冈. 智慧工厂的参考模型与关键技术[J]. 计算机集成制造系统, 2016, 22(1): 1-12.
ZHANG Yi, FENG Yi-ping, RONG Gang. Reference Model and Key Technologies of Smart Factory[J]. Computer Integrated Manufacturing Systems, 2016, 22(1): 1-12.

[3] 覃京燕. 量子思维对人工智能、大数据、万联网语境下的交互设计影响研究[J]. 装饰, 2018(10): 35-39.
QIN Jing-yan. Research on the Impaction of Quantum Thinking on Interaction Design in the Context of AI, Big Data and Internet of Everything[J]. Zhuangshi, 2018(10): 35-39.

[4] 赵其杰, 邵辉, 卢建霞. 基于头眼行为的交互意图检测方法[J]. 仪器仪表学报, 2014, 35(10): 2313-2320.
ZHAO Qi-jie, SHAO Hui, LU Jian-xia. Identification Method of Interaction Intention Based on Head and Eye Behaviors[J]. Chinese Journal of Scientific Instrument, 2014, 35(10): 2313-2320.

[5] 吴琼, 赵毅平. 人机自然交互的前沿探索——阿里巴巴国际 UED 负责人傅利民专访[J]. 装饰, 2018(10): 40-43.
WU Qiong, ZHAO Yi-ping. Frontier Exploration on Natural Human-Computer Interaction: an Interview with Paul Fu, Sr. Director of User Experience, Alibaba Group

[J]. Zhuangshi, 2018(10): 40-43.

[6] 李文峰, 李大娟, 张军会. 语音交互技术在矿山救护队应急救援管理系统中的应用[J]. 工矿自动化, 2019, 45(11): 98-108.
LI Wen-feng, LI Da-juan, ZHANG Jun-hui. Application of Voice Interaction Technology in Emergency Rescue Management System for Mine Rescue Team[J]. Industry and Mine Automation, 2019, 45(11): 98-108.

[7] 廖青林, 王玫, 冯战. 基于情感交互的智能家居产品语音交互设计[J]. 包装工程, 2019, 40(16): 37-42.
LIAO Qing-lin, WANG Mei, FENG Zhan. Voice Interaction Design of Smart Home Products Based on Emotional Interaction[J]. Packaging Engineering, 2019, 40(16): 37-42.

[8] 李悦, 王军锋, 王文军. 基于时间知觉的语音用户界面反馈机制优化[J]. 装饰, 2019(7): 100-103.
LI Yue, WANG Jun-feng, WANG Wen-jun. Optimization of VUI Feedback Mechanism Based on the Time Perception[J]. Zhuangshi, 2019(7): 100-103.

[9] 孙妍彦, 李士岩, 陈宪涛. 情感化语音交互设计—百度 AI 用户体验部门人机交互研究地图与设计案例[J]. 装饰, 2019(11): 22-27.
SUN Yan-yan, LI Shi-yan, CHEN Xian-tao. Emotional Voice Interaction Design: Human Computer Interaction Research Map and Design Case of Baidu AI User Experience Department[J]. Zhuangshi, 2019(11): 22-27.

[10] 兰玉琪, 刘湃. 基于用户体验的交互产品情感化研究[J]. 包装工程, 2019, 40(12): 23-28.
LAN Yu-qi, LIU Pai. Emotionalization of Interactive Products Based on User Experience[J]. Packaging Engineering, 2019, 40(12): 23-28.

[11] 谭征宇, 杨文灵. 面向自然交互的声音通感设计研究[J]. 包装工程, 2018, 39(8): 68-73.
TAN Zheng-yu, YANG Wen-ling. Sound Synaesthesia Design Based on the Natural Interactions[J]. Packaging Engineering, 2018, 39(8): 68-73.

[12] 毕强, 刘健, 鲍玉来. 基于语义相似度的文本聚类研究[J]. 现代图书情报技术, 2016(12): 9-16.
BI Qiang, LIU Jian, BAO Yu-lai. A New Text Clustering Method Based on Semantic Similarity[J]. New Technology of Library and Information Service, 2016(12): 9-16.

[13] 杨洋. 基于语音交互的手持移动端多通道交互研究[D]. 北京: 北京邮电大学, 2017.
YANY Yang. Research on Multi-Channel Interaction of Handheld Mobile Device Based on Voice Interaction[D]. Beijing: Beijing University of Posts and Telecommunications, 2017

[14] 覃京燕. 人工智能对交互设计的影响研究[J]. 包装工程, 2017, 38(20): 27-31.
QIN Jing-yan. Impaction of Artificial Intelligence on Interaction Design[J]. Packaging Engineering, 2017, 38(20): 27-31.

[15] 杨明浩, 陶建华. 多通道人机交互信息融合的智能方法[J]. 中国科学: 信息科学, 2018, 48(4): 433-448.
YANG Ming-hao, TAO Jian-hua. Intelligence Methods of Multi-Modal Information Fusion in Human-Computer Interaction[J]. Scientia Sinica Informationis, 2018, 48(4): 433-448.

（上接第 25 页）

[19] 覃京燕. 人工智能对交互设计的影响研究[J]. 包装工程, 2017, 38(20): 39-43.
QIN Jing-yan. Impaction of Artificial Intelligence on Interaction Design[J]. Packaging Engineering, 2017, 38(20): 39-43.

[20] 胡洁. 人工智能驱动的艺术创新[J]. 装饰, 2019(11): 12-17.
HU Jie. Artificial Intelligence Driven Art Innovation[J]. Zhuangshi, 2019(11): 12-17.

[21] GATYS L A, ECKER A S, BETHGE M. A Neural Algorithm of Artistic Style[J]. CVPR2016, 2015.

[22] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Image Net Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2): 1097-1105.

[23] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-scale Image Recognition[J]. NIPS, 2014.

[24] 王伟凝, 王励, 赵明权, 等. 基于并行深度卷积神经网络的图像美感分类[J]. 自动化学报, 2015, 42(6): 904-914.
WANG Wei-ning, WANG Li, ZHAO Ming-quan, et al. Image Aesthetic Classification Using Parallel Deep Convolutional Neural Networks[J]. ActAutomatica Sinica, 2015, 42(6): 904-914.

[25] 王晓慧, 覃京燕. 基于深度学习的情感化设计[J]. 包装工程, 2017, 38(6): 12-16.
WANG Xiao-hui, QIN Jing-yan. Emotional Design Based on Deep Learning[J]. Packaging Engineering, 2017, 38(6): 12-16.

[26] 厉智, 孙玉宝, 王枫, 等. 基于深度卷积神经网络的服装图像分类检索算法[J]. 计算机工程, 2016, 42(11): 309-315.
LI Zhi, SUN Yu-bao, WANG Feng, et al. Clothing Image Classification and Retrieval Algorithm Based on Deep Convolutional Neural Network[J]. Computer Engineering, 2016, 42(11): 309-315.

[27] 张振焕, 周彩兰, 梁媛. 基于残差的优化卷积神经网络服装分类算法[J]. 计算机工程与科学, 2018, 40(2): 354-360.
ZHANG Zhen-huan, ZHOU Cai-lan, LIANG Yuan. An Optimized Clothing Classification Algorithm Based on Residual Convolutional Neural Network[J]. Computer Engineering and Science, 2018, 40(2): 354-360.