信息无障碍中的智能交互技术

关键词：信息无障碍　智能交互

特邀编辑：史元春
清华大学

信息无障碍 (information accessibility) 是一个学科交叉的技术和应用领域，旨在用信息技术弥补残障人士生理和认知能力的不足，让他们可以顺畅地与他人、物理世界和信息设备进行交互。据中国残联统计，中国现有 8500 万残疾人，是世界上残疾人口最多的国家。其中，听力残疾 2000 万人，视力残疾 1200 万人，各类肢体残疾 2500 万人，智力残疾和精神残疾 1200 万人……随着社会老龄化程度加重，残疾人口数量也在持续增长。互联网和用户终端的普及，使得信息无障碍成为一个越来越值得关注的领域，目标是解决残障人士的信息访问甚至是生活服务问题。

信息无障碍始于个人计算设备，IBM 在 1984 年首次开发了基于桌面操作系统的读屏程序，让盲人可以使用电脑。1997 年，万维网联盟 (W3C) 成立了网络无障碍推动 (WAI) 小组，推动网络页面的信息无障碍。随着人工智能技术的发展，语音识别、图像识别、手语翻译等技术也被应用到信息无障碍领域中，支持更多的残疾用户（比如聋哑人）通信和访问信息设备。微软在 2018 年推出了“人工智能无障碍计划”(AI for Accessibility)，国内的互联网公司阿里巴巴、百度等也积极推出智能读图等无障碍交互应用。

在学术界，专门交流无障碍技术的国际会议是 ASSETS (ACM SIGACCESS Conference on Computers and Accessibility)，1994 年是第一届。如今智能手机上基于触摸的读屏系统原型就是在 ASSETS 2008 上提出的。2017 年，人机交互领域的旗舰会议 ACM CHI (ACM CHI Conference on Human Factors in Computing Systems) 将无障碍列为十大方向之一。

我国于 2004 年举办了第一届中国信息无障碍论坛。2006 年，工信部把信息无障碍工作纳入“阳光绿色工程”，并由中国通信标准化协会 (CCSA) 开始系统化制定信息无障碍标准。2008 年，工信部发布了首个网站设计无障碍技术标准，信息无障碍在“十三五”规划期间被纳入国家发展计划。

从研究和应用水平上看，信息无障碍总体还处于比较初步的状态。在应用上，针对信息访问和设备使用，具有基本功能的技术可以被应用，但效果和效率等可用性指标都不高；在现实生活中，针对听障人士与他人交流、盲人独立出行等，能支撑的新技术还处于原型和概念阶段。我们组织本期专题的目的，就是让读者了解无障碍领域目前存在的问题和研究进展，了解此交叉学科具有的重要研究与应用价值，希望更多的科技工作者能够介入或投身到相关研究中去，从各自的角度参与和贡献力量。

信息无障碍交互技术的现状

残障类型多样，所需要的无障碍技术也不尽相同，这里介绍三类主要的残障类型（视觉障碍、听觉障碍和运动障碍）人群遇到的问题和主要的技术解决方案。

视力残疾用户的需求包括独立出行、识别身边物体、与信息设备交互等。针对独立出行的需求，

目前有基于计算机视觉的道路识别技术，通过立体声场或者震动反馈为视力残疾用户指示方向。但是这些设备目前还不能取代盲杖，还需要更多的技术突破。针对识别物体的需求，主要是利用视频 / 图像转换为文本的技术，包括微软的 Seeing AI 和谷歌的 Lookout 都是此类应用。针对使用手机和电脑的需求，主要采用读屏程序 Screen Reader（苹果手机上的 VoiceOver 或者安卓系统上的 Talkback，均为系统默认自带功能），可以通过语音读出获得焦点的控件信息，这样视力残疾用户通过听就能了解设备界面上的信息内容。

听力残疾用户面临的主要问题是与人交流存在障碍，以及观看视频内容时听不到声音。老年听力障碍是指随着年龄增长，听觉器官的衰老和退变所导致的听觉功能下降，发病率居世界第三位。助听器设备通过放大声音信号，可解决“听不到”的问题；但对于听觉中枢受损的人，声音信号分析能力却难以弥补，解决不了“听得清”的问题。针对听障用户，相关信息无障碍技术包括手语的识别与合成，以及语音识别技术。

运动控制能力缺失的用户，包括上肢残疾，或者患帕金森症、脑瘫、肌肉萎缩、渐冻症等疾病的用户。他们丧失了灵活控制手指运动的能力，而手指是人表达交互意图的主要运动器官，也是电脑和手机的主要操作器官。在构建面向这类用户的信息无障碍交互技术时，其中一个难题是用户的差异性，几乎每个用户的可运动部位及其运动能力都是不同的，给构建适合于个体的通用输入技术带来了挑战。相关的技术方案有眼动，但是用“眼动”作为输入方式时，缺少“确认”操作，容易产生误触发，且操作精度有限。

信息无障碍的主要科学问题

音视频的理解和信息转换（主要针对听障和视障）。视觉和听觉是人们接受信息的主要感官。听障和视障用户因为缺乏某种感官而无法完整理解信息，需要建立音视频的理解技术，用机器算法理解音视频内容的语义，进而转换为用户可用感官能接受的信息类型，包括音频和文字之间的语音识别和文本到语言 (Text To Speech, TTS) 技术，图像到文字和视频到文字的技术。目前，精度是主要问题，尤其是克服多种噪声条件下的高精度实现，对于这些技术的可用性起到关键作用。

图形用户界面到声音界面的编码转换（主要针对视障）。个人电脑和手机都是图形用户界面，信息以可视的方式传递给用户，而视障用户只能通过听觉（触觉为辅）来接收信息，相比于视觉，不仅信息接收的带宽要低很多，而且信息呈现的模式也发生了变化。视觉提供整体和并行的信息获取能力，听觉只能提供局部串行的信息。这也会影响用户对于交互界面的心理模型，进而影响到交互决策。因此，需要研究从图形界面到声音界面的编码转换方法，优化“读屏”的方法。

个性化信息输入和意图理解（主要针对视障和运动障碍）。人体的运动控制系统包括运动执行和反馈两部分。运动障碍用户无法精确灵活地控制手指运动，视障用户由于缺少视觉反馈也不能做精确的输入控制，导致物理运动自由度受限和运动控制精度低的问题。前者需要开发具有个性化能力的输入技术，根据用户实际可以控制的输入范围来映射有效的输入；后者需要实现从有噪声的运动控制数据中提取用户的交互意图。

通过智能交互技术实现信息无障碍

信息无障碍是以用户为中心的交互方案，是对人的交互性能的优化。优秀的信息无障碍技术要适应用户的生理和认知能力，而不是让用户适应技术。为此，要采用智能交互方法来开展研究，从用户角度来设计和创新适用的交互模式，通过智能传感、智能用户意图推理和智能信息呈现来构建信息无障碍的交互界面。

对用户行为和认知能力的准确建模 需要建立用户动作能力和心理模型的计算模型，建立并

引入生理、心理的先验知识来描述用户的信息输入输出能力，对用户意图、表达方式、动作控制能力做统计建模和描述。残障用户的一个重要特点就是个性化，每个用户的信息输入输出能力都不同，除了需要研究合适的模型函数，还要研究个性化参数的计算方法，力求能准确地描述个体残障用户的能力。

智能的感知技术 需要研究高精度的感知技术，感知用户的外在动作，也感知用户的内在心理状态。信息无障碍中，智能感知的应用场景非常丰富，针对不同类型的残障用户、不同的交互任务、不同的使用情景，都需要适合的传感方案。哪些动作或者心理状态是有交互价值的，如何采用成本可控、易于部署的硬件方案都是需要考虑的问题。

智能的意图推理技术 残障用户通过自然动作（手、眼或身体其他部位）表达交互意图。优化信息无障碍的交互体验，需要使用户在表达意图时的生理和心理开销最小化，但这样将导致用户的表达方式和表达动作都是不精确的，在时间和空间上都存在随机性。如何从连续随机的行为数据中提取用户的输入意图，是需要解决的问题。作为操控型的交互界面，需要具有高精度、可理解、结果可预测的特点。

智能的信息呈现 感官残障用户对于交互界面的理解是不完整的，交互决策的心理模型带有随机性。高可用的信息无障碍交互界面，首先需要有对用户信息需求的预测能力，确定信息输出的目标；然后根据用户的信息接收能力将目标信息编码到具体模态上，编码方式涉及多模态融合；最后根据交互情景，对信息呈现的编码方式做动态优化调整，保证用户接收信息的有效性。

本期专题

本期专题邀请了5篇文章。美国康奈尔大学的Enhancing Ability Lab实验室是一个国际著名的面向弱视用户的研究团队，赵宇航和希里·雅曾科 (Shiri Azenkot) 教授撰写的《利用增强现实技术增强弱视人士的视觉能力》详细介绍了增强现实技术是如何改善视弱人士的日常活动的。中国科学院计算技术研究所研究员陈益强和王向东的研究团队长期开展面向聋人和盲人的信息无障碍技术的研究，他们分别撰写了文章《面向聋人的信息无障碍技术——手语识别与合成》与《面向盲人的信息无障碍技术——盲文翻译与识别》。中国残联与清华大学在2016年成立了清华大学无障碍发展研究院，计算机系人机交互实验室发挥其专业能力，在研究院重点开展信息无障碍交互技术研究，易鑫等人共同完成的《基于稳态视觉响应的脑电接口动态分组键盘技术》详细介绍了脑电响应的影响因素和为肢残用户设计的优化交互信息效率的动态分组键盘技术；石伟男和喻纯撰写的《基于字符级纠错的盲人智能键盘》分析了盲人在触屏手机上键入文本时的难题，他们构建的适合于盲人的智能输入法获得CHI 2019最佳论文提名奖。 ■

（本期专题责任编委：**贾珈**）

史元春

CCF会士，CCF常务理事，*CCCF*前专题主编。清华大学教授。主要研究方向为人机交互、普适计算、多媒体、网络教育技术等。

shiyc@tsinghua.edu.cn