人机交互调研报告

人机交互的发展历史，是从人适应计算机到计算机不断地适应人的发展史，随之出现的主要的交互界面形式依次为命令行界面、图形用户界面、触摸交互界面和三维交互界面，可以看出，随着交互界面的演变，交互的自然性逐渐提高，但由于交互接口尺寸的限制和触觉等反馈信道的受限，导致了输入的精度和交互效率反而逐渐降低。这种交互自然性和高效性之间的制约关系，成为了人机交互研究中的难题，如何在两者之间兼顾和平衡，是具有重要理论和实践意义的研究问题。

下面将从技术、应用和前沿研究三方面展开人机交互领域的的调研报告。

1. 技术分析：

本篇基于用户与系统之间进行交流操作所主要使用的触摸、手势、语音和视觉等自然感官的顺序来进行相关技术介绍。具体包括触控交互、声控交互、动作交互、眼动交互、虚拟现实输入、多模态交互以及智能交互等人机交互技术。

1.1 触控交互

触控功能与显示器的一体化模式使得显示器从仅向用户输出可视化信息发展成为了一种交互界面装置。触控技术能让人们通过触摸就能直接与屏幕内容互动，从而不用或进行很少的训练就能有更为便捷的使用体验。在工作原理上，有电容、电阻、光学和声学交互等不同的实现方式。

电阻式触控技术：电阻触摸屏通过压力感应原理来实现对屏幕进行操作和控制。手指触摸屏幕时薄膜下方能形成导电通路，可计算出触点的X和Y轴坐标，从而确定触摸点的位置。

电容式触控技术：手指触摸电容式触摸屏时，因为工作面上有高频信号而手指相当于导体，所以会引发触摸屏表面上的电流分布变化，测量计算四个方向电流比例，可得出触点坐标。

红外触控技术：红外触控依靠的是在屏幕外框安装红外发射管和红外接收管，当手指触摸屏幕时，红外光线被阻断，控制端根据红外接收端电压的变化可计算出触点位置。

表面声波触控技术：表面声波触控原理与红外触控相似，不同的是用安装在边角上的超声波换能器来替代红外，根据手指是否阻挡声波能量的传播来计算触点位置。

1.2 声控交互

声控交互主要包含语音识别、自然语言处理及语音合成三项关键技术。

1.2.1 语音识别

语音识别主要由四个部分组成：特征提取、声学模型、语言模型和解码器搜索。

特征提取是为了从模拟的语音信号汇总采样得到合适的声学特征参数供后续声学模型训练使用，好的声学特征应当具有比较优秀的区分特性、尽量低的参数维度和较好的鲁棒性。当前最常用的声学特征提取办法是基于倒谱分析的特征参数提取。

声学模型在语音识别中至关重要，描述了声学基元产生特征序列的变换过程。给定一个声学特征矢量，依据声学模型和概率的办法来确定与特征序列对应的状态序列。在声学模型的研究中两个比较核心的问题是声学基元的选择和隐马尔可夫模型问题。

语言模型刻画人类语言表达的方式习惯，着重描述了词与词在排列结构上的内在联系。当前最常使用的语言模型是基于N-gram的统计语言模型。

解码器是在完成了声学特征提取，声学模型和语言模型训练后，利用相关搜索算法找出最优词序列的工具，可以想见，如果不作任何限制，搜索空间随着词表大小和语音中可能出现的词数目是以指数级增长的。所以，解码器搜索效率的高低直接关系到语音识别系统的实用程度。必须通过一些有效的优化算法，提高解码效率，把原来超大规模的搜索问题压缩到计算机可以有效处理的程度。

1.2.2 语音合成

语音合成是将一系列的输入文字序列信号经过适当处理后，产生出具有尽可能高自然度的语音输出，从而使计算机能够像人一样说话的技术。

语音合成可以按照设计的主要思想分成规则驱动方法和数据驱动方法。前者主要思想是根据人类发音物理过程从而制定一系列规则来模拟这一过程，后者则是在语音库中的数据上利用统计方法如建模来实现合成的方法。规则驱动方式包括共振峰合成、发音规则合成，数据驱动方式包括波形拼接合成、基于隐马尔可夫模型合成和深度神经网络方法。

1.3 动作交互

1.3.1 发展现状

动作交互包含几方面问题：基于动作的目标获取技术、自然的动作命令映射方法和自然交互动作的识别方法。

1.3.2 手势识别

手势可定义为人手或者手和手臂的组合所产生的各种姿态和动作，手势的识别需要依次进行图像的获取、手的检测和分割、手势的分析、静态或动态的手势识别这几个环节。

手势分割是手势识别过程中关键的一步，目前最常用的手势分割法主要包括基于单目视觉的手势分割和基于立体视觉的手势分割。

手势分析是完成手势识别的关键技术之一，通过手势分析可以获得手势的形状特征或运动轨迹。手势分析的主要方法有以下几类：边缘轮廓提取法、质心手指等多特征结合法以及指关节式跟踪法等。

手势识别是将模型参数空间里的轨迹（或点）分类到该空间里某个子集的过程。其包括静态手势识别和动态手势识别，常见方法主要有：模版匹配法、神经网络法和隐马尔可夫模型法。

1.3.3. 姿势识别

姿势识别通俗地讲就是研究一种时变数据的分类问题，其实质就是研究如何从样本中学习获取一组典型的身体姿势的参考序列。常用算法有三类：基于模版匹配的身体姿势识别方法、基于状态空间的身体姿势识别方法、基于语义描述的身体姿势识别方法。

基于模版匹配的方法是将姿势序列转化为静态的模型来处理。其首先需要获得一个姿势模版库，然后计算需要测试的身体姿态的数据与模版库中的姿态的数据进行匹配的相似度。其研究重点是如何找到一个比较好的模版库。

基于状态空间的方法是将姿势转化为状态和联合概率。将姿势序列视作选取的静态姿态在特定状态节点之间的一个状态切换，通过计算每个状态到达其他状态的联合概率，选取最大的联合概率作为分类的准则。

基于语义描述的方法利用制定的语法格式将场景中人物的活动情况用填空的形式填入到对应的语法格式中，其实质就是用自然语言达到对场景描述的目的。

1.4 眼动交互

眼动交互研究的问题包括眼动行为的特征分析及参数化、可用性分析、眼动数据可视化及可视分析、眼动跟踪算法及相应眼动仪装置的研发。近年来，随着人工智能技术的发展，眼动跟踪作为感知和理解用户的关键环节，引起了相关学者的关注，目前，针对眼动跟踪AI系统的研究工作有：基于群智感知的眼动计算与分析、基于大数据学习的眼动跟踪和眼动数据与脑电数据融合的智能交互。

1.4.1 常见交互方式

眼动，即眼球的运动。早期研究者们发现人们阅读过程中的视线不仅存在注视和平滑扫视，还存在一系列短暂的跳动和停顿。这一发现使得人们开始进行眼动行为的相关研究和探索。目前常用的几种眼动交互方式有驻留时间触发、平滑追随运动、眨眼、眼势（一系列有序的视线行程）等。

1.5 虚拟现实输入

近年来虚拟现实技术迅速普及，目前已开发有多种适用于虚拟现实的文本输入技术，从输入机制角度可分为实体键盘技术、虚拟键盘技术和新型输入技术等。

1.5.1 实体键盘技术

实体键盘有多种实现方式，例如基于QWERTY实体键盘技术，基于移动实体的输入技术、虚实混合技术等。

1.5.2 虚拟键盘技术

虚拟键盘文本输入是目前最常用的输入方式之一，解决了对实体键盘的依赖性问题。但目前的虚拟键盘仍然保持着QWERTY键盘的布局模式，利用用户熟悉度，降低用户学习成本。但是虚拟键盘的输入也利用了不同的确认方式，包括手柄射线、头部射线、手部、手指、手腕手势、触碰眼动、手写等方式。目前的实现方式主要有基于手持控制器的输入技术、基于头部的输入技术、基于手部的输入技术、基于触摸屏的输入技术、基于视点的输入技术、基于空中敲击的输入等。

1.5.3 新型输入技术

新型的输入技术有手部输入技术、圆形键盘输入技术和立体输入技术。

手部输入技术通过在食指上佩戴小型触摸表面，将字母分配在2x3的网格中，通过拇指点击表面不同区域来实现选中输入。

圆形键盘输入则将一个圆圈分为7个切片，每个切片包含4个字符，用户使用操纵杆的双指拨杆与该布局交互-右指拨杆用于在圆形键盘周围移动，而左指拨杆用于选择字符。

立体输入则更加适配3D物体实现，其输入键盘使用的是3D立体键盘，将字母排列成3x3x3的3D阵列，包含27个单元格，其中26个代表26个英语字母，中间一个是空白单元格。用户使用控制器在单元格中导航以选择字符。

1.6 多模态交互

多模态交互指的是组合不同形式的输入模式（例如，语音、手势、触摸、凝视等），其目标是向用户提供与计算机进行交互的多种选择方式，以支持自然的用户选择。与传统的单一界面相比，多模态界面可以被定义为多个输入模态的组合，这些组合可以分为6种基本类型：互补型、重复型、等价型、专业型、并发型和转化型。

1.7 信息无障碍中的智能交互技术

信息无障碍技术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。信息无障碍的主要科学问题包括：音视频的理解和信息转换、图形用户界面到声音界面的编码转换、个性化信息输入和意图理解。信息无障碍要适应的是用户的生理和认知能力，而不是让用户适应技术，所以要采用智能交互的方法来开展研究。从智能交互的角度来看，有以下几点主要内容：对用户行为和认知能力的准确建模、只能的感知技术、智能的意图推理技术和智能的信息呈现。

2.应用领域分析：

2.1智能终端领域

2.1.1智能手机终端

从功能机到智能机的过渡使得用户与手机的交互方式从实体按键转换为了触摸屏，当下主流的交互方式仍采用的是触摸屏幕方式，随着用户市场需求和手机厂商的双向作用，手机上的实体按键也越来越少，并且各大手机厂商也努力在交互方式上也在寻求一定的创新。从虚拟导航键的出现代替实体的菜单键，用画面的形式可以达到同样的功能并且在交互体验上有了新的提升，虽然会占用一定的屏幕空间，但是同时也省去了机身的空间占用。再到后来的指纹解锁、人脸识别解锁等功能的出现让用户的交互体验有了新的层次的上升，同时现在各大手机厂商均在手机上内置了智能语音助手用来减少使用时的复杂操作，然而AI语音助手的无论是从灵敏度还是人性化的角度体验虽然还有待提升，但是也给用户和终端的交互提供了更多可能。

2.1.2智能车载终端

智能车载终端一直是嵌入式应用的一个重要领域，随着汽车执行功能的变化，人们迫切需要的不只是一个代步工具，而是期望能够将生活的更多方面融入于此，因此汽车与用户的交互也逐渐从实体操作装置转变为更加人性化的方式，目前触控、语音、手势甚至瞳孔识别的部分功能也应用到了车载终端领域。

2.2可穿戴设备

可穿戴移动终端领域一直是国内外人机交互研究的热点领域，近年来的机器学习方式来实现的手势识别和分类也成为了可穿戴领域的主流解决方案，微软研究院曾据此提出了判别式分类器，能够在深度图像和红外图像上快速实现手势的识别分类，但是此种应用有场景和时间的限制，所以还仍处于研究阶段。同时通过可穿戴设备可以实现生理信号的采集，当前关于此领域的研究主要过渡到了新型纳米材料的研究，借此以实现穿戴设备的心电、肌电等信号的采集。

2.3智能家居领域

2.3.1手机交互及跨屏交互端

目前最为广泛的智能家居交互方式还是通过手机终端实现的，大部分是借助于手机app来取代传统的远程控制或是定时开关，现在许多厂商在自己旗下的产品间尝试建立生态链，以情景模式的方式实现一键控制各类家居设备。

2.3.2语音交互和手势交互端

语音交互方式又分为直接语音与间接语音的方式，亚马逊的echo和谷歌的GoogleHome就是采用的直接语音方式进行交互。手势交互方式则主要是通过感应用户体感或者通过摄像头识别手势来接收和响应指令，当然这也是他原理性的局限之一。

2.4游戏领域

电子游戏的发展正逐渐向“自然化交互方式”转变，包括多点触摸、体感手势识别、语音识别、重力加速感应和脑电波识别的多种技术交叉应用，最具代表性的应用就是“体感游戏”，与传统的按键控制不同，它主要是基于体感技术实现的利用肢体动作来完成一系列指令，所以也具有很多优势如：加强人类感觉和动作通道的能力，减轻转化的负担、提供更多操作的可能，提高游戏的趣味性和玩家的参与性以及更真实更多维度的感官体验。

除了上述的应用领域以外，人机交互技术还广泛存在于教育领域，如应用了虚拟现实等技术的多媒体教学系统。包括医学领域也有多用途高效率的应用实现，如“VR+”电子病历系统、远程医学教学，“AR+”外科手术导航、微创手术等。

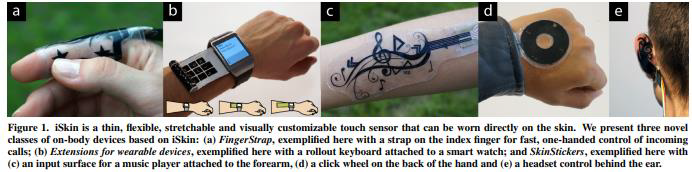
随着信息处理和信息交互工具的不断创新发展，人机交互的领域也在不断地拓宽，用户体验层次也在不断地丰富。相信随着科学技术的进步，人机交互技术能够为人类带来更多的便利，产生更多意想不到的惊喜。

3.前沿研究成果列举：

1. iSkin: Flexible, Stretchable and Visually Customizable On-Body Touch Sensors for Mobile Computing

作者：Martin Weigel，Tong Lu，Gilles Bailly，Antti Oulasvirta，Carmel Majidi，Jürgen Steimle

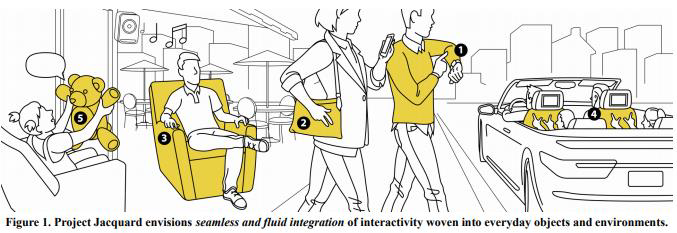
该文中提出了一种可穿戴在皮肤上的传感器iSkin，它使用与人体相兼容的硅胶制作而成，可设计成各种不同的样式和大小。它能够穿戴在人体的不同位置，包括手指、前臂，甚至是耳垂的后方。iSkin的最大特点是薄，柔软并且能折叠弯曲，即使发生了严重的形变也不会影响到它的功能。它内嵌电容式和电阻式传感器，支持多点触控功能，可以识别划动等动作。这项研究使得新型的体内装置成为可能，包括手指穿戴设备、对传统可穿戴设备的扩展以及触摸式智能贴纸。



1. Project Jacquard: Interactive Digital Textiles at Scale

作者：Ivan Poupyrev，Nan-Wei Gong，Shiho Fukuhara，Mustafa Emre Karagozler，Carsten Schwesig，Karen E. Robinson

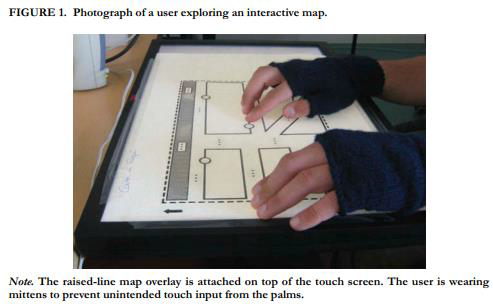
这篇文章介绍了由谷歌发明的神奇布料 Project Jacquard（提花织物），文中提出的这种新颖的交互式纺织布料，利用熟知的纺织原料与导电纤维编成纱线，之后这些具有优越拉伸性能的纱线被织进布料中，整个过程只要稍微改一改普通织布机的程序，就能轻松做到，并不需要额外的工序。为了与智能手机等设备互联，Project Jacquard包含了可传导信号、导电并且可以和各种材质的纱线纺织在一起的特殊纱线。此外嵌入到衣服当中的还有像纽扣一样大小的电路线圈，以及带芯片的独立主板；Project Jacquard可以用来捕捉人体的触摸输入以及手势操作。文中还描述了纱线、纺织品、服装和用户交互性的发展，并提出了可穿戴的交互式布料目前存在的机遇和挑战。



1. Interactivity Improves Usability of Geographic Maps for Visually Impaired People

作者：Anke Brock,Philippe Truillet,Bernard Oriola,Delphine Picard,Christophe Jouffrais

在这篇文章中，作者展示了交互式地图优于带有盲文图例的触觉纸质地图，它涉及三个可用性维度中的两个，并与第三个维度等效：效率，根据学习不同类型空间信息所需的时间来衡量。视觉障碍者使用触觉地图来获取空间的心理表征，但它们存在严重的局限性（信息量有限、盲文文本等）。这项研究提出了一个可用性的经典凸起线地图与互动地图组成的多点触摸屏、凸起线覆盖和音频输出。这两张地图都由24名盲人参与者进行了测试。他们将可用性衡量为效率、有效性和满意度。研究结果显示，交互式地图是改善视力受损人群地图探索和认知地图绘制的一个很好的解决方案。以简单的听觉触觉互动取代盲文，可显著提高效率及使用者的满意度。



李可然 张岩

2020年11月18日

参考文献：

[1] 朱莉. 高可靠电阻式触摸屏的研究与实现[D]. 东南大学, 2015.

[2] 袁保宗, 阮秋琦, 王延江, 刘汝杰, 唐晓芳. 新一代(第四代)人机交互的概念框架特征及关键技术[J]. 电子学报, 2003(S1): 1945-1954.

[3] 人机交互发展历史. https://wenku.baidu.com/view/faeda09d0a1c59eef8c75fbfc77da26924c596ee.html

[4] 郭小爽. 人机交互中的动态手势识别及应用研究[D]. 西安电子科技大学, 2014.

[5] 武霞, 张崎, 许艳旭. 手势识别研究发展现状综述[J]. 电子科技, 2013, 26(06):171-174.

[6] 刘昕. 基于眼动的智能人机交互技术与应用研究[D]. 南京大学, 2019.

[7] 史元春. 信息无障碍中的智能交互技术[J]. 中国计算学会通讯, 2020, 16(3): 8-10. [8]iSkin: Flexible, Stretchable and Visually Customizable On-Body Touch Sensors for Mobile Computing

作者：Martin Weigel，Tong Lu，Gilles Bailly，Antti Oulasvirta，Carmel Majidi，Jürgen Steimle

[9]Project Jacquard: Interactive Digital Textiles at Scale

作者：Ivan Poupyrev，Nan-Wei Gong，Shiho Fukuhara，Mustafa Emre Karagozler，Carsten Schwesig，Karen E. Robinson

1. Interactivity Improves Usability of Geographic Maps for Visually Impaired People

作者：Anke Brock,Philippe Truillet,Bernard Oriola,Delphine Picard,Christophe Jouffrais