

# 读书报告——Going,Going,Gone

## 1.引言

随着科技的迅速发展,虚拟现实技术(VR)已经成为了现实生活中不可或缺的一部分。虚拟现实技术给我们带来了不一样的娱乐体验,其中多人共同探索虚拟世界、多人共同完成一项活动这一主题更是增加了社交成分,但是当前虚拟世界的移动技术还存在一定的问题。

探索虚拟世界中多用户运动的意图交流是本文的中心点,它既介绍了现有一些多人移动技术的不足之处,还通过实验的办法,向我们介绍了三种意图沟通方式,探索在虚拟世界中保持个体活动自由的同时实现多人之间的移动交流。

通过这三种意图沟通方法,给解决VR世界中多人移动技术提供了重要的思路。这对于虚拟现实技术应用具有重要意义,可以优化虚拟世界中多人共同体验的感受,在丰富社交趣味的同时增加团队协作的能力。

## 2.相关工作

### 2.1 单用户定位

单用户定位工作指的是研究和开发用于在虚拟现实环境中让单个用户进行移动和导航的技术和方法。它的工作旨在研究单个用户在虚拟现实环境中的移动技术。单用户定位技术主要包括传送、基于手势、视线和声音的移动等,它选择多样,能够精确地定位用户的移动,但是不足之处便是缺乏身体参与感,用户通过手柄和按键控制,而且有社交交互限制,它是为单个用户设计的,缺乏多人互动。

### 2.2 VR 中的群体定位

群体定位指的是在多个用户之间协调和管理他们在虚拟环境中的位置和移动。旨在促使多个用户在同一虚拟场景中共同移动、交互和协作。群体定位技术可以应用于各种VR场景,包括虚拟会议、团队合作、游戏等。它和单用户定位不同,增强了社交互动性,提高了团队的合作和协作,但是也面临着协调复杂性、资源消耗和可能导致混乱等挑战,这增加了系统的复杂性和开发的难度。

### 2.3 运动的社交方面

虚拟现实中的运动具有丰富的社交层面,涉及到与他人的社交互动、群体协作、社交氛围等方面。运动的社交也促进了VR技术的发展,Weissker等人提出了多用户定向传送技术的概念,为多用户协作创造了良好的交互体验;研究表明,在虚拟环境中进行身体同步运动可以增加用户之间的社交亲近感,合作意愿和信任度。但是运动社交性还存在一些问题,比如虚拟环境中忽略了非语言沟通的重要性,用户的个人隐私空间的考虑,需要更多的解决方案。

## 3.研究方法

在本文的实验中设计了一个虚拟现实迷宫任务,供两名参与者完成。任务务分为个体和集体两个阶段。在个人阶段,参与者需要自由移动,不能有相应的交流,穿过迷宫找到目标点之一。在集体阶段,他们需要协作返回起点。参与者可以口头交流。同时,该实验也相应地设置了自变量和因变量。自变量为不同级别的可

视化水平，这包括了目标、方向和群体；因变量有 IOS、(用来评估对另一名玩家的心理亲近感)GEQ(用来衡量玩家对对方的关注程度)。整个实验内容流程包括分配工作站和头戴式显示器、熟悉控制器分配、完成阶段性任务、填写问卷、手机反馈和得出实验结论多个步骤。

本实验的适应性在于在集体阶段的实验过程中以分别以领导者和跟随者的角度探究了三种意图沟通方法和两种基准方法的效果，这样使得实验结果更具有全面性，同时实验内容的简单有趣性使得该实验可适用于不同人群；然而，该实验也具有局限性，比如在虚拟场景的视觉上没有做到简约可视化、没有解决运动信息的隐私性等问题，这还需要提出相应的解决方法。

## 4.实验结果

在实验研究中发现，对于多人共同完成的 VR 探索项目中，对于跟随玩家来说，方向技术(指导用户在虚拟环境中的定向或移动)比目标技术指向特定目标或对象的定位、选择和操作技术更有助于感知其他玩家的位置并更容易保持在一起。所以在未来多人 VR 发展中可以多注重利用多个目标的相对位置来实现用户交互。

同时，实验结果表明，单用户移动的自由度会对联合运动的自由度产生限制影响。单用户移动中，用户可以轻松地跳转到不同位置，而不必经历真实世界中的移动过程。这可能会导致用户对虚拟环境的空间感知和体验的负面影响，使得团体运动时的沟通、协作和空间感知受到限制。在未来设计 VR 联合运动时，需要考虑单用户移动和团体运动对于用户自由度、空间感知以及沟通和协作的影响。

结合实验结果和现实应用，多用户 VR 技术未来发展将以方向技术为主要的技术用来实现多个用户之间的定位移动，目前的多人技术还是以两位用户为基础，该项技术需要具有可扩展性，这仍然是 VR 技术发展路上的一大挑战。最后，处理技术发展之外，未来的多用户交互中还需要考虑到信息隐私这部分问题。

## 5.我的感悟

本文主要回顾了多人 VR 技术的先前相关工作，分析了当下相应技术的不足与对于方案。同时，本文以一个具体的实验研究来得出未来相应科技的发展趋势。

读完这篇文章，首先我对虚拟现实多人虚拟现实交互的理解有了进一步的认知，多人虚拟现实交互是指多个用户能够在虚拟现实环境中实时互动、共享体验的技术和应用。这篇文章作者通过实验得出了不同技术对用户体验的影响，同时我们可以从中得出相应的结论和总结出一些建议。

在我看来，在多人虚拟现实交互中，多用户之间交互的体验是很重要的。但是实际上单用户移动对联合运动会产生限制影响，所以我们需要综合考虑单用户移动和团体运动对于用户自由度、空间感知以及沟通和协作的影响；其次，每个用户之间的协调和沟通是非常重要的，所以在未来利用更加高效的方向技术可以更容易地指引玩家找到位置或者是导航。

总而言之，这篇论文让我对虚拟现实多人用户运动意图传达的现状和方法有了更深入的了解，同时也启发了我对存在问题和优化策略的一定思考。

# 调研报告——从视听探究 AI 技术在 HCI 领域中的运用

## 1.引言

人机交互技术是人类与计算机系统之间进行信息交换与互动的过程中的技术手段，先进的人机交互技术可以使人们能够更加自然、方便地操作计算机。人机交互技术可以体现在硬件方面，比如鼠标、手柄、触摸屏等等，还可以体现在软件方面，比如合理的图形用户界面、语音播放等。人机交互技术的发展源远流长，其在 1975 年就被最早提出，在 80 年代开始蓬勃发展，在这个发展过程中，人们制定了设计人机交互界面的多项原则，并且随着时代科技的发展，人机交互技术将这些最新的科技技术融合起来，使得人机交互技术不断进化。

近几年来，随着人工智能异军突起，市面上出现了许多令人惊奇的人工智能模型，这些模型有着强大的功能，比如自然语言处理、结果预测和图像生成等等，这些功能也同样被人机交互技术所吸收，出现了许多新奇的产品。

## 2.相关工作和背景概述

在传统的观念中，人机交互需要的是一个实体介质，比如键盘，屏幕等等，这需要人们对计算机的 IO 部分进行触碰，将这些操作转为电信号后由计算机进行反映，但是由人工智能带来的语音识别技术让人们可以通过说话的方式控制计算机或者其他电子设备。

计算机视觉和计算机听觉是 AI 技术在 HCI 领域中的两个重要方面，它们分别涉及计算机对视觉和听觉信息的处理和理解，以实现更智能、更自然的人机交互体验。计算机视觉涉及计算机对图像、视频等视觉信息的理解和处理。在虚拟现实(VR)和增强现实(AR)系统中，计算机视觉技术可用于实时跟踪用户的头部运动和手势，从而实现更加沉浸式和自然的交互体验。另外，计算机视觉还可用于人脸识别技术，实现用户身份验证和个性化服务；计算机听觉涉及计算机对声音、语音等听觉信息的处理和理解。语音助手(如 Siri、Alexa、Google Assistant 等)利用计算机听觉技术实现了语音识别和自然语言理解，使得用户能够通过语音与计算机进行交互。下面我们来仔细探究视听方面 AI 在 HCI 中的应用。

## 3.具体介绍和专家观点

图像识别与分析是人工智能在 HCI 中被大家所熟知的一点，从最简单的手写数字识别，到人脸识别，再到人体的体态与手势识别。这些功能在日常生活中运用极其广泛，门禁、运动员的动作分析等等，为生活增加了许多便利。手势识别技术在近年来得到了广泛的应用，并且在各个领域都有着不断涌现的新应用。AI 听觉方面的 HCI 涵盖了语音识别、自然语言处理、情感识别、声音合成、智能音

频分析等多个方面，这些技术的应用使得人机交互更加智能、自然和个性化，为用户提供了更加丰富和便捷的交互体验。

如果我们将语音识别与图像识别，手势识别结合起来，那有会怎么样？在 HCI 领域中，计算机视觉和计算机听觉的结合可以实现更加全面和智能的人机交互体验。通过综合利用视觉和听觉信息，计算机可以更准确地理解用户的意图和需求，从而提供更加个性化、自然和高效的交互服务。

在最近，苹果公司推出了首个空间计算设备——Apple Vision Pro，这是一台革命性的空间计算设备，将数字内容无缝融入真实世界，让用户处在当下并与他人保持连接。Apple Vision Pro 打造无边际画布，让 app 突破传统显示屏的限制，为用户带来全新的 3D 交互体验，以最自然、最直观的输入方式来控制眼睛、双手与语音。这是一台集合计算机视觉、计算机听觉为一体的人机交互设备机器，为每个用户带来身临其境的即时体验。

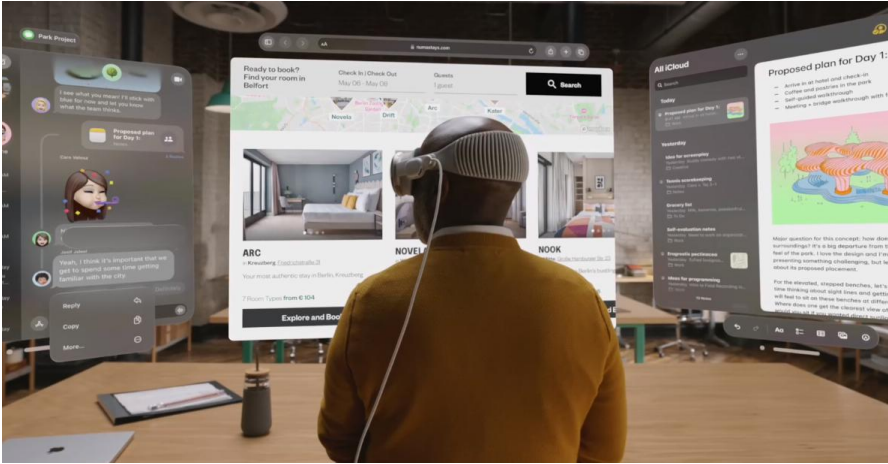


图 1 结合视觉和听觉为一体的 Apple Vision Pro

这是因为苹果公司在计算机视觉和计算机听觉方面一直处于前沿，才能推出多项技术和功能，如今推出了 Apple Vision Pro。让我分别拿苹果公司计算机视觉和计算机听觉两种技术用来举例，它们分别是苹果公司计算机视觉中实景增强(AR)技术中的 ARKit 框架和苹果公司计算机听觉中的自然语言处理系统。

ARKit 是苹果公司推出的增强现实(AR)开发框架，旨在让开发者能够在 iOS 设备上创建高性能的增强现实体验。ARKit 技术的原理和大多数 AR 技术的原理大致相同，首先 ARKit 使用设备的摄像头、运动传感器和深度传感器来获取环境的实时数据。这些传感器提供了设备位置、方向、移动速度以及周围环境的深度信息等数据；再次通过分析传感器数据，识别环境中的平面、物体和人体等场景元素，并构建出环境的 3D 空间模型；接着根据环境的理解和用户的交互，将虚拟对象叠加在真实世界中的适当位置。

ARKit 框架能够实时分析环境中的平面、物体和人体等元素，并构建出环境的 3D 空间模型、提供了强大的虚拟对象跟踪和稳定性功能，能够准确地跟踪设备的位置和方向，并保持虚拟对象与真实世界的稳定对齐、实时获取环境的深度信息，进一步提高了虚拟对象的放置和交互的准确性，为 Apple Vision Pro 的研发打下了坚实的基础。

	ARCore	ARKit	Magic Leap	HoloLens
Device tracking	✓	✓	✓	✓
Plane tracking	✓	✓	✓	
Point clouds	✓	✓		
Anchors	✓	✓	✓	✓
Light estimation	✓	✓		
Environment probes	✓	✓		
Face tracking	✓	✓		
2D Image tracking	✓	✓	✓	
3D Object tracking		✓		
Meshing		✓	✓	✓
2D & 3D body tracking		✓		
Collaborative participants		✓		
Human segmentation and occlusion		✓		
Raycast	✓	✓	✓	

图 2 ARKit 在多功能上的优势

上文所说苹果公司在计算机听觉处理上也有着强大的技术系统。计算机听觉的处理方面，需要利用多种技术和方法，包括声学模型、语言模型、声学特征提取、语言特征提取等。在苹果公司开发出的 SIRI 智能手机的语音助手是其在计算机听觉技术的基础，它所搭载的深度学习框架网络用来训练大规模的语音识别模型，以提高准确性和性能。

在语音识别任务中，循环神经网络(RNN)和长短期记忆网络(LSTM)是两种常见的深度学习模型，用于处理序列数据。RNN 接受连续的语音信号作为输入，并逐步输出对应的文本或语音命令。但是传统的 RNN 存在梯度消失和梯度爆炸等问题，导致对长序列数据的建模效果不佳。在语音识别中，LSTM 可以更好地捕捉语音信号中的长期依赖关系，从而提高识别的准确性和性能。通过引入记忆单元和门控机制，LSTM 可以有效地处理长序列数据，避免了梯度消失和梯度爆炸的问题。

如下图，我们利用所训练的两个模型来对一段相同的语言进行识别，结果我们可以看见，LSTM 所带来的识别效果随着时间的推移准确率会比 RNN 所带来的识别效果好。同样的，虽然苹果公司 SIRI 背后所用的训练神经网络没有被完全地公开，但是毋庸置疑的是强大的神经网络训练系统使得语音识别在用户交互的准

确性上有极大的提升。

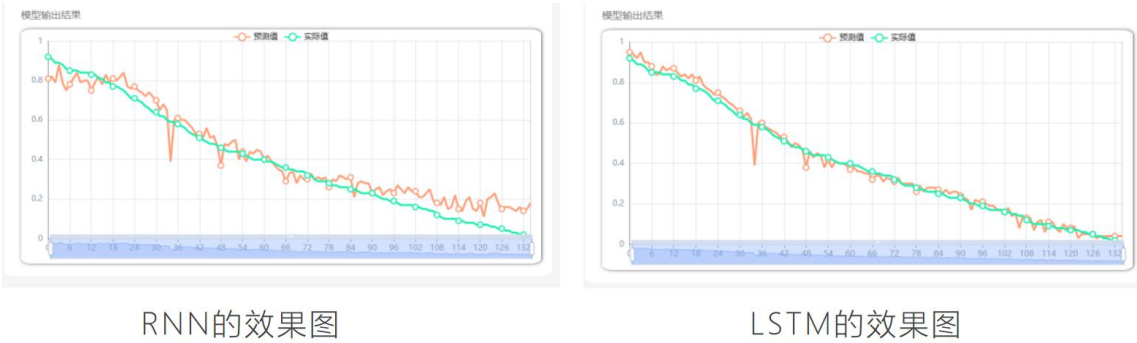


图3 在语音识别中两种神经网络的对比

当然，Apple Vision Pro 在快速火热的同时，也引来了不少专家对其进行测评。国内有三家公司首要参加了 Apple Vision Pro 的初体验。根据这三家科技产品体验媒体人的真机体验，我们发现三家媒体均认为 Vision Pro 具有“改变认知”级别的超震撼交互体验，愿意进一步体验或成为第一批用户。大家都认为 Vision Pro 属于划时代的产品，给 VR/AR/MR 行业带来了新的行业标准，并异口同声的表示：“未来可期”。数码测评师钟文泽也表示 Vision Pro 在硬件层面可以在 4K 的显示效果下完成图片渲染，在软件层面搭载了众多 APP，可以给用户以完整的交互体验。

#### 4.总结与个人体会

人工智能在人机交互技术中已经被广泛应用，或许能够为人类再次带来一场伟大的工业革命。在这篇调研报告中我们主要把注意力放在人工智能所带来的计算机视觉和计算机听觉上，以苹果公司推出的新产品 Apple Vision Pro 为例子，分析了整个最新的用户交互技术产品背后所搭载的技术。

不管是 Apple Vision Pro 还是别的用户交互技术产品，我们都可以发现，人工智能是一项极为高端的技术，对它的直接操控和调整优化是由技术员执行的，只有通过人机交互技术的人工智能才能真正地被大众所使用。像 Apple Vision Pro 产品，它背后的技术网络是相当复杂的，但是表层所给用户带来的体验方式却是简单的，正如 Richard H. R. Harper 在他的文章《The Role of HCI in the Age of AI》指出，未来不仅仅是人工智能，而是必须由人机交互技术所搭载的人工智能。人工智能给带来了理论基础，人机交互技术与人工智能的结合才是未来发展的趋势。

本篇调研的主要方向还是 AI 技术中视觉、听觉方面在 HCI 中的应用。从上文的举例分析中不难看出，计算机视觉、听觉可能还存在着一些问题。在技术方面，可能还对环境的理解的方法上还需要改进，目前的计算机视觉系统往往只能对图像进行浅层的特征提取和物体检测；同时在个人问题上，人机交互技术也需要关注用户数据的隐私保护和安全性。这些都是未来的努力方向。