

# 引入智能虚拟讲解员的云展厅构建与交互反馈研究\*

刘舫<sup>1</sup>, 吕天<sup>1</sup>, 刘心阁<sup>1</sup>, 叶盛<sup>1</sup>, 郭锐<sup>1</sup>, 张烈<sup>2</sup>, 马翠霞<sup>3</sup>, 王庆伟<sup>4</sup>, 刘永进<sup>1</sup>

<sup>1</sup>(普适计算教育部重点实验室, 清华大学计算机科学与技术系, 北京 100084)

<sup>2</sup>(清华大学美术学院交互媒体研究所, 北京 100084)

<sup>3</sup>(计算机科学国家重点实验室, 人机交互北京市重点实验室, 中国科学院软件研究所, 中国科学院大学, 北京 100190)

<sup>4</sup>(首都体育学院, 管理与传播学院, 北京 100091)

通讯作者: 刘永进, E-mail: liuyongjin@tsinghua.edu.cn

**摘要:** 奥林匹克遗产是全人类的宝贵财富, 利用科技与文化艺术相融合, 对实现北京冬奥会及奥林匹克遗产和精神文化的多元呈现和高效传播至关重要。线上云展厅作为信息时代数字博物馆发展的重要趋势, 虽然在单项数字博物馆与交互系统技术研究方面已具备较好的基础, 但尚未形成体系化、智能化、交互友好的冬奥会奥林匹克数字博物馆系统。本文面向北京 2022 年冬季奥林匹克运动会, 提出了一种具有交互反馈的云展厅系统构建方式, 通过构建具有虚拟讲解员的智能交互云展厅, 进一步探索了交互反馈在知识传播型的数字博物馆中对于非物质文化遗产的传播作用。为了探索视听交互反馈对云展厅的奥林匹克精神文化传播的影响, 同时提升知识传播型云展厅中的用户体验, 我们对 32 名参与者进行了用户实验。研究表明, 本文构建的云展厅能够极大地促进奥林匹克文化与精神的传播; 同时, 在云展厅中引入视听交互反馈能够提高用户的感知控制, 从而有效地提升用户体验。

**关键词:** 奥林匹克云展厅; 交互反馈; 智能虚拟讲解员; 感知控制

**中图法分类号:** TP399

中文引用格式: 刘舫, 吕天, 刘心阁, 叶盛, 郭锐, 张烈, 马翠霞, 王庆伟, 刘永进. 引入智能虚拟讲解员的云展厅构建与交互反馈研究. 软件学报, 2022. <http://www.jos.org.cn/1000-9825/0000.htm>

英文引用格式: Liu F, Lv T, L XG, Guo R, Zhang Lie, Ma CX, Wang QW, Liu YJ. Research on the Construction and Interactive Feedback of the Exhibition-hall with the Introduction of Intelligent Virtual Agent. Ruan Jian Xue Bao/Journal of Software, 2022 (in Chinese). <http://www.jos.org.cn/1000-9825/0000.htm>

## Research on the Construction and Interactive Feedback of the Exhibition-hall with the Introduction of Intelligent Virtual Agent

Liu Fang<sup>1</sup>, Lv Tian<sup>1</sup>, Liu Xin-Ge<sup>1</sup>, Ye Sheng<sup>1</sup>, Guo Rui<sup>1</sup>, Zhang Lie<sup>2</sup>, Ma Cui-Xia<sup>3</sup>, Wang Qing-Wei<sup>4</sup>, Liu Yong-Jin<sup>1</sup>

<sup>1</sup>(Department of Computer Science and Technology, MOE-Key Laboratory of Pervasive Computing, Tsinghua University, Beijing 100084, China)

<sup>2</sup>(Interactive Media Institute, Academy of Arts&Design, Tsinghua University, Beijing 100084, China)

<sup>3</sup>(State Key Laboratory of Computer Science and Beijing Key Lab of Human-Computer Interaction, Institute of Software Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing 100190, China)

<sup>4</sup>(Capital University of Physical Education and Sports, School of Sports Management and Communication, Beijing 100091, China)

**Abstract:** The Olympic heritage is the precious wealth of the world. The integration of technology, culture and art is crucial to the realization of diversified presentation and efficient dissemination of the heritage of the Beijing Winter Olympics. As an important form of digital museum, online exhibition-halls have a good foundation in the research of interactive technologies, but as far as we know, there is no systematic, intelligent, interactive and friendly Winter Olympics digital museum has been formed. We present an online exhibition-hall

\* 基金项目: 国家重点研发计划项目《奥林匹克云展厅智能交互系统关键技术研发》(2021YFF0307702)资助

收稿时间: 2022-05; 修改时间: 2022-09-07; 采用时间: xxxx-xx-xx

construction method with interactive feedback for the Beijing 2022 Winter Olympic Games. By constructing an interactive exhibition-hall with an intelligent virtual agent, we have further explored the role of interactive feedback in the dissemination of intangible cultural heritage in a knowledge dissemination digital museum. In order to explore the influence of audio-visual interactive feedback on the Olympic spirit culture dissemination in the exhibition hall, and at the same time improve the user experience, we conducted a user experiment with 32 participants. The results show that our exhibition hall constructed in this paper can greatly promote the dissemination of Olympic culture and spirit, and at the same time, the introduction of audio-visual interactive feedback in the exhibition hall can improve the user's perceptual control, thereby effectively improving the user experience.

**Key words:** Beijing Winter Olympics exhibition hall; interactive feedback; intelligent virtual agent; perceived control

## 1 引言

数字博物馆通过对文化遗产进行数字化采集并以多媒体的方式进行展示,可以为参观者提供丰富的学习与传承文化遗产知识的机会,目前已成为非物质文化遗产传播的重要途径<sup>[1,2]</sup>。在过去的几十年里,通信技术与人机交互技术的发展促使了各式各样的数字博物馆与线上展厅的出现,包括网站、手机 APP 应用、微信公众号和小程序等形式。数字博物馆与线上展厅的构建主要包括对展品内容的数字化采集、整理、组织、存储、展示、传播等环节<sup>[3]</sup>。中国国家博物馆王春法馆长指出,基于多媒体技术、数字显示技术和人机交互等新型科技技术优化线上展厅的展示与传播是智慧博物馆建设的重要内容<sup>[4]</sup>。数字博物馆的展品形态丰富,包括文本、图像、音视频、三维模型、动画等。在数字博物馆的智慧展示方面,已有相关研究利用多模态交互、虚拟现实(Virtual Reality, VR)、增强现实(Augmented Reality, AR)、三维重建与显示等人机交互与可视化技术改善用户的观展体验。一方面,数字博物馆的发展呈现出展品内容、交互方式与观展设备的多样性、丰富性趋势,虽然新型交互方式能够显著改善用户的观展体验,目前仍然缺乏对数字博物馆中视听交互反馈与用户体验之间的相互影响机制的研究。另一方面,对于知识传播型的数字博物馆,其构建目标归根结底是进行文化和精神的传播与教育<sup>[5]</sup>,线上展厅是知识存储、展示和传递的重要途径<sup>[6]</sup>。因此,如何在保障数字博物馆文化遗产传播功能的前提下,利用新型交互技术改善用户的观展体验是一个至关重要的问题。

中国首都北京在继 2008 年举办夏季奥林匹克运动会后,在 2022 年再次成功举办冬季奥林匹克运动会,成为世界上第一个“双奥之城”。国际奥委会以及历届主办城市都高度重视奥运遗产的保护与宣传工作,北京冬奥组委专门制定了《北京 2022 年冬奥会和冬残奥会遗产战略计划》,通过精心规划收集了丰富的冬奥和冬残奥遗产成果。共青团中央与北京冬奥组委发布《北京 2022 年冬奥会和冬残奥会青少年行动计划》,动员青少年弘扬奥运精神、宣传冬奥文化。在北京冬奥会奥林匹克文化在线宣传方面,北京 2022 年冬奥会和冬残奥会组织委员会官方网站<sup>[7]</sup>开设了“教育”模块,以“主题活动知识课堂”的形式展示了冬奥知识和奥林匹克文化相关的图像、视频与文本。周立超等人<sup>[8]</sup>在北京市电厂路小学依托“小小冬奥博物馆”,通过展示冬奥相关的学生创意作品来传播奥林匹克教育。诸如此类的一系列活动,虽然产生了积极的影响,但直到本文所构建的云展厅之前,尚未出现专门针对北京 2022 年冬奥会的体系化、智能化、交互界面友好的智能数字博物馆。开发具备多模态交互、面向多场景、覆盖多人群的冬奥会奥林匹克云展厅,在 2022 冬奥会赛期与赛后积极开展并促进北京冬奥会奥林匹克知识传播,有助于弘扬奥林匹克精神和展示中国奥运贡献。

本文的目标是利用多媒体线上布展技术与交互反馈技术相结合的方式,为北京 2022 年冬季奥林匹克运动会构建一个内容丰富的智能交互云展厅,有助于加强对奥林匹克文化与精神的展示与传播。如图 1 所示,(a)为云展厅首页,(b)为云展厅导览页,用户可选择进入相应展廊,(c)(d)为“项目”主题展厅下的“冰壶”展廊内展示页面。视觉与听觉是人机交互中最重要的两个通道,本文通过引入一个具有视听交互反馈功能的智能虚拟讲解员来实现奥林匹克云展厅的多模态用户交互与反馈(如图 1(b)、(c)中右下角所示)。在视觉反馈方面,研究基于个性化虚拟说话人生成的虚拟讲解员角色快速构建技术与基于语音驱动的讲解说话动画渲染技术,实现使用用户上传的任意单张正面人脸照片在线生成虚拟讲解员角色的功能,支持用户自主地控制智能虚拟讲解员的人脸外貌、面部表情等视觉特征。针对虚拟讲解员形象生成,首先输入一张人脸照片,采用人脸三维形变统计模型(3D Morphable Model, 3DMM 模型),通过设计损失函数预测人脸的参

数化 BFM (Basel Face Model) 系数; 然后通过云展厅预置的光照、人脸姿势、人脸纹理等配置, 结合上述 BFM 系数在云展厅中生成符合用户需求的虚拟形象。在听觉反馈方面, 设计虚拟讲解员的讲解与问答两种对话反馈模式, 研究基于冬奥会特定领域下结构化知识图谱建立的智能问答技术, 考察不同听觉反馈模式对云展厅的知识传播能力以及用户体验的影响机制。具体地, 系统首先将用户的语音输入转化为文本句子进行解析与实体匹配, 基于本文构建的知识图谱系统检索并生成问题的文本答案; 然后利用文字转语音 (TextToSpeech, TTS) 技术将文本的答案转化成语音序列, 并通过提取语音序列中的音素推理得到人脸表情参数的变化序列; 最后, 基于该变化序列, 利用预设的光照等配置对虚拟形象进行渲染得到逼真的虚拟形象说话讲解的效果。云展厅中的虚拟讲解员不仅具有逼真的人脸说话与交互效果, 而且能够实时交互。

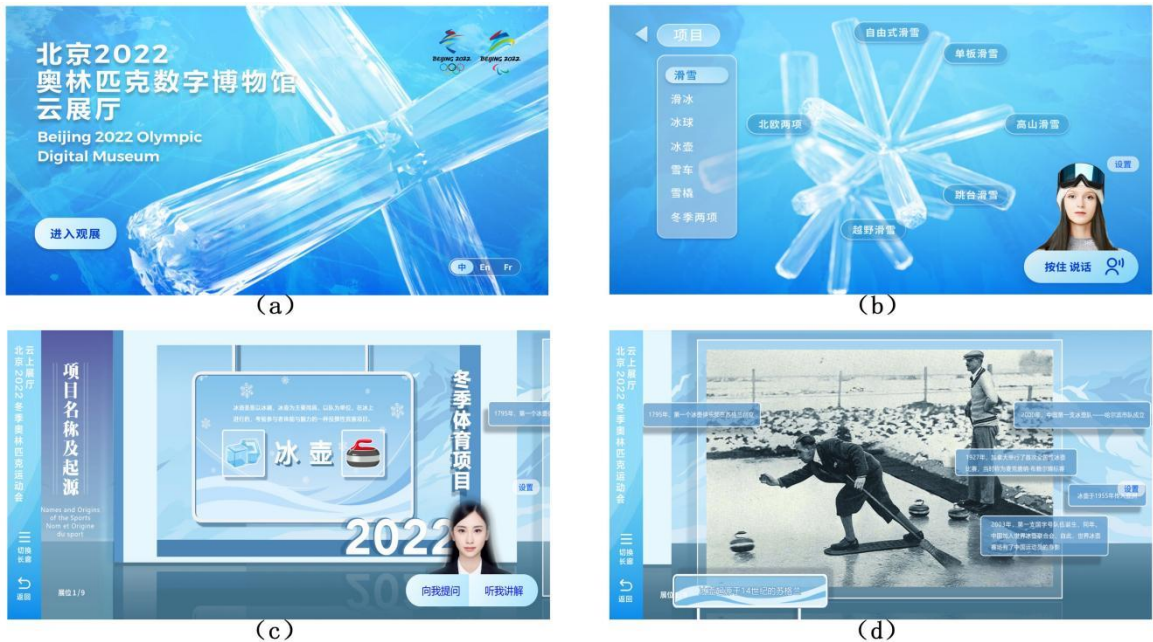


图1 OlympicHall2022 奥林匹克云展厅。(a): 云展厅首页; (b): 云展厅导览页; (c-d): “冰壶”展厅页面。云展厅更多细节请见在线网站 <https://exhibition-hall.cn/>。

近年来, 在交互设计中越来越多地涉及到对感知控制 (perceived control, 或 sense of control) 因素的考量, 包括界面设计<sup>[9]</sup>、App 研发<sup>[10]</sup>、残疾人设施<sup>[40]</sup>和社交机器人<sup>[11]</sup>等。感知控制指用户相信自身可以决定自己的内部状态和行为、能够影响自身的环境和/或带来预期结果的信念<sup>[11]</sup>。在交互领域对感知控制的研究主要包括探索其对交互系统性能与用户体验的影响。进一步地, 本文从探索智能交互云展厅中视听反馈与用户的感知控制之间的关系出发, 探讨在数字博物馆中的交互反馈与用户体验之间的关系, 从而更好地为知识传播型数字博物馆的设计提出建议。本文基于构建的北京 2022 冬奥会云展厅 (以下简称云展厅) 搭建了三个基线测试系统, 并邀请了 32 名参与者进行用户实验研究。通过分析参与者的实验行为与问卷调查结果, 本文得到的主要结论如下:

- **交互反馈对知识传播的影响:** 与不具有智能虚拟讲解员的对照组实验系统相比, 参与者浏览云展厅之后能够掌握更多的冬奥知识, 在数字博物馆中引入视听交互反馈能够有效地提高系统的知识传播能力。
- **感知控制:** 与对照组系统相比, 云展厅能够在视觉与听觉上提供即时的、准确的反馈, 参与者在浏览云展厅时具有更强的感知控制。在数字博物馆中引入交互反馈能够增强用户的感知控制, 从而提供更好的浏览体验。
- **用户的整体评价:** 参与者对云展厅的用户体验各指标评价均较高, 云展厅在“参与意向”指标上得分显

著高于对照组系统,表明参与者更愿意参与到具有智能虚拟讲解员的系统中。

## 2 相关工作

### 2.1 奥林匹克与数字博物馆

数字博物馆是展示奥林匹克历史、文化与精神的高效途径,其形式包括网站、虚拟现实系统等。国际奥委会奥林匹克研究中心长期以来对世界奥林匹克知识、奥林匹克运动最新信息、国际奥委会档案、官方出版物、期刊藏品等进行了收集与整理,建立了世界奥林匹克图书馆,包括3万多本实体出版物与1万多个电子文档<sup>[12]</sup>。Kubo等人<sup>[13]</sup>对举办奥运会的社会及经济影响进行了系统的分析,并建立了一个在线虚拟博物馆<sup>[14]</sup>展示了1896-2028年间夏季奥运会的收藏品、纪念品、稀物品、重要事迹和信息。He等人<sup>[15]</sup>从奥林匹克数字博物馆的运动、人类和娱乐/教育三大主题出发,创建了一个虚拟的数字奥运博物馆(VDOM),通过解决面向奥运任务的数据存储和检索、数字博物馆的建模和渲染、用虚拟人进行运动的虚拟演示和基于虚拟现实的运动模拟等问题,扩展了传统体育博物馆的主要功能。Pan等人<sup>[16]</sup>基于动画与跨媒体信息检索技术构建了一个虚拟现实奥林匹克数字博物馆,并通过引入虚拟形象实现博物馆导览功能。进一步地,他们开发了能够应用于奥林匹克数字博物馆的VR运动模拟系统,由虚拟保龄球、虚拟乒乓球和虚拟网络马拉松组成<sup>[17]</sup>。Qi等人<sup>[18]</sup>提出一种重建、压缩和保护真实展品3D模型的方法,实现了基于图像的3D建模、3D数据编解码器和3D水印生成技术,并应用于数字奥林匹克博物馆。尽管已有少量研究建立了奥林匹克运动相关的数字虚拟展厅,但目前奥林匹克数字博物馆大多局限于单点的运动项目或单方面的介绍,且不涉及探索云展厅中交互反馈对云展厅知识传播能力及用户体验的影响,尚未出现专门针对冬奥会历史、文化与精神展示的内容丰富的智能交互云展厅。

### 2.2 交互系统与反馈机制

随着人机交互和人工智能的发展,多模态交互技术在多个交互情境中都得到了广泛的应用,包括可穿戴设备<sup>[19,20,21]</sup>、触觉交互<sup>[22]</sup>、虚拟现实(VR)交互<sup>[23,24]</sup>等。反馈作为人机交互中的重要环节,其形式随交互任务与情境的变化而改变。Liao等人<sup>[25]</sup>在笔式交互系统中整合了视、听、触等多模态反馈模式,并将反馈划分为三个级别,即基于发现的反馈、基于状态的反馈与基于任务的反馈。其它的基于多模式反馈的笔式交互系统包括RealPen<sup>[26]</sup>、Pen+touch<sup>[27]</sup>、Conte<sup>[28]</sup>等。Srinivasan等人<sup>[29]</sup>提出了InChorus,探讨不同平板电脑中数据可视化的多模态交互与反馈设计方法。Zhao等人<sup>[30]</sup>研究了虚拟现实设备中基于语言和非语言行为的多模态反馈设计。Freeman等人<sup>[31]</sup>应用多模态反馈优化空中手势系统任务。

多模态交互与反馈技术也广泛地应用于数字博物馆。Masque等人<sup>[32]</sup>通过在头戴式显示器(HMD)内集成皮肤拉伸模块,为用户提供二维面部皮肤拉伸反馈,并应用于VR博物馆导览中。Sidenmark等人<sup>[33]</sup>在虚拟现实博物馆中引入头眼交互,为用户提供更好控制的、更灵活的注视点快速选择方法。Clarke等人<sup>[34]</sup>基于交通博物馆的用户体验进行研究,提出了功能性、时间性、物理性和间接性四种展览设计需要考虑的因素。NetLogo<sup>[35]</sup>是一个支持博物馆互动游戏的多点触控系统,被用于探索面向知识传播型博物馆的交互与协作机制。Hamilton等人<sup>[36]</sup>为在线学习提出了一种协作直播媒体策展的新媒体系统,并集成了包括流式视频、屏幕共享、音频和文本聊天等多模态直播媒体,用于知识共享与传播。

多模态交互与反馈技术越来越多的应用于包括数字博物馆在内的各种交互系统,然而,目前仍然缺乏专门针对在知识传播型博物馆中引用交互反馈的影响的研究,本文在奥林匹克云展厅中引入视听反馈,并探究视听反馈对云展厅用户体验的提升效果。

### 2.3 感知控制与用户体验

近年来,在人机交互系统设计中越来越重视对感知控制因素的考量,包括界面设计<sup>[37]</sup>、App研发<sup>[38]</sup>、残疾人设施<sup>[39]</sup>、社交机器人<sup>[40]</sup>等方面。感知控制反映个人对其所从事的行为进行控制的感知程度,早期应用于用户健康<sup>[41]</sup>、压力处理<sup>[42]</sup>等领域,例如,Skinner等人<sup>[43]</sup>通过个体差异实验分析证明儿童的感知控制和他们

的学业成绩直接相关。Roy 等人<sup>[44]</sup>研究了人工智能领域中的感知控制和自动化性能之间的关系，发现相比于依靠机器自动化，一定程度的控制感更能提升用户的满意度。

在交互领域对感知控制的研究主要涉及探索其对交互系统性能与用户体验的影响。Painting with Bob<sup>[45]</sup>是一个为新手画师设计的绘画创作系统，表明用户的感知控制与系统的灵活性、创造性有着强烈的关系。Lyngs 等人<sup>[46]</sup>调查了 Facebook 用户界面对用户行为和用户感知控制的影响，发现可以通过一定的任务干预提高用户的控制感。GreenByMe<sup>[47]</sup>是一个旨在提供个性化服务的手机应用，其用户研究表明个性化设计能够影响用户的感知控制体验。Jhaver 等人<sup>[48]</sup>讨论了 Airbnb 房东在算法评估中的感知控制体验，发现感知控制的缺乏可能会导致房东产生焦虑。Sundar 等人<sup>[49]</sup>探索定制性、交互性和导航性等通信技术能否提升博物馆的用户体验，设计了一套详细的测量用户感知控制程度的问卷调查量表。

本文从数字博物馆的用户界面设计出发，研究感知控制与用户体验的影响机制，具体探讨在云展厅内引入智能虚拟讲解员的视听反馈如何影响用户的感知控制感，从而影响奥林匹克云展厅的知识传播能力与用户体验。

## 2.4 虚拟说话人人脸动画生成及应用

数字化展厅中的虚拟人包括传统的只有说话音频的讲解员、卡通式虚拟讲解、人形讲解助手、二维虚拟说话人、三维虚拟说话人等各种形式，能够在展览空间内对用户体验产生积极的影响<sup>[50]</sup>。Duguleană 等人<sup>[51]</sup>创建了一个能够与用户对话的智能虚拟代理 (agent)，以此提高历史博物馆内的信息可访问性。Ye 等人<sup>[52]</sup>研究虚拟现实导航领域的虚拟代理，通过优化虚拟代理的位置和方向，控制虚拟代理运动，并将该技术应用于虚拟博物馆、工厂和学校体育馆等各种场景。Andrea 等人<sup>[53]</sup>探索在虚拟博物馆中，用户更喜欢由虚拟指南还是虚拟代理陪同，通过实验表明将两种方式结合能达到更高的用户可接受度。本文在云展厅内引入的虚拟讲解员，重点涉及虚拟人的人脸构建与说话动画生成技术。

虚拟说话人生成通常是基于语音或视频帧序列驱动，生成特定的人脸三维动画说话视频。本文研究云展厅内的虚拟讲解员人脸动画的生成方法，基于用户选定或者上传的一张带有人脸的照片生成三维人脸模型，并基于语音生成说话动画视频。现有语音驱动的虚拟说话人生成方法大多基于深度学习框架学习人脸视觉与音频的联合特征表达，并基于卷积神经网络模型生成视频<sup>[54, 55]</sup>。Chen 等人<sup>[56]</sup>设计了一种级联生成式对抗神经网络 (GAN) 的方法来生成面部的说话视频，与直接学习从音频到视频帧的映射相比，Chen 等人首先将音频转换为高层次的结构特征，然后再生成视频帧，该方法对不同的面部形状、视角、面部特征和嘈杂的音频条件具有鲁棒性。Song 等人<sup>[57]</sup>重点关注整个生成视频上保持嘴唇和面部运动的平滑过渡，提出一种基于循环对抗神经网络的虚拟说话人生成方法。Zhang 等人<sup>[58]</sup>研究单样本 (One-shot) 的说话人脸生成方法，构建了一个大型高分辨率视听数据集，并提出了一种光流引导的说话人脸生成框架，利用面部可变形模型将框架拆分为两个级联模块，能够合成较高分辨率的视频。其它虚拟说话人人脸动画生成的研究方向包括姿态可控的生成方法<sup>[59]</sup>、说话人表情可定制的生成方法<sup>[60]</sup>、多模态生成方法<sup>[61]</sup>等。

## 3 智能虚拟讲解员与反馈设计

云展厅的目标是为用户提供一个内容丰富、引人入胜的北京 2022 冬奥会数字博物馆。在云展厅系统设计与构建方面，通过引入智能虚拟讲解员，实现视听交互反馈，进一步提升云展厅的知识传播能力与用户的观展体验。

### 3.1 面向虚拟讲解员交互反馈的人脸动画生成

云展厅中采用经典的人脸三维形变统计模型 (3DMM 模型) 和深度学习算法相结合的方式生成人脸模型，即输入一张人脸照片，进行三维重建得到正确的形象。记人脸照片为  $I$ ，输入到一个 ResNet<sup>[62]</sup> 结构的卷积神经网络之中，预测人脸的参数化 BFM 系数，包括人脸的身份参数  $\alpha \in R^{80}$ 、表情参数  $\beta \in R^{64}$ 、纹理参数  $\gamma \in R^{80}$ 、光照参数  $\delta \in R^9$  以及人脸的姿态参数  $p \in R^6$ 。记神经网络模型为  $f(\cdot)$ ，则该模型的作用形式化地表示为：

$$(\alpha, \beta, \gamma, \delta, p) = f(I) \quad (1)$$

由于具有人脸 BFM 系数标注的数据集较少,本文的人脸重建方法受到了 Deng 等人<sup>[63]</sup>的启发,将神经网络模型采取无监督训练的方式,基于已有的大量无标注人脸图片学习人脸网格参数的预测:将模型预测得到的部分参数,利用可微渲染得到重新渲染的人脸,并将该结果贴回原图得到图像  $I'$ ,通过比较  $I$  与  $I'$  并利用多种损失函数,从而驱动网络模型中参数的更新。

在训练过程中,预测人脸 BFM 参数的网络在 CelebA-HQ<sup>[64]</sup>、FFHQ<sup>[65]</sup>等数据集上进行训练,从而保证模型在各类年龄、性别、光照、姿势条件下均有较好的表现。本文主要使用了像素级图像损失函数、关键点级损失函数、特征层级损失函数、系数约束损失以及纹理平滑损失。

像素级图像损失函数专注于原图与渲染后图片面部区域的相似性比较,可以通过直接比较渲染后的图片与输入原图的差异得到:

$$L_p = \frac{\sum_{i \in F} M_i \cdot \|I_i - f(I)_i\|_2}{\sum_{i \in F} M_i} \quad (2)$$

其中  $i$  代表像素的标号,  $F$  为重建后人脸投影到二维照片对应的区域,  $I$ 、 $f(I)$  分别为原图、原图经过神经网络后再进行可微渲染后得到的图像。 $M_i$  为一个经过预训练的贝叶斯分类器对原图进行预测的注意力掩模,其功能是预测  $I$  中每一个像素  $i$  位于人面部的概率。

关键点级损失函数能够控制人脸轮廓、形状的一致性,具体可以将重建得到的三维人脸网格中的关键点提取出来,并投影到图像平面上,并与原图中的对应关键点求欧氏距离,作为损失函数的值:

$$L_l = \frac{1}{N} \sum_{i=1}^N w_i \left\| l(I)_i - l(f(I))_i \right\|_2 \quad (3)$$

其中  $N$  是关键点的个数,  $l$  为提取关键点的函数,  $i$ 、 $w_i$  分别为关键点的标号及其对应的权重。该权重的设置是为了使得某些关键点,例如嘴唇部分、鼻子部分等具有更显著的影响。这是因为人们关注面部不同区域的程度不同,例如嘴唇发生细微的变化人们便能感受到,而面部边缘轮廓有细微变化则较难感知。

为了保证模型不会落入局部最优点,并且希望能够从原图、重新渲染的图片的特征层级进行差异的对比,模型还引入了特征层级损失函数,通过使用一个预训练的人脸编码神经网络,将原图与重新渲染的图片进行特征编码,并比较特征编码的差异:

$$L_f(I) = 1 - \frac{g(I) \cdot g(f(I))}{\|g(I)\| \cdot \|g(f(I))\|} \quad (4)$$

其中  $g(\cdot)$  为 人脸特征编码神经网络。

系数约束损失以及纹理平滑损失分别为:

$$L_c(I) = w_\alpha \|\alpha\|^2 + w_\beta \|\beta\|^2 + w_\delta \|\delta\|^2 \quad (5)$$

$$L_t(I) = \sum_{c \in \{r, g, b\}} \text{var}(T_{c,R}(I)) \quad (6)$$

其中  $w_\alpha$ 、 $w_\beta$ 、 $w_\delta$  分别为人脸几何、表情与纹理的权重系数,用于约束系数大小;而  $R$  为通过人脸 BFM 几何得到的面部区域,  $T$  代表得到的纹理,  $T_{c,R}$  代表面部区域纹理的某一个通道,因此,通过计算面部区域纹理的像素颜色的方差作为损失,即可有效地约束面部纹理的平滑性。

最终的损失函数为上述损失函数的加权求和,其中权重为训练前指定的超参数。

在应用过程中,本系统可以基于用户选定或者上传的一张带有人脸的照片,预测得到人脸 BFM 参数,并利用其中的身份参数  $\alpha$  确定人脸的形状、轮廓,从而使得云展厅中的人物形象与用户选定的形象保持一致;其余参数则采用预设参数来保证人脸纹理、光照、姿势等方面虚拟说话人的合理性。

为了进一步让静态的虚拟人脸模型开口说话成为动态的、高清的讲解动画,综合运用已有的基于语音驱动的人脸说话动画渲染技术<sup>[66,67]</sup>。引入说话动画渲染技术包含两个阶段:(1)基于语音的表情参数映射和(2)说话动画渲染。模型在第一阶段使用参数来控制虚拟讲解员的头部运动,对于轻量级的应用采用先验参数控制的模式,先从语音中提取音素,再从音素映射到人脸表情运动单元,最后从运动单元映射到表情参数,中



间插值过渡帧以达到唇形同步和脸部肌肉运动的自然效果。如果输入的是文字讲稿，云展厅系统使用现有的文字转语音（TTS）技术，将讲稿转为音频输入。模型在第二阶段利用深度学习方法对虚拟讲解员说话动画进行渲染。这一阶段利用三维人脸重建技术，从视频中先恢复出参数序列，把视频对应的音频（及其特征）和参数序列当作训练数据，学习它们到音频和表情参数的对应关系，从而实现端到端的语音驱动控制，使得生成的说话视频具有真实自然的效果。最后，为了增强虚拟讲解员的清晰度、真实性，在三维重建与虚拟讲解员动画生成框架中引入了具有高时效性的图像修复模块<sup>[68]</sup>：该模块将此前渲染得到的虚拟讲解员形象进行逐帧图像修复，即针对在三维重建与说话动画渲染时产生的模糊、瑕疵等进行修复，得到更加真实、自然的虚拟讲解员的说话视频。我们使用在 FlickrFaces-HQ（FFHQ）<sup>[69]</sup>数据集上预训练的图像修复模块。图 2 展示了虚拟讲解员人脸动画生成交互流程与用户界面设计，用户可以对虚拟讲解员形象的性别、造型、人脸进行个性化设置。

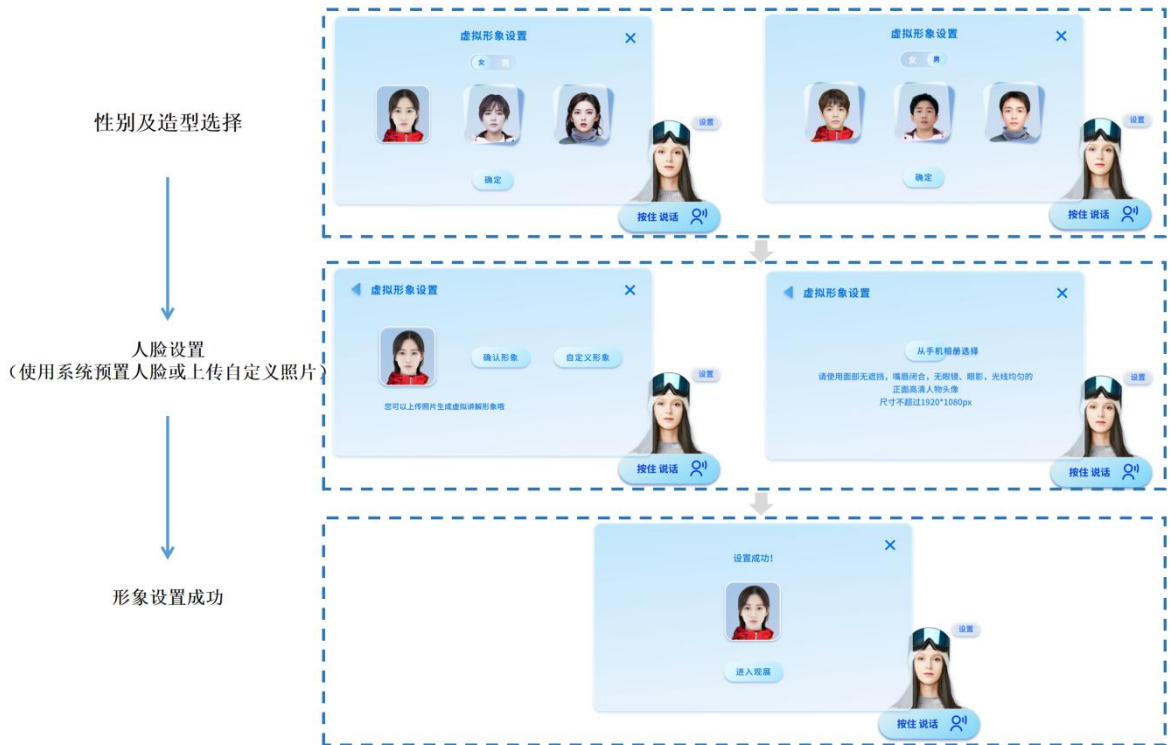


图 2 虚拟讲解员个性化人脸设置流程：用户可以对虚拟讲解员形象的性别、造型、人脸进行个性化设置

### 3.2 虚拟讲解员语音讲解与问答反馈

研究基于冬奥会特定领域结构化知识图谱建立的问答技术，主要包括冬奥会领域知识图谱的建立与问答模型的建立。据我们所知，目前还没有专门针对奥林匹亚和冬奥会的知识图谱，本文提出构建冬奥领域知识图谱，同时研究针对此特定领域知识的问答技术，对特定的实体、关系等信息进行特别处理。图 3 (a) 中展示了虚拟讲解员讲解与问答的交互流程。

在面向冬奥会领域知识图谱构建方面，使用基于实体和关系抽取的构建方法，即利用结构化数据进行实体和关系抽取处理，从而获得完整的冬奥知识图谱。面向问答的信息获取首先需要用户对问题的语义进行解析，得到结构化的问答意图表示，通常由对话动作和槽值两部分构成，其中对话动作是用户问题的目标，表示用户当前的问题想要获得哪类信息；槽值表示用户问题中的约束信息，用以进行匹配。对用户问题语义解析的衡量指标为问答意图识别准确率：对于某个用户输入，只有当对话动作和槽值都预测正确时，才认为这

个意图被正确预测, 否则认为预测错误。表 1 为结构化对话意图样例。

用户问题	高山滑雪包括哪些器械?
对话动作	查询器械
槽值	项目=高山滑雪

表 1 结构化对话意图样例

在问答模型构建方面, 构建知识与记忆引导的问答模型, 将输入信息与特定领域知识图谱相关联, 再利用知识表示模型将结构化的知识图谱表示为向量后与输入信息一起进行基于问答推理网络的知识感知编解码。在问答中, 问答模型需要基于领域知识, 对用户问题进行解析, 提取出其中和知识相关的实体和关系。然后基于得到的结构化语义, 在知识图谱中进行精准匹配, 获得正确答案。为了优化用户使用问答功能的体验, 云展厅系统中采用了问答问题示例卡片设计, 引导用户提出冬奥会云展厅相关专业领域内的问题, 如图 4 (b) 所示, 卡片列举了几类典型的冬奥会知识问答类型示例。

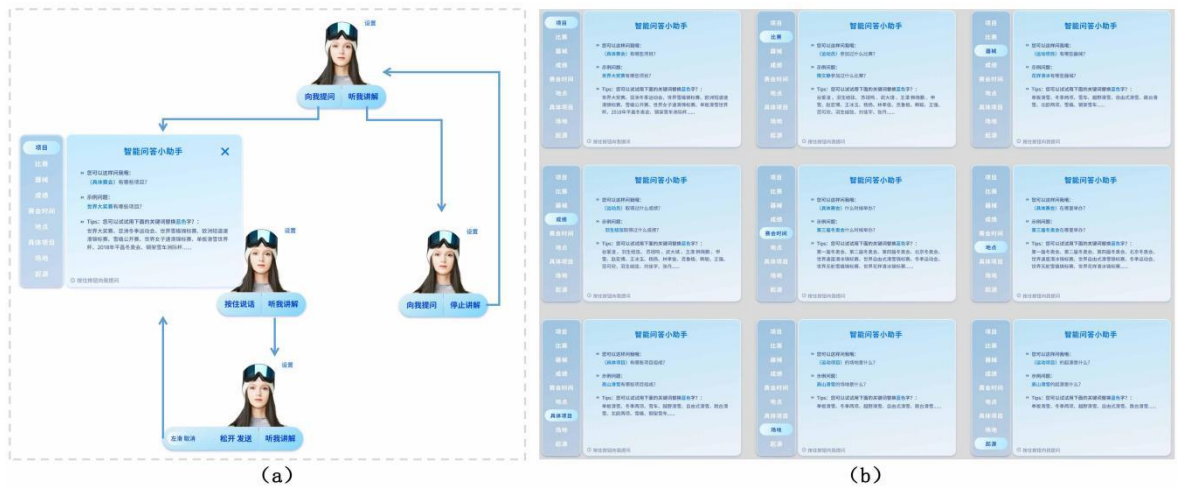


图 3 (a) 虚拟讲解员语音讲解与问答交互流程; (b) 问答问题示例卡片

## 4 云展厅系统设计与实现

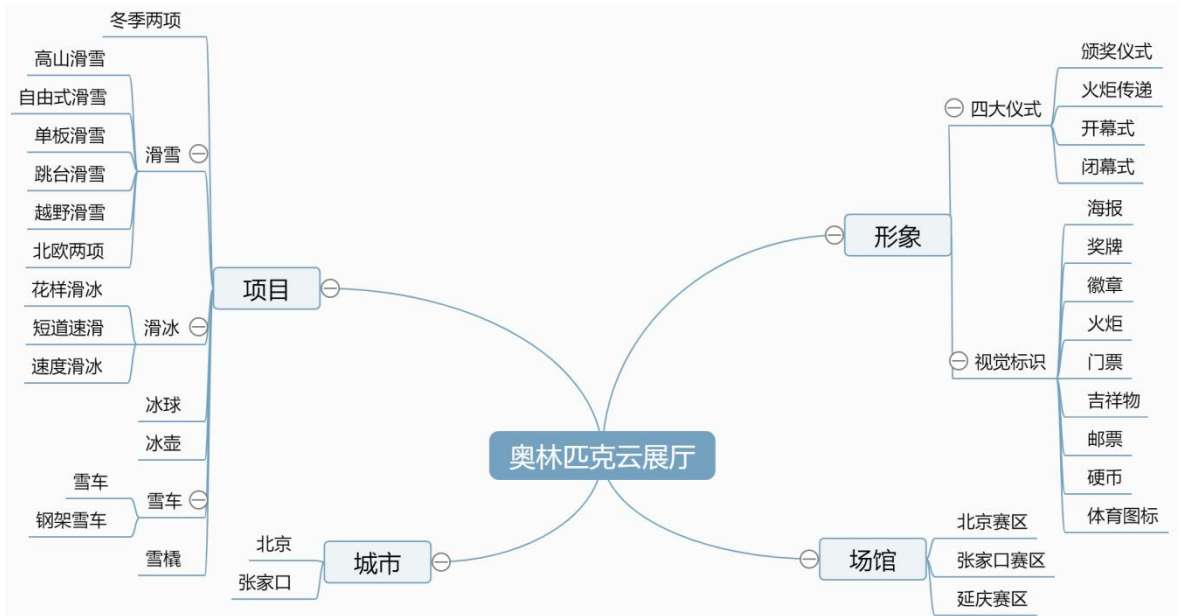
在云展厅的内容选择与布展方面, 我们首先收集了超过 5 万条的文本、图像、音视频、三维模型、动画等多媒体素材, 涵盖冬奥历史、发展、运动项目、人物、场馆、事迹、精彩集锦等多方面的冬奥文化与精神, 然后通过美术专业人士人工整合与再加工形成展品, 并利用展品制作展板 (每个展板内含有 1-5 个展品), 最后精选出 284 个展板用于本文提出的云展厅。进一步地, 引入一种通过图片多层叠加的方式来模拟 3D 效果的 2.5D 在线展览技术, 在不需要 3D 引擎进行实时渲染、不增加系统资源和内存消耗、保证云展厅运行速度的情况下, 改善了二维展览模式画面立体性和丰富程度, 给用户带来身临其境的参展体验。

### 4.1 多媒体布展内容组织

为更好地收藏与展示奥林匹克文化以及北京 2022 年冬奥会遗产, 云展厅收集并整理了多媒体冬奥会素材与展品, 并对其进行了组织与整合。云展厅中包括“项目”、“场馆”、“形象”、“城市”四大主题展厅, 每个主题展厅内都展示了相关的简介概况、历史、发展、文化等内容。项目主题展厅由冬奥会七大运动项目 (冬季两项、滑雪、滑冰、冰球、冰壶、雪车、雪橇) 组成, 每个运动项目分展厅下设冬奥运动项目画廊, 总共包含 15 个具体的运动项目 (花样滑冰、短道速滑、速度滑冰、冰球、冰壶、雪车、钢架雪车、雪橇、北欧两项、高山滑雪、自由式滑雪、单板滑雪、跳台滑雪、越野滑雪和冬季两项)。在每个运动项目画廊内, 详细展览了运动项目的名称及起源、项目特点及文化、项目规则及其演变、项目进奥运的时间、取得每个项目冠军最



多的国家、取得每个项目冠军最多的运动员、各个项目的纪录保持者、各个项目的看点等相关内容。类似地，场馆主题展厅设有北京、延庆、张家口三个赛区分展厅；城市主题展厅设有两个奥运城市（北京和张家口）分展厅；在形象主题展厅展示了冬奥会的视觉标识（包括海报、奖牌、徽章、火炬、门票、吉祥物、邮票、硬币和体育图标）和四大仪式。云展厅内容组织架构如 4 所示。



#### 4.2 模拟3D效果的在线展览方法

我们引入了一种通过图片多层叠加的方式来模拟 3D 效果的在线展览技术，后台布展系统同时控制不同层级元素的位置和速度，来模拟 3D 空间的视角，最终提供给用户具有空间纵深视觉效果的观展页面。展区展位分为超近景层、近景层、中景层、远景层。首先，获取展区长廊的各布置展位的展位素材，然后将所有图层中的多媒体展位素材叠加，通过计算、压缩和集成，以不同层级的展位素材叠加显示在二维平面空间页面上，最后通过设置目标位置和速度控制不同层次的展品移动，模拟 3D 空间的透视和立体效果。云展厅展览效果如图 5 所示。



### 5 用户实验

本文设计了一个用户实验来探究基于虚拟讲解员的视听交互反馈对数字博物馆的知识传播能力以及参

观者的用户体验的影响。一方面,考察云展厅是否能够帮助用户更好地获取以及掌握奥林匹克文化和奥林匹克运动会的相关知识;另一方面,关注用户对云展厅系统的浏览体验与评价,重点探索交互反馈对用户感知控制等因素的影响机制。用户实验中探索的反馈的丰富性感受、感知控制、交互体验、用户意向各指标大多是主观感受,本文采用问卷调查这一主流测量方式进行研究。

### 5.1 参与者

通过在校园网站上发布用户实验招聘通知招募了 32 名参与者(18 名男性和 14 名女性)。参与者的平均年龄为 26.8,方差为 3.14(男性:平均值=28.1,方差=3.02,女性:平均值=25.2,方差=2.57)。实验前,我们向参与者详细介绍了云展厅的构建背景、实验目的以及实验流程,并允许用户试用 5-10 分钟,以便熟悉系统。特别地,实验前告知参与者云展厅的构建目的是为了宣传与弘扬奥林匹克文化和遗产。实验在参与者的个人电脑上进行,云展厅在实验期间确保正常工作。每位参与者的实验时长大约持续 30 分钟。

### 5.2 实验步骤

用户实验中,我们基于云展厅实现了四个基线系统,通过控制虚拟讲解员“有/无人脸画面”与“讲解/问答”两组变量得到云展厅中不同的视听反馈模式。把参与者分为四组(每组 8 人),分别浏览 A、B、C、D 四个云展厅基线系统。四个基线系统中的云展厅布置与展品内容完全一样(如图 6 所示),其区别如下:

- 云展厅 A(即本文提出的奥林匹克云展厅数字博物馆):虚拟讲解员同时具有人脸视觉反馈与问答功能。
- 云展厅 B:虚拟讲解员不具有人脸视觉反馈,具有问答功能。
- 云展厅 C:虚拟讲解员具有人脸视觉反馈,不具有问答功能。
- 云展厅 D:虚拟讲解员不具有人脸视觉反馈,也不具有问答功能。

在用户实验开始之前,每位参与者需要完成一个冬奥知识测试;用户实验完成之后,每位参与者再次完成另一个冬奥知识测试,并填写调查问卷。用户实验前后的两次冬奥知识测试均包含 8 道奥林匹克知识与文化相关的选择题,且题目各不相同。

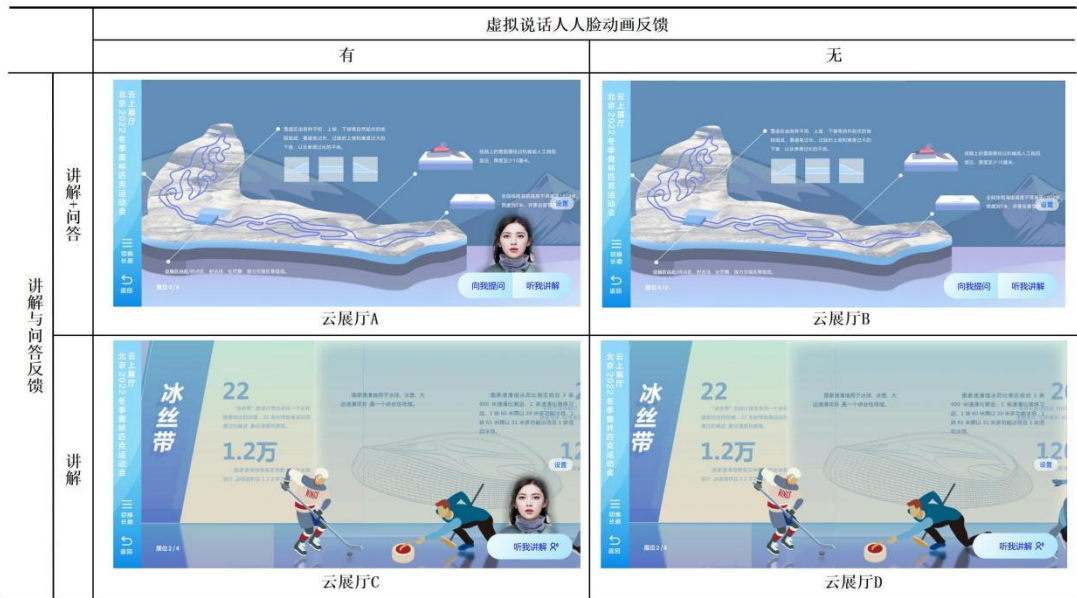


图 6 用户实验基线系统界面示例

### 5.3 测量指标

**(1) 云展厅的知识传播能力:** 用户使用云展厅前后参加奥林匹克知识测试的成绩的变化可以反映用户从云展厅中获取知识的效率, 从而体现云展厅的知识传播能力。实验中, 通过比较各参与者使用各云展厅基线系统前后参加冬奥知识测试的平均成绩的变化来反映各个系统的知识传播能力。用户浏览云展厅之后较浏览前的冬奥知识测试成绩提高的幅度越大, 从一定程度上说明该系统具有越强的知识传播能力, 更能有效地帮助用户获取与记忆奥林匹克知识。

**(2) 反馈的丰富性:** 测量用户对各云展厅基线系统中视听交互反馈的感知丰富度。设计 5 个与云展厅反馈丰富性相关的问题并邀请参与者参加量表评分 (实验过程中涉及的问卷量表评分都采用 7 分制的李克特量表 (Likert scale), 参与者在 1-7 分之间打分, 1 分表示非常不同意, 7 分表示非常同意), 问题如下:

- 该系统可以进行对话。
- 该系统是交互式的。
- 该系统在视觉和听觉模态上具有反馈。
- 该系统提供即时的反馈。
- 该系统提供准确的反馈。

**(3) 感知控制:** 通过测量用户的感知控制指标以探索在云展厅中引入交互反馈对用户的控制感的影响。本文采用 Sundar 等人<sup>[70]</sup>提出的 6 个选项测量参与者的感知控制程度:

- 我能控制我与系统界面的交互。
- 系统环境可以对我发起 (或执行) 的操作做出响应。
- 在交互过程中, 我能够自由地做出选择。
- 我可以自由决定如何访问该系统。
- 我与系统的交互显得很自然。
- 我觉得我可以掌控自己的系统使用体验。

**(4) 用户体验:** 关注用户对云展厅系统的整体满意度与体验, 包括对云展厅系统设计、展品数量与质量、用户界面与用户体验等, 本文采用常用的对交互系统评测的量表选项<sup>[71,72]</sup>, 具体如下:

1) 有用性; 2) 易用性; 3) 易学性; 4) 有效性; 5) 高效性; 6) 令人舒适的; 7) 交流友好的; 8) 用户友好的; 9) 连贯性; 10) 充实的; 11) 有趣的; 12) 令人满意的。

**(5) 用户意向:** 最后, 引入三个选项衡量用户对云展厅的整体意图感受, 分别反映用户是否愿意与他人分享云展厅、是否愿意进一步使用云展厅以及是否愿意更多地参与该系统:

- 推荐意向 (我愿意与他人分享该云展厅)。
- 查询意向 (我希望进一步使用云展厅获取更多信息)。
- 参与意向 (我感觉参与到了该系统中)。

### 5.4 实验结果分析

通过用户实验, 我们共收集到 32 位参与者的冬奥知识测试答题情况与问卷调研结果。我们首先对问卷结果进行了单因素方差分析 (one way ANOVA), 显著性水平设置为  $p < 0.001^{***}$ ,  $p < 0.01^{**}$ ,  $p < 0.05^*$ , 即如果  $p < 0.05$ , 则认为比较对象存在显著差异; 进一步地, 我们进行邓肯多重检验 (Tukey's HSD test) 来确定云展厅 (A 组) 与各基线系统 (B 组、C 组、D 组) 的均值之间差异是否显著, 我们设置了 0.05 的显著性水平来拒绝原假设。详细的实验结果分析如下。

#### (1) 云展厅的知识传播能力

参与者浏览各云展厅基线系统 (A 组为云展厅, B 组、C 组、D 组三组为对照组) 前后参加冬奥知识测试的成绩变化情况如图 7 所示, 可以看出, 使用本文提出的云展厅 (A 组) 对参与者的冬奥知识成绩提高的帮助最大, 视听反馈方式在一定程度上有助于提高云展厅的知识传播能力。

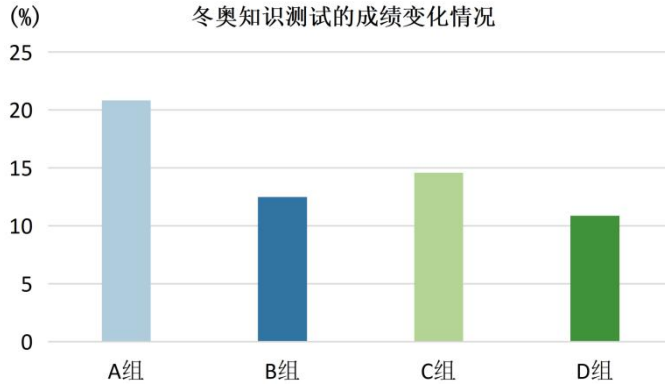


图7 用户浏览云展厅前后参加冬奥知识测试的成绩变化情况。用户浏览 A 组云展厅后参加冬奥知识测试的成绩提高的幅度更大

(2) 反馈的丰富性

对参与者浏览本文提出的云展厅 (A 组) 与各基线系统 (B 组、C 组、D 组) 后对云展厅“反馈的丰富性”相关问卷调查结果进行方差分析,如图 8 所示,可以看出,除“该系统是交互式的”指标外,A 组的云展厅与 B 组、C 组、D 组三个基线系统的其他指标均存在显著差异。进一步地,表 1 中展示了云展厅与各基线系统之间的两两之间的 Tukey' s HSD 检验结果,可以看出,在“该系统可以进行对话”、“该系统在视觉和听觉模态上提供多模态反馈”、“该系统提供即时的反馈”和“该系统提供准确的反馈”四个指标上,A 组的云展厅与 D 组存在显著的差异;在“该系统是交互式的”指标上差异不显著,但也具有同其它选项类似的趋势。进一步地,在“该系统在视觉和听觉模态上提供多模态反馈”和“该系统提供即时的反馈”两个指标上,A 组的云展厅与 B 组存在显著的差异。云展厅中虚拟讲解员的人脸与问答能够在视觉与听觉上为用户提供丰富的交互反馈。

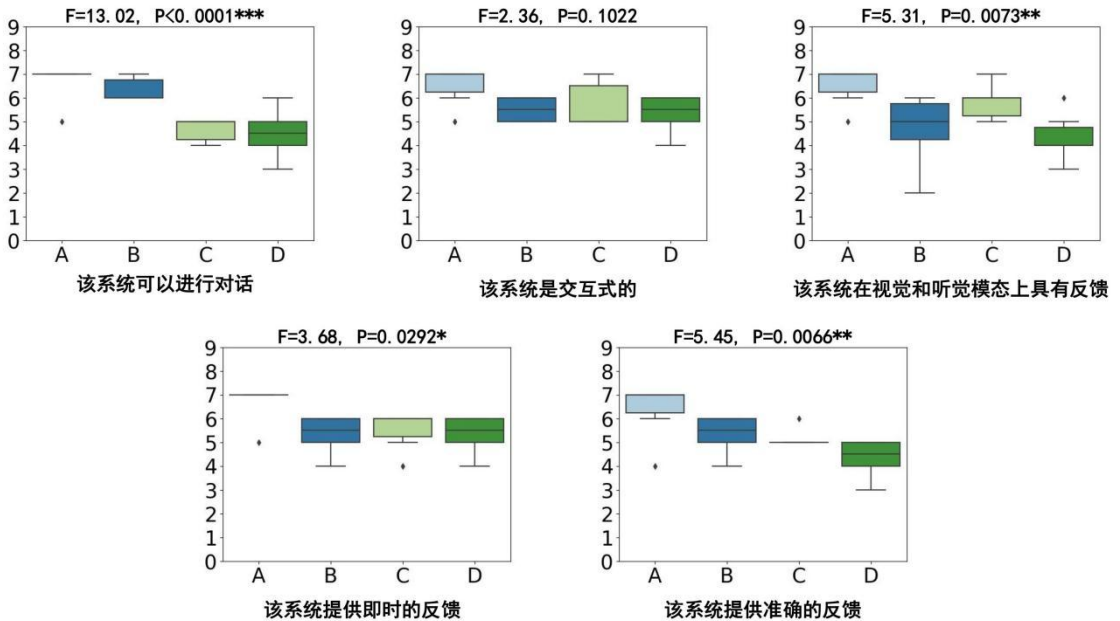


图8 用户浏览云展厅和各基线系统问卷调查中“反馈的丰富性”各指标的箱线图

表 1 Tukey' s HSD 检验结果

测量指标	问卷项目	A 组-B 组		A 组-C 组		A 组-D 组	
		Statistic	p-value	Statistic	p-value	Statistic	p-value
反馈的丰富性	该系统可以进行对话	0.333	0.871	2.000	<b>0.001***</b>	2.167	<b>&lt;0.001***</b>
	该系统是交互式的	1.000	0.189	0.833	0.327	1.167	0.101
	该系统在视觉和听觉模态上提供多模态反馈	1.833	<b>0.036*</b>	0.667	0.707	2.167	<b>0.011*</b>
	该系统提供即时的反馈	1.333	<b>0.049*</b>	1.167	0.098	1.333	<b>0.049*</b>
	该系统提供准确的反馈	1.000	0.217	1.167	0.121	2.000	<b>0.003**</b>
感知控制	我能够控制我与系统界面的交互	1.833	<b>0.004**</b>	0.333	0.890	1.833	<b>0.004**</b>
	系统环境可以对我发起(或执行)的操作做出响应	1.167	0.056	0.500	0.648	1.500	<b>0.010**</b>
	在我与系统的交互过程中,我能够自由地做出选择	1.000	0.268	0.500	0.784	2.167	<b>0.003**</b>
	我可以自由决定如何访问该系统	1.167	0.172	0.833	0.436	2.333	<b>0.002**</b>
	我与系统的交互显得很自然	1.167	0.299	0.833	0.579	2.000	<b>0.027*</b>
	我觉得我自己可以掌控自己的系统使用体验	1.667	<b>0.045*</b>	1.333	0.136	1.833	<b>0.025*</b>
用户体验	很有用	0.833	0.366	0.167	0.987	0.167	0.987
	很容易使用	-0.500	0.751	-0.500	0.751	0.667	0.553
	很容易学习	-0.833	0.171	-0.833	0.171	-0.833	0.171
	很有效果	0.333	0.817	0.000	1.000	0.667	0.324
	效率很高	0.500	0.669	1.167	0.066	0.833	0.258
	令人舒适	0.500	0.839	0.667	0.688	0.167	0.992
	交流友好的	1.500	<b>0.017*</b>	0.833	0.286	0.333	0.882
	用户友好的	0.667	0.398	0.500	0.631	0.500	0.631
	很连贯	1.333	0.185	1.667	0.070	1.000	0.411
	很充实	1.000	0.205	0.500	0.738	0.833	0.347
	很有趣	0.333	0.825	0.833	0.171	0.833	0.171
令人满意	1.000	<b>0.015*</b>	0.000	1.000	0.333	0.683	
用户意向	推荐意向	0.500	0.720	-0.167	0.985	0.167	0.985
	查询意向	1.000	0.249	1.333	0.079	0.833	0.398
	参与意向	0.167	0.988	0.333	0.913	1.833	<b>0.009**</b>

### (3) 感知控制

图 9 展示了参与者对云展厅“感知控制”各指标问卷调查的方差分析结果,从图中可以看出,“感知控制”各指标均存在显著差异。进一步地,从表 1 中可看出,A 组云展厅的“感知控制”各指标均高于 D 组,且存在显著差异;A 组与 B 组、C 组的指标的差异性虽然不显著,但均有高于它们的趋势。进一步地,在“我能够控制我与系统界面的交互”和“我觉得我自己可以掌控自己的系统使用体验”两个指标上,A 组云展厅与 B 组存在显著的差异。一方面,视觉与听觉的反馈能够明显提高用户浏览云展厅时的感知控制(A 组-D 组);另一方面,通过对比(A 组-B 组)与(A 组-C 组)可看出,视觉反馈具有更强的提升用户控制感的能力。



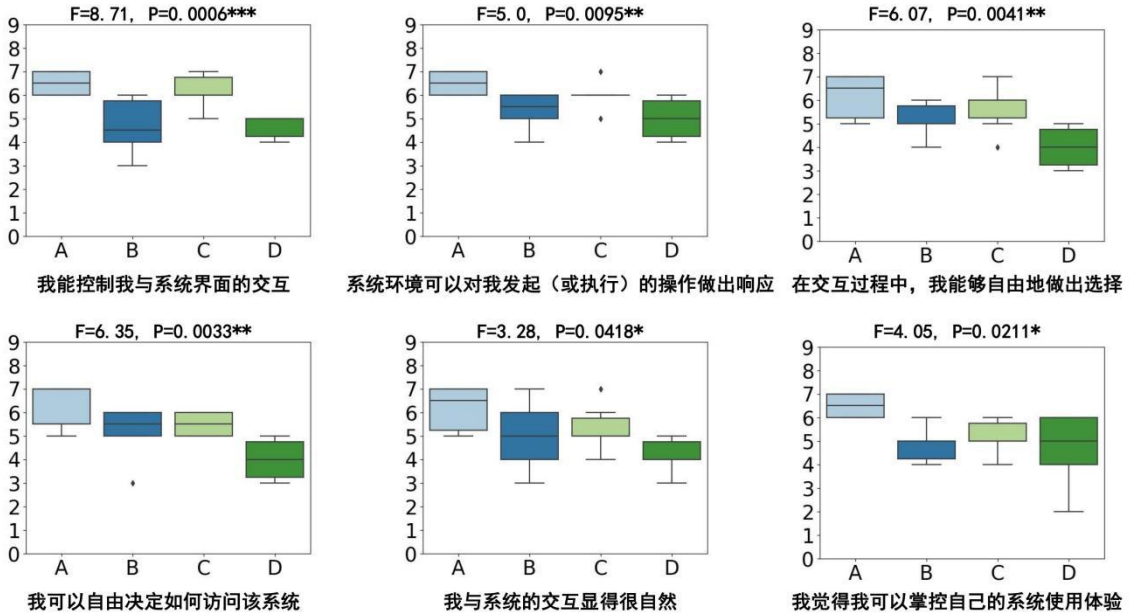


图9 用户浏览 OlympicHall2022 和各基线系统问卷调查中“感知控制”各指标的箱线图

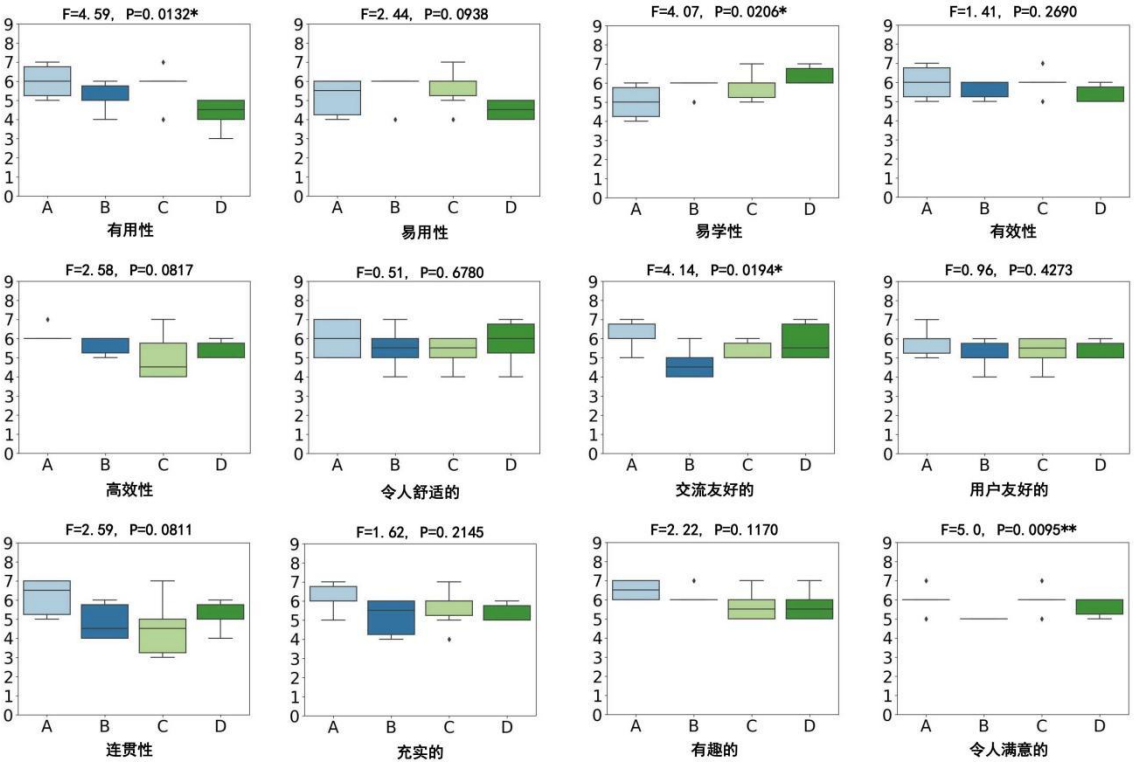


图10 用户浏览 A 组云展厅和各基线系统问卷调查中“用户体验”各指标的箱线图

#### (4) 用户体验

参与者对云展厅“用户体验”各指标问卷调查的方差分析结果如图 10 所示,我们发现除“有用性”、“易学性”、“交流友好的”和“令人满意的”四个指标外,四个云展厅基线系统的其它指标不存在显著差异;从表 1 中可看出, A 组云展厅与 D 组基线系统在用户体验相关的 12 个指标上均无显著差异,而与 B 组基线系统在“交流友好的”“令人满意的”两个指标上有明显差异。

参与者对 A 组云展厅与各基线系统(B 组、C 组、D 组)的用户体验的区别不明显,究其原因,首先可能是由于除虚拟讲解员的设置不同以外,各基线系统的展厅布置及展品素材与云展厅完全相同,而用户的注意力集中于观展,暂时未察觉到各系统间的“有效性”、“高效性”、“连贯性”、“有趣的”等指标的差异。其次,我们在实验前向参与者详细介绍了云展厅,并允许用户试用与熟悉系统,极可能导致用户在正式实验中对系统的“易学性”及“用户友好的”指标上体验相同。最后,由于目前的用户实验无法反映用户与云展厅交互的长期体验,而短期内用户对“令人舒适的”、“充实的”等云展厅性能体验不明显,造成各系统在这两个指标上无明显差异,未来我们将持续接收用户对云展厅体验的反馈并改进云展厅。

#### (5) 用户意向

在“用户意向”的三个问题上,对问卷调查进行方差分析的结果如图 11 所示,结合表 1 中 Tukey's HSD 检验结果可知, A 组云展厅的得分明显高于 D 组基线系统,存在显著差异,用户能更多地感觉到参与到了 A 组云展厅中。A 组云展厅与其他各基线系统间的“推荐意向”与“查询意向”无显著差异。

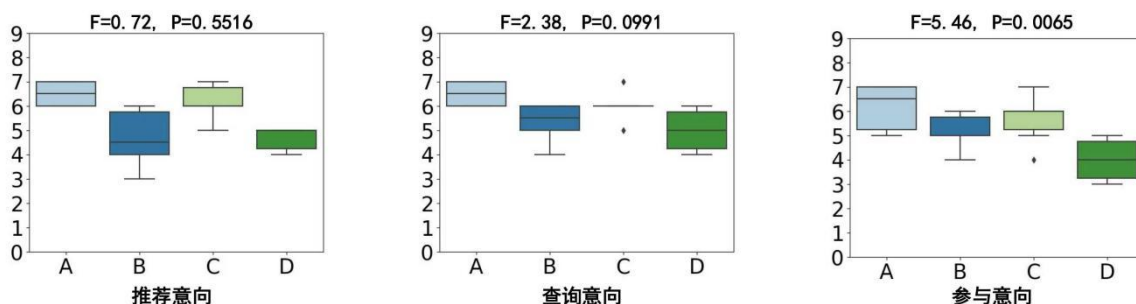


图 11 用户浏览 A 组云展厅和各基线系统问卷调查中“用户意向”各指标的箱线图

## 6 结语与展望

本文介绍了一个专门为北京 2022 年冬季奥林匹克运动会设计的数字博物馆,通过引入虚拟讲解员角色,构建具有视听反馈功能的智能交互云展厅。云展厅中集成了面向特定奥运展馆的智能问答、支持个性化虚拟说话人生成的虚拟讲解员角色快速构建、文本或语音驱动的讲解动画生成以及快速模拟三维立体视觉效果的在线展览等技术,全面、细致地整合了冬奥会相关历史、发展、运动项目、人物、场馆、事迹、精彩锦集等多方面素材,有助于促进对冬奥会精神文化遗产的收藏与传播。冬奥会期间,云展厅于 2022 年 2 月 13 日至 2022 年 2 月 20 日上线试运行;试运行期间,浏览总次数 25,482,平均每日浏览次数 3,185.25,访问 IP 地址来自中、美、日、英四个国家,其中国内覆盖 29 个省份或直辖市、共计 111 个城市。

本文从感知控制的角度研究交互反馈对数字博物馆知识传播功能的影响,通过设计基线系统与用户实验,发现具有视听反馈的云展厅系统能够有效地提高云展厅的知识传播能力与用户的感知控制,用户更愿意参与到具有视听反馈的云展厅系统中。未来的工作包括对云展厅的功能进行扩展,例如,在云展厅中引入更多的交互反馈形式(例如如手势交互),进一步研究数字博物馆中多模态反馈对用户感知控制及交互体验的影响。此外,我们还将不断更新与完善云展厅中的冬奥会文化与知识素材。本文所构建的奥林匹克数字博物馆云展厅,能够在北京冬奥会结束后作为冬奥遗产持续发挥宣传普及的重要作用,更加深入的普及宣传北京冬奥遗产成果。

## References:

- [1] Li X, Chen W, Wu Y. The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology The Adjunct Publication of the 32nd Annual ACM Symposium. ACM, 2019. 42–43.
- [2] Martin, Kit. Multitouch NetLogo for museum interactive game. In Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing, 2018, pp. 5-8.
- [3] 马晓娜, 图拉, 徐迎庆. 非物质文化遗产数字化发展现状[J]. 中国科学: 信息科学, 2019, 49(02):121-142.
- [4] 王春法. 关于智慧博物馆建设的若干思考[J]. 博物馆管理, 2020(3):12.
- [5] Kuflik T, Stock O, Zancanaro M, et al. A visitor's guide in an active museum: Presentations, communications, and reflection[J]. Journal on Computing & Cultural Heritage, 2011, 3(3):1-25.
- [6] Harada T, Hideyoshi Y, Gressier-Soudan E, et al. MUSEUM EXPERIENCE DESIGN BASED ON MULTI-SENSORY TRANSFORMATION APPROACH[C]. 15th International Design Conference. 2018. 2221–2228.
- [7] <https://www.beijing2022.cn/education/index.htm>
- [8] 周立超, 薛东. 以“小小冬奥博物馆”为载体实施奥林匹克教育的实践研究[J]. 体育教学, 2021, 41(12):3.
- [9] Ayobi, Amid, Paul Marshall, Anna L. Cox, and Yunan Chen. Quantifying the body and caring for the mind: self-tracking in multiple sclerosis. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017, pp. 6889-6901.
- [10] Ayobi, Amid, Paul Marshall, and Anna L. Cox. Trackly: A Customisable and Pictorial Self-Tracking App to Support Agency in Multiple Sclerosis Self-Care. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1-15.
- [11] Luria, Michal, Guy Hoffman, and Oren Zuckerman. Comparing social robot, screen and voice interfaces for smart-home control. In Proceedings of the 2017 CHI conference on human factors in computing systems, 2017, pp. 580-628.
- [12] Duval, Antoine. The Olympic charter: A transnational constitution without a state?. Journal of Law and Society, 2018, 45, pp.S245-S269.
- [13] Kubo, Hiromi. The social and economic impact of hosting the Olympic Games: A guide to online resources. 2014, pp. 75(1):24-27.
- [14] <https://olympic-museum.de/>
- [15] Kuflik T, Stock O, Zancanaro M, et al. A visitor's guide in an active museum: Presentations, communications, and reflection[J]. Journal on Computing & Cultural Heritage, 2011, 3(3):1-25
- [16] Pan, Zhigeng, et al. Virtual reality in the digital Olympic museum. IEEE Computer Graphics and Applications, 2009, 29(5), 91-95.
- [17] Pan, Zhigeng, and Ruwei Yun. Digital olympic museum and sports simulation. Virtual Reality & Augmented Reality in Industry. Springer, Berlin, Heidelberg, 2011. 121-132.
- [18] Qi, Yue, Su Cai, and Shen Yang. 3D modeling, codec and protection in digital museum. Second Workshop on Digital Media and its Application in Museum & Heritages (DMAMH 2007). IEEE, 2007. pp. 231-236.
- [19] Ali, Ghazanfar, Hong-Quan Le, Junho Kim, Seung-Won Hwang, and Jae-In Hwang. Design of seamless multi-modal interaction framework for intelligent virtual agents in wearable mixed reality environment. In Proceedings of the 32nd International Conference on Computer Animation and Social Agents, pp. 47-52. 2019.
- [20] Cauchard, J. R., Cheng, J. L., Pietrzak, T., & Landay, J. A. ActiVibe: design and evaluation of vibrations for progress monitoring. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016, pp. 3261-3271.
- [21] Peiris, Roshan Lalitha, Yuan-Ling Feng, Liwei Chan, and Kouta Minamizawa. Thermalbracelet: Exploring thermal haptic feedback around the wrist. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1-11.
- [22] Zhu, Mengjia, Amirhossein H. Memar, Aakar Gupta, Majed Samad, Priyanshu Agarwal, Yon Visell, Sean J. Keller, and Nicholas Colonnese. Pneusleeve: In-fabric multimodal actuation and sensing in a soft, compact, and expressive haptic sleeve. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1-12.
- [23] Anderson, Fraser, Tovi Grossman, Justin Matejka, and George Fitzmaurice. YouMove: enhancing movement training with an augmented reality mirror. In Proceedings of the 26th annual ACM symposium on User interface software and technology, 2013, pp. 311-320.
- [24] Nguyen, Cuong, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. CollaVR: collaborative in-headset review for VR video. In Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, 2017, pp. 267-277.

- [25] Liao, Chunyuan, François Guimbretière, and Corinna E. Loeckenhoff. Pen-top feedback for paper-based interfaces. In Proceedings of the 19th annual ACM symposium on User interface software and technology, 2006, pp. 201-210.
- [26] Cho, Youngjun, Andrea Bianchi, Nicolai Marquardt, and Nadia Bianchi-Berthouze. RealPen: Providing realism in handwriting tasks on touch surfaces using auditory-tactile feedback. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology, 2016, pp. 195-205.
- [27] Hinckley, Ken, Koji Yatani, Michel Pahud, Nicole Coddington, Jenny Rodenhouse, Andy Wilson, Hrvoje Benko, and Bill Buxton. Pen+ touch= new tools. In Proceedings of the 23rd annual ACM symposium on User interface software and technology, 2010, pp. 27-36.
- [28] Vogel, Daniel, and Géry Casiez. Conté: multimodal input inspired by an artist's crayon. In Proceedings of the 24th annual ACM symposium on User interface software and technology, 2011, pp. 357-366.
- [29] Srinivasan, Arjun, Bongshin Lee, Nathalie Henry Riche, Steven M. Drucker, and Ken Hinckley. InChorus: Designing consistent multimodal interactions for data visualization on tablet devices. In Proceedings of the 2020 CHI conference on human factors in computing systems, 2020, pp. 1-13.
- [30] Zhao, Zhenjie, and Xiaojuan Ma. Situated learning of soft skills with an interactive agent in virtual reality via multimodal feedback. In Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology, 2020, pp. 25-27.
- [31] Freeman, Euan, Stephen Brewster, and Vuokko Lantz. Do that, there: an interaction technique for addressing in-air gesture systems. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016, pp. 2319-2331.
- [32] Wang, Chi, Da-Yuan Huang, Shuo-wen Hsu, Chu-En Hou, Yeu-Luen Chiu, Ruei-Che Chang, Jo-Yu Lo, and Bing-Yu Chen. Masque: Exploring lateral skin stretch feedback on the face with head-mounted displays. In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, 2019, pp. 439-451.
- [33] Sidenmark, Ludwig, and Hans Gellersen. Eye&head: Synergetic eye and head movement for gaze pointing and selection. In Proceedings of the 32nd annual ACM symposium on user interface software and technology, 2019, pp. 1161-1174.
- [34] Clarke, Loraine, Eva Hornecker, and Ian Ruthven. Fighting fires and powering steam locomotives: Distribution of control and its role in social interaction at tangible interactive museum exhibits. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1-17.
- [35] Martin, Kit. Multitouch NetLogo for museum interactive game. In Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing, 2018, pp. 5-8.
- [36] Hamilton, William A., Nic Lupfer, Nicolas Botello, Tyler Tesch, Alex Stacy, Jeremy Merrill, Blake Williford, Frank R. Bentley, and Android Kerne. Collaborative live media curation: Shared context for participation in online learning. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1-14.
- [37] Ayobi, Amid, Paul Marshall, Anna L. Cox, and Yunan Chen. Quantifying the body and caring for the mind: self-tracking in multiple sclerosis. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017, pp. 6889-6901.
- [38] Ayobi, Amid, Paul Marshall, and Anna L. Cox. Trackly: A Customisable and Pictorial Self-Tracking App to Support Agency in Multiple Sclerosis Self-Care. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1-15.
- [39] Pradhan, Alisha, Kanika Mehta, and Leah Findlater. Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In Proceedings of the 2018 CHI Conference on human factors in computing systems, 2018, pp. 1-13.
- [40] Luria, Michal, Guy Hoffman, and Oren Zuckerman. Comparing social robot, screen and voice interfaces for smart-home control. In Proceedings of the 2017 CHI conference on human factors in computing systems, 2017, pp. 580-628.
- [41] Wallston, Kenneth A., Barbara Strudler Wallston, Shelton Smith, and Carolyn J. Dobbins. Perceived control and health. *Current Psychology*, 1987, 6(1), pp. 5-25.
- [42] Skinner, Ellen A., and Melanie J. Zimmer-Gembeck. Perceived control and the development of coping. 2011, pp. 35-59.
- [43] 44 Skinner, Ellen A., Melanie J. Zimmer-Gembeck, James P. Connell, Jacquelynne S. Eccles, and James G. Wellborn. Individual differences and the development of perceived control. *Monographs of the society for Research in Child Development*, 1998, i-231.
- [44] Roy, Quentin, Futian Zhang, and Daniel Vogel. Automation accuracy is good, but high controllability may be better. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1-8.

- [45] Benedetti, Luca, Holger Winnemöler, Massimiliano Corsini, and Roberto Scopigno. Painting with bob: Assisted creativity for novices. In Proceedings of the 27th annual ACM symposium on User interface software and technology, 2014, pp. 419-428.
- [46] Lyngs, Ulrik, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotko, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 'I Just Want to Hack Myself to Not Get Distracted' Evaluating Design Interventions for Self-Control on Facebook. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1-15. 2020.
- [47] Chen, Tsai-Wei, and S. Shyam Sundar. This app would like to use your current location to better serve you: Importance of user assent and system transparency in personalized mobile services. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018. pp. 1-13.
- [48] Jhaver, Shagun, Yoni Karpfen, and Judd Antin. Algorithmic anxiety and coping strategies of Airbnb hosts. In Proceedings of the 2018 CHI conference on human factors in computing systems, 2018, pp. 1-12.
- [49] Sundar, S. Shyam, Eun Go, Hyang-Sook Kim, and Bo Zhang. Communicating art, virtually! Psychological effects of technological affordances in a virtual museum. International Journal of Human-Computer Interaction, 2015, 31(6), pp.385- 401.
- [50] Schmidt, Susanne, Gerd Bruder, and Frank Steinicke. Effects of virtual agent and object representation on experiencing exhibited artifacts. Computers & Graphics 83 (2019): 1-10.
- [51] Duguleană, Mihai, Victor-Alexandru Briciu, Ionuț-Alexandru Duduman, and Octavian Mihai Machidon. A virtual assistant for natural interactions in museums. Sustainability 12, no. 17 (2020): 6958.
- [52] Ye, Zi-Ming, Jun-Long Chen, Miao Wang, and Yong-Liang Yang. PAVAL: Position-Aware Virtual Agent Locomotion for Assisted Virtual Reality Navigation. In 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 239-247. IEEE, 2021.
- [53] Bönsch, Andrea, David Hashem, Jonathan Ehret, and Torsten W. Kuhlen. Being Guided or Having Exploratory Freedom: User Preferences of a Virtual Agent's Behavior in a Museum. In Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, pp. 33-40. 2021.
- [54] Chung, Joon Son, Amir Jamaludin, and Andrew Zisserman. You said that?. arXiv preprint arXiv:1705.02966 (2017).
- [55] Zhou, Hang, Yu Liu, Ziwei Liu, Ping Luo, and Xiaogang Wang. Talking face generation by adversarially disentangled audio-visual representation. In Proceedings of the AAAI conference on artificial intelligence, vol. 33, no. 01, pp. 9299-9306. 2019.
- [56] Chen, Lele, Ross K. Maddox, Zhiyao Duan, and Chenliang Xu. Hierarchical cross-modal talking face generation with dynamic pixel-wise loss. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7832-7841. 2019.
- [57] Song, Yang, Jingwen Zhu, Dawei Li, Xiaolong Wang, and Hairong Qi. Talking face generation by conditional recurrent adversarial network. arXiv preprint arXiv:1804.04786 (2018).
- [58] Zhang, Zhimeng, Lincheng Li, Yu Ding, and Changjie Fan. Flow-guided one-shot talking face generation with a high-resolution audio-visual dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3661-3670. 2021.
- [59] Zhou, Hang, Yasheng Sun, Wayne Wu, Chen Change Loy, Xiaogang Wang, and Ziwei Liu. Pose-controllable talking face generation by implicitly modularized audio-visual representation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4176-4186. 2021.
- [60] Zeng, Dan, Han Liu, Hui Lin, and Shiming Ge. Talking face generation with expression-tailored generative adversarial network. In Proceedings of the 28th ACM International Conference on Multimedia, pp. 1716-1724. 2020.
- [61] Yu, Lingyun, Jun Yu, Mengyan Li, and Qiang Ling. Multimodal inputs driven talking face generation with spatial-temporal dependency. IEEE Transactions on Circuits and Systems for Video Technology 31, no. 1 (2020): 203-216.
- [62] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.
- [63] Deng, Yu, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 0-0. 2019.
- [64] Liu, Ziwei, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision, pp. 3730-3738. 2015.
- [65] Karras, Tero, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4401-4410. 2019.



- [66] Yi, Ran, et al. Audio-driven talking face video generation with learning-based personalized head pose. arXiv preprint arXiv:2002.10137 (2020).
- [67] Yi et al. Animating Portrait Line Drawings from a Single Face Photo and a Speech Signal. ACM SIGGRAPH 2022.
- [68] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. 2021. Towards real-world blind face restoration with generative facial prior. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 9168–9178.
- [69] Karras, Tero, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4401-4410. 2019.
- [70] Sundar, S. Shyam, Eun Go, Hyang-Sook Kim, and Bo Zhang. Communicating art, virtually! Psychological effects of technological affordances in a virtual museum. *International Journal of Human-Computer Interaction*, 2015, 31(6), pp.385- 401.
- [71] Oh, Changhoon, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1-13. 2018..
- [72] Lund, Arnold M. Measuring usability with the use questionnaire<sup>12</sup>. *Usability interface* 8, no. 2 (2001): 3-6.