扩展现实中的多模态信息物理交互：混合式博士论文答辩

Ahmad Alhilal，香港科技大学，中国香港；阿尔托大学，埃斯波，芬兰

Kit Yung Lam，卢塞恩计算机科学与信息技术学院，罗特克罗伊茨，瑞士Lik-Hang Lee，香港理工大学，中国香港Xuetong Wang，香港科技大学，中国香港Sijia Li，香港科技大学（广州），中国广州Matti Siekkinen，阿尔托大学，埃斯波，芬兰Tristan Braud，香港科技大学，中国香港Pan Hui，香港科技大学（广州），中国广州；香港科技大学，中国香港

arXiv:2603.15392v1 [cs.MM] 16

摘要——学术活动（例如博士论文答辩）通常局限于现场共同出席或单一的视频会议，导致参与形式僵化、参与体验割裂。我们提出一种多模态框架，打破这种二元局面，支持多种参与方式——从现场出席到沉浸式虚拟现实（*VR*）或浏览器访问，并报告了我们利用该框架举办首场采用扩展现实（*XR*）的混合式博士论文答辩的研究结果。该框架整合全身动作追踪，使用户虚拟化身的动作和手势保持同步，从而能够与现场参与者自然互动，并在虚拟世界中向远程与会者传达肢体语言和手势。该框架利用 *WebXR* 实现跨平台和即时访问，且易于设置。用户反馈分析表明，用户对虚拟现实体验评价积极，也证明了该框架能够有效支持各种混合式活动。

Social 学术活动正日益转向虚拟或混合形式。

扩展现实（XR）通过创建共享虚拟空间，让参与者借助虚拟化身进行互动，从而提升这些体验。 [1], [2], 这种方式兼具远程参与的灵活性与现场互动的丰富性。对于论文答辩等正式学术活动而言，这一能力尤为宝贵。 [3], 在这些活动中，既要维持庄重的仪式感，又要让远程参与者加入，至关重要。本文介绍了我们实施的首场混合式博士论文答辩，该答辩于2023年在香港科技大学举行。

传统论文答辩长期以来一直局限于线下举行，要求候选人、考官和观众齐聚同一地点。传统论文答辩虽有助于营造正式氛围、促进直接互动和细腻的非语言交流，却也受到地域和后勤条件的限制。此外，候选人在选择论文评审委员会（TEC）时也会遇到困难，尤其是当导师或外部评审因其他事务无法亲自出席时。

远程协作工具的兴起带来了虚拟替代方案 [3], ，但这些方案往往难以复现现场答辩的互动动态，存在参与方式僵化、非语言表达能力有限以及临场感割裂等问题。随着学生按时完成学位的压力日益增大，混合式答辩

XXXX-XXX © IEEE 数字对象标识符
10.1109/XXX.0000.0000000

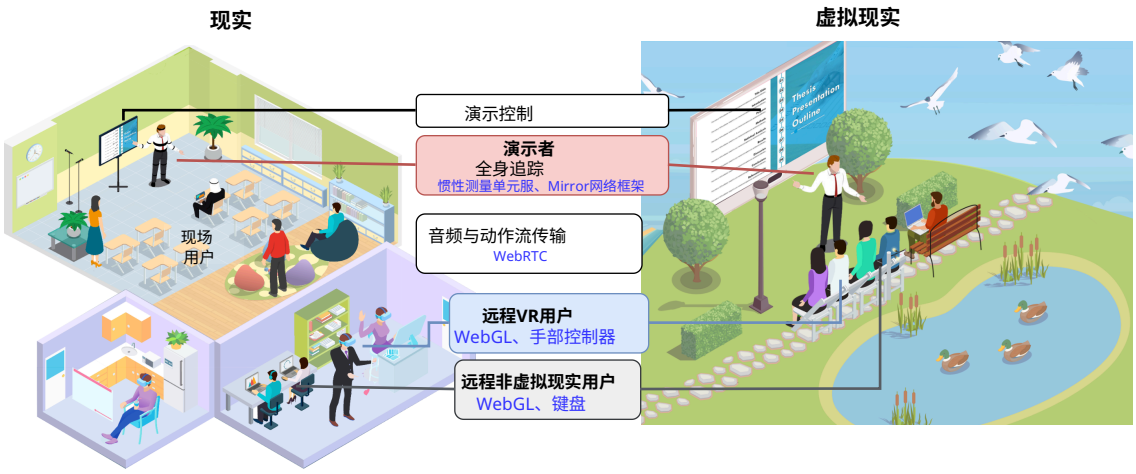


图1：以图示方式描述如何通过多模态信息物理交互组织混合式论文答辩。

已成为应对地理限制、同时确保严格评审进度的必要方式。对于学术答辩而言，眼见为实。从观众的角度来看，全身化身能够独特地保留丰富的非语言交流，而这对于有意义的学术交流至关重要——从细微的姿势变化到方向提示，这些都是上半身虚拟化身无法传达的 [4]。因此，逼真的虚拟临场感与交流对于增强候选人表达的清晰度、增进其与观众的联系至关重要。答辩委员会成员参加答辩的频率不高，因此加入过程必须轻松便捷，并通过直观、无需学习的界面实现。参与过程应无障碍，无需虚拟现实硬件或安装额外软件 [5][6]。这既方便时间有限的考官和导师，也能鼓励更多研究生出席。满足这些要求后，远程参与者便能以与现场出席者相近的还原度体验候选人的交流与演示。

本文介绍了我们的框架（见图 1），该框架用于首次混合式论文答辩，将现场出席与沉浸式虚拟现实参与（通过 WebXR/WebRTC）以及全身虚拟化身同步相结合。我们的框架在现有视频会议和社交虚拟现实平台的基础上更进一步，主要有三项贡献：（*i*）人类-化身交互框架：通过全身动作追踪与复现，实现自然的非语言交流，并以我们此前的工作为基础 [7]，（*ii*）混合活动架构：原生支持虚实参与并行进行，在保持现场答辩的正式氛围与互动动态的同时开展答辩，同时提供不受地理位置限制的灵活性；以及（*iii*）无障碍访问：通过 WebXR 和 WebRTC 技术实现直观的远程参与，无需事先进行技术准备，从而消除采用障碍（例如对专用硬件或软件专业知识的要求）。因此，这种混合形式通过结构化考核和自然的非语言线索，保留了传统答辩的严谨性，同时维持仪式的正式感（例如穿着化身正式着装并做出全身手势）及空间交互，确保线下和虚拟参与者都能感受到同等的学术庄重性。

在本文其余部分，我们将研究虚拟替代方案并讨论其局限性（第1节）。随后，我们将介绍在实际 MR 混合式论文答辩中使用的多模态交互框架（第2节）。接下来，我们将评估用户对该框架性能的感知，以及该框架在开展论文答辩各项活动方面的影响（第3节）。在第4节中，我们将讨论该框架如何适应混合式论文答辩的各个阶段。

1. 背景与挑战

我们首先考察混合式论文答辩面临的独特挑战，然后讨论现有平台、其适用性及局限性。

1.1. 挑战

由于TEC成员通常只是偶尔参加答辩，因此必须借助便捷的工具，让他们能够直观地加入活动，无需经历任何学习过程。技术应能促进远程口头和非口头交流。至关重要的是，技术——

TA 表1：社交虚拟现实平台在可访问性、准备工作和虚拟形象特征方面的比较

平台	可访问性	准备工作	虚拟形象类型
Rec Room 1	跨平台（虚拟现实/桌面端/移动端） 需要VR头显或桌面客户端	账户设置、虚拟形象定制	全身、上半身， 卡通风格
VRChat 2	需要VR头显或桌面客户端	账户设置、虚拟形象定制	全身（用户生成）
Meta Workrooms 3	需要Meta Quest头戴设备	Meta账户、工作空间设置， 虚拟形象定制	上半身， 精准手部追踪
Spatial 4	跨平台（虚拟现实/桌面端/移动端） 浏览器访问	账户设置、房间配置， 虚拟形象定制	全身（用户生成）
我们的平台	**跨平台（虚拟现实/桌面端/移动端） 浏览器访问**	**一键操作** 精选虚拟形象创建	**全身（基于照片） 全身追踪**

——应始终不引人注意，让候选人能够全神贯注地与现场参会者交流。

部署挑战：大多数观众和TEC成员都没有VR头显，也没有时间在自己的设备上安装客户端应用，无论设备是VR设备、移动端还是PC。他们希望无需事先准备就能参加活动，而任何技术障碍都可能让他们打消参加的念头。这些额外要求可能妨碍他们参与，并降低他们在答辩过程中的参与度。论文评审委员会成员通常日程繁忙，也很少参加答辩。因此，加入虚拟论文答辩应当直观易用，且只需付出最少的努力。

演讲者参与度：论文答辩以候选人为中心，流畅的沟通与互动至关重要。因此，博士候选人的讲话和身体动作应当被准确捕捉，并在虚拟世界中重现。为确保高度参与，准确捕捉并重现博士候选人的身体动作至关重要。肢体语言和手势——例如手部动作和身体姿势——是传达想法、增进沟通的有力工具。这些动作能让信息更丰富、更清晰，有助于与观众建立联系，并让要点更令人印象深刻 [8]。与此同时，候选人必须与现场考官保持充分互动，以维护传统答辩的严谨性（结构化考核）、正式性（仪式规范）和真实性（全神贯注）。

1.2. 背景

视频会议平台与混合式参与。自 COVID-19疫情以来，Zoom、Microsoft Teams 和 Google Meet 等视频会议平台（VCP）已被广泛用于远程学术活动，包括论文答辩。哥伦比亚大学等高等教育机构已实施混合灵活式（HyFlex）学习——一种整合现场参与和基于在线视频会议平台参与的多模态方法。该模式允许学生在每次课程中灵活选择偏好的参与方式（现场、同步远程或异步），同时保持一致的学习成果 [9]。在其他机构中，哥伦比亚大学信息技术部门（CUIT）已在校园教室中部署集成式视听系统，包括摄像机和麦克风，以建立支持现场与远程参与者同时参与的混合灵活式学习环境 [10]。尽管基于混合灵活式视频会议平台的论文答辩能够促进实体与虚拟参与，但也会带来成本。候选人必须从关键的口头报告准备和论文最终完善工作中分散宝贵的时间和注意力，以处理混合式活动组织。此外，管理人员还需承担配备技术设备的答辩室以及筹集额外资金等相关成本 [10]。此外，由于参会者窗口众多，且受限于摄像机角度或屏幕共享，无法充分观察非语言线索，因此在论文答辩期间监测参与者的参与度十分困难 [11]。此外，虚拟沟通规范可能会打乱自然互动，导致发言轮次不明确，并错失提问或插话的机会。此外，远程参与者必须安装并熟悉视频会议平台，这往往会打断学位论文答辩的流程和正式性 [12]。

社交虚拟现实平台。表1 对比了主流社交虚拟现实平台在可访问性、准备工作和虚拟形象特征方面的表现，以评估其是否适用于论文答辩环境。Rec Room [1]提供以游戏为中心的社交空间，但缺少结构化轮流发言协议等必要功能，

[1]https://recroom.com/

正式化身和文档共享——这些都是学术答辩所必需的功能。VRChat [2],Horizon Workrooms、[3],以及 Spatial [4] 都能够在虚拟现实空间中复现论文答辩。Park等人提出了一种解决方案，通过探索在虚拟会议中将手势用于交流和控制，丰富非言语交流 [13], [14]。然而，这些平台存在四项主要的参与障碍：(1) 硬件和软件要求（VRChat 和 Rec Room 需要 VR头显或桌面客户端；Horizon Workrooms 需要 Meta Quest头戴设备），(2) 必须创建账户，(3) 房间配置，以及 (4) 所有参与者都必须定制虚拟形象 [5], [6]。我们的框架通过以下方式降低这些障碍：(*i*) 集中设置：仅由演示者负责虚拟房间配置和虚拟形象生成；(*ii*) 身份自动化：使用基于照片的全身化身，无需参与者手动定制；以及 (*iii*) 简化访问： 无需预先安装软件，即可通过浏览器一键加入。

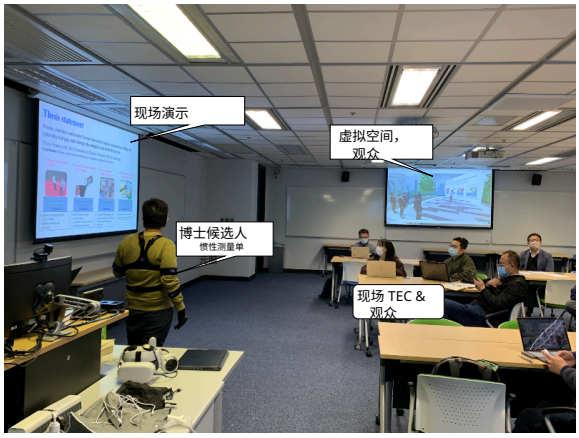
本文介绍了一种新颖的框架，该框架开创了首场结合实体与虚拟参与模式的 XR 论文答辩。该框架专门应对用户（论文答辩委员会成员和观众）普遍面临的时间限制和技术素养方面的挑战，同时为演讲者保留完整的虚拟现实功能，例如全身追踪与实时化身协调。

2. 现实世界中的混合式论文答辩实施

本节介绍我们如何实施首场在混合现实环境中举行的博士论文答辩，并在两种参与方式（现场和远程）下整合了三种参与者角色——演示者、TEC成员和观众。如图1和图2所示，该框架同时支持实体参与（在传统答辩室现场参会）和虚拟参与（远程与会者通过沉浸式化身具象化进入虚拟空间）。

2.1. 博士论文答辩流程

在香港，典型的论文答辩包括候选人的口头报告，随后是与 TEC 成员和观众进行的公开讨论、与 TEC 成员进行的闭门讨论，以及候选人离场后的审议。最终决定是



(a)现场论文答辩，画面展示了身穿惯性测量单元服的博士学位候选人、考官、观众、一台用于现场演示的显示器，以及另一台显示虚拟空间的显示器。



(b)显示器上显示了虚拟空间，其中包括演示内容、演讲者的虚拟化身以及其他用户的虚拟化身。

图2：香港科技大学的真实世界混合现实博士答辩。

由TEC主席传达，通常以仪式性的握手结束。对于口头报告，候选人的惯性测量单元服捕捉全身动作，使远程与会者能够观察自然手势，同时与现场考官保持直接的目光接触和互动。对于公开和闭门讨论，虚拟现实中的空间视听功能能够定位提问参与者，复现实体答辩室的互动动态。TEC主席与候选人之间的握手也会在虚拟世界中得到复现，以恢复仪式礼节。接下来，我们将讨论如何利用技术和系统配置，在虚拟现实中远程复现这些体验。

[2]https://hello.vrchat.com/[3]
https://www.oculus.com/workrooms[4]
https://spatial.io/

2.2. 物理设置与用户角色

图1 和 图2 展示了混合现实论文答辩中按角色划分的设备和硬件配置（例如，演示者的动作捕捉服、远程观众界面），以及所捕捉的多模态数据流（音频、动作）。图2a展示了现场论文答辩的实体设置，其中集成了双显示器：主显示器显示候选人的演示内容，副显示器呈现虚拟答辩环境。博士候选人穿戴惯性测量单元服，以同步虚拟空间中的全身虚拟化身动作。该设置使现场参会者能够同时查看演示内容（主显示器）和虚拟互动（副显示器），而WebRTC音频流（后文详述）则支持与远程参与者进行实时通信。中继服务器部署在云端，用于将用户的多模态数据中继至远程用户。

接下来，我们将讨论四种参与者角色，这些角色根据参与模式（现场/远程）和访问类型（VR头显/基于浏览器）进行分类。

演示者 是为其论文进行答辩的博士候选人。他在此次活动中拥有最高级别的交互权限。在本文讨论的答辩中，我们使用了 Perception Neuron 3 Body Kit 动作捕捉设备[5]，配备了 17 个 Neuron传感器。这些传感器附着在他的头部、肩部、脊柱以及上肢和下肢。它们追踪并捕捉候选人的身体动作和手势。17 个传感器对应 17 组位置与旋转数据，这些数据通过 Unity3D 上的 C#脚本处理，并映射到演讲者的虚拟化身的 59 个关节，以控制其身体动作。此外，候选人还手持遥控器，在答辩期间控制 PowerPoint演示文稿（PPT）的播放进度。由此产生多模态数据：*59* 个虚拟化身关节的变换数据（位置和旋转）、与 *PPT* 相关的命令以及音频。

远程虚拟现实用户是指使用VR头显远程参加虚拟答辩的参与者。他们进入虚拟场地（见图2b），并使用手部控制器控制虚拟化身的动作。VR模式下的应用会运用逆向运动学（IK），根据空间中选定的位置反向计算关节朝向。虚拟化身上肢的动作根据其手部位置计算。与Meta Workrooms类似，由于腿部未被追踪，该应用会根据移动速度触发下半身动画——站立或行走。最终，虚拟化身的手势和变换数据以9个IK关节的变换数据形式呈现。因此，远程虚拟现实用户会传输两类数据：*9*个*IK*关节的变换数据、下半身动画状态和音频。

远程非VR用户指通过笔记本电脑上的网页浏览器加入的参与者。他们使用键盘输入（WASD）移动自己的虚拟化身，并使用鼠标调整朝向。由于他们的手部和腿部动作未被追踪，应用程序会根据输入强度动态混合预设动画：无输入时保持静止姿态，按键时则执行手臂摆动的行走动作。这些用户会传输三类数据：变换数据（位置和旋转）、动作和动画状态，以及音频。

现场参与者亲自参加了答辩，包括论文答辩委员会成员、导师和普通观众。现场参与者的音频被录制，并与演示者的声音一同传输。

2.3. 实施

下面详细介绍技术实现，包括通过 Avatar SDK 使用特定角色的虚拟化身、通过 Mirror 实现实时身体动作同步、通过 WebRTC 传输音频流，以及通过 WebXR 实现跨平台支持。

虚拟化身 我们为博士论文答辩设计了针对不同角色的虚拟化身，为演讲者、导师和考官生成了照片级真实感三维虚拟化身，并准确还原了他们的真实外貌。在每场答辩开始前，我们通过专用门户网站收集考官和导师的照片，然后使用 Avatar SDK [6] 根据这些照片生成逼真的 3D 虚拟形象。登录后，每位考官和导师都会自动获得与其对应的个性化虚拟化身。对于观众，我们提供了一组选定的预设通用类人虚拟化身，供他们进入虚拟环境时选择。无论是考官、导师还是观众，任何用户登录系统后，其虚拟化身都会自动分发给所有其他已连接客户端，使所有参与者都能在共享虚拟空间中看到彼此的虚拟形象并进行互动。由于博士答辩以候选人为中心，演讲者的虚拟化身默认会预先加载到所有参与者的本地 XR 环境中。

数据传输/同步 该框架使用 Mirror[7]，（一款免费开源的游戏网络库）实现虚拟化身和幻灯片同步

[5] https://neuronmocap.com/pages/perception-neuron-3

[6] https://avatarsdk.com/ [7] https://mirror-networking.com/

。它会传输演示者的全身动作（59 个关节变换）和当前幻灯片、VR用户的逆运动学数据（9 个关节）及下半身动画状态，以及非VR用户的变换和肢体动画状态。远程应用程序会处理这些多模态数据，实时渲染动画化身。

虚拟幻灯片我们将 PowerPoint 幻灯片转换为单张图像，并将其作为纹理导入 Unity3D。随后，我们使用这些纹理创建材质，并将其映射到 VR平面上，作为演示画面。最后，我们实现了一个幻灯片放映 C#脚本，通过键盘输入使用“下一张”和“上一张”控件浏览纹理。演示者的虚拟现实应用会实时传输幻灯片编号，以便在所有已连接客户端之间同步。演示者同时控制虚拟和实体幻灯片，确保两种现实中的内容一致。

音频 该框架利用 WebRTC[8] 与远程用户进行音频通信。它支持双向音频通道，以便在开放问答环节中进行问答。在本文讨论的答辩中，所有参与者——包括考官、导师和远程用户（虚拟现实和非虚拟现实用户）——默认均处于静音状态。现场参会者（考官、导师和观众）与候选人身处同一物理空间，因此他们的音频会与演示者的声音一同被采集并传输。在问答环节中，用户可以取消静音并提问。

跨平台与网页访问该框架利用 WebXR API [9] 在网页浏览器中提供虚拟现实体验。这为扩展现实提供了原生的跨平台支持，无需针对特定平台（桌面端或虚拟现实）开发应用程序。为了在我们的 Unity3D 环境（C# 代码）中处理动作数据，该框架集成了 SimpleWebXR库 [10]，以实现 Unity3D 与 WebXR JavaScript API 之间的接口。使用 VR头显的远程参与者可以通过兼容 WebXR 的浏览器直接加入，手部控制器则支持虚拟手部动作等自然交互。SimpleWebXR库通过 WebSockets 在 WebXR 与 Unity3D 之间进行通信，而 WebGL 则提供硬件加速渲染，从而实现可通过标准网页浏览器访问的完整虚拟现实体验。

3. 用户研究

我们开展了一项用户研究，以评估参与者对虚拟博士答辩的看法，尤其关注候选人在数字环境中的表现如何被评价。

3.1. 参与者。

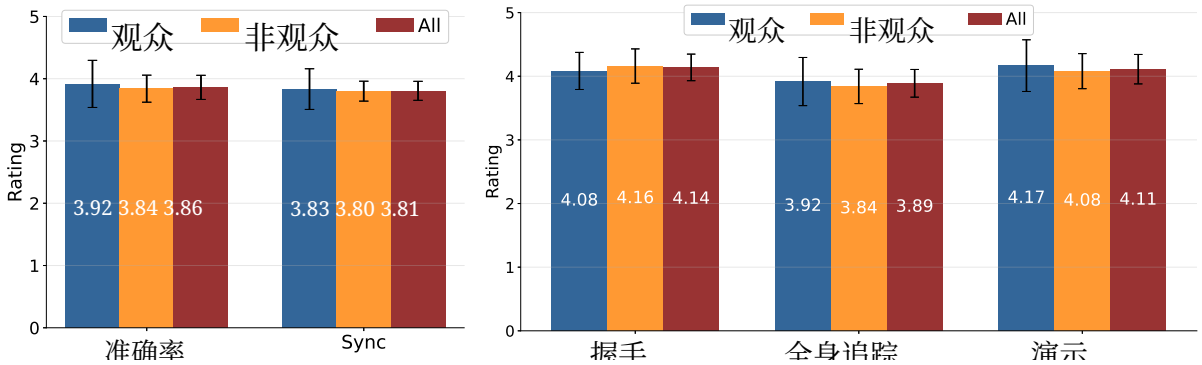
我们对36名参与者（12名答辩出席者、24名未出席者）进行了在线调查，其中男性22名、女性14名，年龄在19至42岁之间。根据他们对该技术的熟悉程度，可将其分为以下几类：不熟悉（2/36）、略微熟悉（9/36）、中等熟悉（12/36）和非常熟悉（13/36）。该调查旨在收集他们对这一框架的反馈。

3.2. 流程。

由于博士答辩以候选人为中心，我们的评估重点关注参与者对候选人虚拟表现的感知，包括握手等正式的多用户互动。在总结致辞环节，一位现场考官佩戴VR头显，与候选人握手并向其表示祝贺，将物理与虚拟元素融为一体。随后，出席者（n=12）根据与现场答辩相比的体验进行评分。

未到场者（n=24）观看了一段视频，视频展示了一位穿着 IMU suit 的真人，以及其同步化身动作在虚拟现实中三种典型使用场景下的表现：（*i*）**握手**：两人握手；（*ii*）**演示**：一人使用 PowerPoint 幻灯片进行演示；以及（*iii*）**全身追踪**：一人进行自由动作（手势和四肢动作）。观看每段视频后，参与者通过在线调查填写了李克特量表问卷，评价自己的体验。我们请所有研究参与者使用五点评分量表，评价他们感知到的化身重现准确率，评分范围从“非常不准确”到“非常准确”：（*1*：非常不准确，*2*：不准确，*3*：比较准确，*4*：准确，*5*：非常准确）。他们还使用五点评分量表，评价身体动作与虚拟化身动作之间的同步程度，评分范围从“完全不同步”到“完全同步”：（*1*：完全不同步，*2*：不同步，*3*：比较同步，*4*：同步，*5*：完全同步）。由于参与者具有扩展现实技术方面的经验，他们还评价了多模态交互在以下方面的前景：

[8] https://webrtc.org [9] https://immersiveweb.dev/ [10] https://github.com/Rufus31415/Simple-WebXR-Unity



（a）对表演的感知。　　（b）对防御活动功能性的感知。

图3：用户感知的MR论文答辩表现：参会者评分（n=12，现场/远程）与非观众评分（n=24）。

针对每种使用场景，使用五点评分量表评价多模态交互能否实现成功的虚拟答辩，评分范围从“肯定不能”到“肯定能”。

3.3. 结果。

图3 展示了各参与者组（参会者、非参会者和全体参与者）的对比结果，并给出了平均评分及其95%置信区间。分析表明，所有参与者的体验均持续为正面，其中观众组（参会者）的评分尤其高：准确率（3.92/5）、同步（3.83/5）、握手（4.08/5）、全身追踪（3.92/5）和演示（4.17/5）。非观众组（非参会者）也报告了类似的积极结果：准确率（3.84/5）、同步（3.80/5）、握手（4.16/5）、全身追踪（3.84/5）和演示（4.08/5）；全体参与者的汇总结果同样稳健：准确率（3.86/5）、同步（3.81/5）、握手（4.14/5）、全身追踪（3.89/5）和演示（4.11/5）。

总体而言，结果表明参与者对虚拟现实空间中的表现评价良好。参与者报告称感知准确率较高（M=3.86，95% CI±0.15），演示者与虚拟化身的动作同步性也较高（M=3.81，95% CI ±0.16）。MR论文答辩平台实现了稳健的动作准确率与同步性，证实其实时追踪及虚拟形象与演讲者协调能力可靠。这些指标表明，平台成功将漂移降至最低；这对于实现类似面对面互动的自然非语言交流至关重要，进而提升了参与者在虚拟现实中的参与度。

活动专项评估进一步凸显了MR论文答辩对答辩关键活动的有效性和适用性：（*i*）握手互动（M=4.14，95% CI ±0.14）得分最高，成功重现了自然的社交手势，尤其是在仪式性时刻（例如论文委员会主席向候选人表示祝贺）；（*ii*）全身追踪（M=3.89，95% CI ±0.17）表现始终良好，能够支持动态动作，且没有明显延迟，使非语言交流自然流畅，媲美面对面互动；（*iii*）演示功能（M=4.11，95% CI ±0.13）获得了最稳定的评分，凸显了该平台在支持虚拟答辩核心任务方面的可靠性，这些任务有助于学术讨论，包括幻灯片导航、批注和图表/行引用。

这些发现共同验证了我们的多模态框架在支持虚拟论文答辩所有关键方面的有效性，既展现了技术稳健性（所有指标均为 $>$ 3.8/5），也体现了运行可靠性（95% CI 均为 $\leq$ ±0.17）。从参与者反馈所反映的普遍可访问性来看，多位与会者表示，他们点击链接后即可加入论文答辩“无需教程或任何设置，点击链接后便可立即加入。”一位委员会成员表示，“我原以为会遇到技术障碍，但通过 *Chrome* 浏览器使用起来非常顺畅”，这凸显了平台无需专用硬件。上述体验符合我们的设计目标，即实现真正的无障碍参与，让即使是首次使用者也能无需任何准备便参与其中。

4. 讨论

4.1. 流程总结

我们首先分析本框架如何简化博士论文答辩流程，并收集参与者对其有效性的反馈。

口头报告。高保真动作捕捉能够在演示过程中自然呈现肢体语言，从而提升沟通清晰度和

虚拟环境中的观众参与度。更重要的是，基于IMU的方法无需使用VR头显或桌面界面，使候选人能够与现场考官保持直接的目光接触并进行互动。

开放式/闭门讨论。VR环境使演讲者能够立即找到提问的参会者，通过以下方式再现实体答辩室的动态情境：(1) 空间三维音频，便于在三维空间中定位发言者；(2) 优化的视觉布局，使演讲者通过自然转动头部即可看清所有虚拟化身。

结果宣布。答辩成功完成后，委员会主席可以与候选人进行虚拟或现场握手。用户研究的参与者对这一手势以及虚拟现实中的全身追踪表示认可。为吸引远程观众，该框架会通过互联网将委员会主席和候选人的实时身体和手部动作数据传输至远程用户的应用程序。随后，应用程序会渲染虚拟化身之间同步的握手动画。这保留了虚拟现实答辩的仪式感——这是相较于缺乏空间交互的传统视频平台的一项关键优势。

4.2. 性能与感知

我们的框架将全身虚拟化身交互与面向演讲者的显示器相结合，从而能够与远程参与者进行自然的非语言交流。用户研究显示，参与者对该框架的性能评价良好，并指出演讲者虚拟化身动作的准确率和同步性很高。针对具体活动的评估证实，该框架能够有效支持答辩的核心任务，在虚拟演示、全身追踪准确率和仪式性握手方面均获得了较高评价。候选人无需佩戴VR头显或使用桌面端界面，因此可以全神贯注于实体答辩空间，同时其动作和手势会被复制并在虚拟现实中呈现。这使候选人能够全神贯注于现场参会者，同时与远程与会者进行互动。

4.3. 推进最先进技术。

该框架弥补了视频会议平台的不足，包括感知到的正式程度较低、非语言交流不足以及参会者可见性较差。具体而言，该框架纳入了四项关键解决方案：（*i*）沉浸式三维环境通过基于照片级真实感三维虚拟化身的交互，重现实体答辩场景，恢复仪式感；（*ii*）借助全身手势追踪和空间音频实现的具身交互，保留非语言交流线索，例如用于阐释概念的示范性手势、用于突出重点的强调性动作，以及有意的姿势变化；（*iii*）三维空间音频和高分辨率音频可重现真实世界的声学保真度；以及（*iv*）自然的三维交互：提问者靠近时，语音和虚拟化身的大小会随距离调整，从而在保持完全可见的同时，摆脱僵化的视频网格。与 HyFlex 相比，我们的框架降低了部署复杂度 [10]，，只需双显示屏配置——一个屏幕用于现场演示，另一个用于虚拟环境。与依赖多个摄像头、麦克风和专用视听基础设施的 HyFlex 不同，我们的方法最大限度地减少了硬件依赖。我们的框架还通过支持演示者与虚拟化身的交互、虚实同步参与，以及借助零安装网络技术实现直观、无障碍的访问，进一步推动了社交虚拟现实平台的发展。

4.4. 局限性与未来方向

演讲者化身框架无法捕捉面部表情，因此与面对面互动相比，远程互动中的情感联结有所减弱。虚拟幻灯片的实现存在两项主要局限。首先，将 PPT 幻灯片转换为图像会丢失交互元素（动画、切换效果和嵌入式媒体）。其次，同时控制实体幻灯片和虚拟幻灯片会增加演讲者的认知负荷，并带来两种现实之间出现不一致的风险。该框架要求演讲者穿戴并仔细校准惯性测量单元服，这增加了技术复杂性，可能会打乱答辩的自然进程。虽然这套服装本身并不令人感到不适，但准备过程需要花费时间并投入技术精力，而这些本可用于专注演示。基于视觉的方法（如 3D 高斯泼溅（TaoAvatar [15]））可以免去可穿戴传感器，仅通过摄像头即可自动完成全身捕捉。这样一来，现场参与者将完全察觉不到这一过程。

5. 结论

本文介绍了我们为混合式博士论文答辩设计的多模态扩展现实框架，该框架首次支持远程参与。该框架采用 WebXR 技术，支持跨平台访问，无需任何技术前置条件，从而让所有人都能参与。框架结合实时运动追踪，实现全身化身同步

并通过 WebRTC 提供音频流传输，从而实现自然的口头交流和非语言交流。这种双通道方式在虚拟现实中再现了答辩的核心活动——从口头报告和自然手势到正式仪式——同时不损害面对面互动的真实感。用户研究显示，参与者对该框架的性能及其在实体空间和虚拟空间中维持参与度的能力持积极看法。未来的改进将着重于以基于视觉的捕捉方式取代惯性测量单元套装、支持交互式虚拟幻灯片，并增加面部追踪以增强情感联结，从而在提升真实感的同时简化框架。

6. 致谢

本研究部分得到广州市南沙区科学技术局根据合同编号 2022ZD01 提供的资助，以及香港科技大学（广州）MetaHKUST 项目的支持。此外，本研究还部分得到香港创新科技署在创新及科技基金框架下提供的资助，资助编号为 ITS/319/22FP。我们衷心感谢 AvatarSDK 的慷慨赞助，其为我们提供了为期一年的虚拟形象生成 API 使用权限。

7. 参考文献

1. H. Dong 和 J. S. A. Lee，《从多媒体通信视角看元宇宙》，*IEEE MultiMedia*，第29卷，第4期，第123–127页，2022年。2. S.-C. Chen，《面向元宇宙的多媒体研究》，*IEEE MultiMedia*，第29卷，第1期，第125–127页，2022年。3. D. Dermawan、R. P. Wibawa 和 M. D. E. Susanti，《COVID-19疫情期间提案／论文答辩中虚拟会议使用情况分析》，2020年。4. Meta。《Meta Horizon 虚拟化身最佳实践》。Meta for Developers。在线；2025-07-10。 [在线]。可访问：https://developers.meta.com/horizon/design/bp-avatars/5. A. Alhilal、K. Shatilov、G. Tyson、T. Braud 和 P. Hui，《元宇宙中的网络流量：以社交虚拟现实为例》，载于*2023*年 *IEEE* 第*43*届分布式计算系统国际会议研讨会（*ICD- CSW*），2023年，第109–114页。6. K. Shatilov、A. Alhilal、T. Braud、L.-H. Lee、P. Zhou 和 P. Hui，《玩家尚未准备好 101：元宇宙中的混合模式活动组织教程》，载于《第一届元宇宙系统与应用研讨会论文集》，系列：MetaSys '23。美国纽约：计算机协会，2023年，第14–20页。 [在线]。可访问：https://doi.org/10.1145/3597063.3597360 7. K. Y. Lam、L. Yang、A. Alhilal、L.-H. Lee、G. Tyson 和 P. Hui，《元宇宙中的人类—虚拟形象交互：全身交互框架》，载于《第4届 ACM 亚洲多媒体国际会议论文集》，系列：MMAsia '22。美国纽约：计算机协会，2022年。 [在线]。可访问：https://doi.org/10.1145/3551626.3564936 8. A. Adkins、A. Normoyle、L. Lin、Y. Sun、Y. Ye、M. Di Luca 和 S. Jörg，《从猜字游戏的视角看，精细手部动作对虚拟角色交流有多重要？》《ACM 图形学汇刊》，第42卷，第3期，2023年5月。[在线]。可访问：https://doi.org/10.1145/3578575 9. B. Beatty，《混合灵活课程设计：实施学生主导的混合课程》。EdTech Books，英国伦敦，2019年。10. C. C. for Teaching and Learning。（2023年）混合式／HyFlex 教学与学习。哥伦比亚大学。[在线]。可访问：https://ctl.columbia.edu/resources-and-technology/teaching-with-technology/teaching-online/hyflex/ 11. A. Raes、L. Detienne、I. Windey 和 F. Depaepe，《同步混合式学习系统性文献综述：发现的研究空白》，《Learning Environments Research》，第23卷，第269–290页，2020年。12. S. Allen 和 P. Williams，《管理学、领导力和商业领域的远程博士论文口头答辩：优势、教学法与实践》，《Management Teaching Review》，第7卷，第3期，第208–225页，2022年。 [在线]。可访问：https://doi.org/ 10.1177/23792981211013024 13. G. W. W. Park，《设计赋能远程会议参与者的机制，以促进更有效、更有帮助的会议》，载于《CHI 人机交互国际会议扩展摘要论文集》，系列：CHI EA '25。美国纽约：计算机协会，2025年。 [在线]。可访问：https://doi.org/10.1145/3706599.3721097。

14 G. W. W. Park, A . Tang 和 F . Chevalier，《Jollygesture：探索虚拟现实演示中的双用途手势和手势引导》，载于第*50*届图形界面会议论文集，GI '24 系列。美国纽约州纽约：计算机协会，2024 年。 [在线]。可获取：https://doi.org/10.1145/3670947.3670965

15. J.Chen、J.Hu、G. Wang、Z.Jiang、T.Zhou、Z. Chen、

以及 C. Lv，“Taoavatar：通过 3D 高斯泼溅实现增强现实中的实时逼真全身会说话的虚拟化身”，载于计算机视觉与模式识别会议论文集（*CVPR*），2025年6月，第 10 723–10734 页。

Ahmad Alhilal 是芬兰阿尔托大学的博士后研究员。他曾在香港科技大学（HKUST）担任博士后研究员。目前的研究兴趣包括网络感知流媒体、可扩展计算以及云渲染移动/XR游戏。Alhilal博士在HKUST获得计算机科学与工程博士学位。他是IEEE计算机学会会员。联系邮箱：ahmad.alhilal@aalto.fi。

Kit Yung Lam 是瑞士罗特克罗伊茨卢塞恩计算机科学与信息技术学院的研究助理。他的研究兴趣包括增强现实、多模态交互和沉浸式媒体体验。Lam博士在香港科技大学获得计算机科学与工程博士学位。联系邮箱：kityung.lam@hslu.ch。

Lik-Hang Lee 是香港特别行政区香港理工大学的助理教授。目前的研究兴趣包括混合现实、人机交互、大语言模型和教育技术。Lee博士在香港科技大学获得计算机科学与工程博士学位。他是IEEE高级会员。联系邮箱：lik-hang.lee@polyu.edu.hk

Xuetong Wang 是香港科技大学（HKUST）的博士学位候选人。她的研究兴趣包括人机交互、扩展现实和人工智能。Xuetong 获得了香港科技大学电子工程专业的工程学学士学位。联系邮箱：xwangdd@connect.ust.hk。

Sijia Li 是英国剑桥的一名独立研究人员。她的研究兴趣包括人机交互、教育人工智能和高效学习。Sijia 获得了香港科技大学（广州）计算媒体与艺术专业的硕士学位。请通过 sli308@connect.hkust-gz.edu.cn 联系她。

Matti Siekkinen 是芬兰埃斯波阿尔托大学的一名大学讲师。他的研究兴趣包括云加速虚拟现实与增强现实、分布式多媒体系统和移动计算。Siekkinen 博士获得了法国 Eurecom／尼斯-索菲亚·安蒂波利斯大学的计算机科学博士学位。请通过 matti.siekkinen@aalto.fi 联系他。

Tristan Braud 是香港科技大学（HKUST）整合系统与设计学部的助理教授，并领导扩展现实与沉浸式媒体（XRIM）实验室。他的研究探索泛在计算与以人为中心的系统设计的交叉领域，重点关注将增强现实和虚拟现实作为主要交互媒介。Braud 博士在法国格勒诺布尔阿尔卑斯大学获得计算机科学博士学位。联系邮箱：braudt@ust.hk

Pan Hui 是香港科技大学（广州）计算媒体与艺术讲席教授、元宇宙与计算创造力中心主任，同时也是香港科技大学新兴跨学科领域讲席教授。他的研究兴趣包括数据驱动系统设计和沉浸式人类-数据交互。Hui 教授在剑桥大学计算机实验室获得博士学位。他是英国皇家工程院国际院士（FREng）、欧洲科学院成员（MAE）、IEEE会士，也是计算机协会（ACM）杰出科学家。联系邮箱：pan-hui@ust.hk。