Engineering For Social Change
A. Cooper et al. (Eds.)

doi:10.3233/ATDE240947

Extended Reality Methods for Transdisciplinary Asynchronous Engineering

Anjela MAYER[a,1], Jivka OVTCHAROVA[a], Jean-Rémy CHARDONNET[b], Sebastian AMANN[c], Nicolai BEISHEIM[c] and Xianbiao JIANG[c]
[a]*Karlsruhe Institute of Technology, Germany*
[b]*LISPEN, HESAM Université, Arts et Metiers Institute of Technology, France*
[c]*Albstadt-Sigmaringen University, Albstadt, Germany*

Abstract. In the evolving landscape of engineering collaboration, Extended Reality (XR) demonstrates transformative potential for transdisciplinary work. XR approaches are promising in enhancing flexibility and efficiency in diverse, global work settings by facilitating remote and asynchronous collaboration. This paper introduces XR methods designed to address challenges such as communication gaps and project misalignment in asynchronous collaboration. Building upon our previous work, which introduced Avatar Replay for non-present users, this paper presents an extension that enables interaction with these users through their avatars. This development promotes a more dynamic and responsive approach to asynchronous collaboration. Additionally, the paper shares insights from a test scenario using a Head-Mounted Display (HMD) with an improved prototype in a transdisciplinary guidance scenario. The findings of this research highlight the significant potential of XR in enhancing collaboration and understanding across disciplines, particularly in asynchronous collaboration between experts and non-experts in the engineering domain. This enhanced interaction across realities not only promises increased societal acceptance of immersive technologies but also signifies a transformative shift in the way we work together, seamlessly blending virtual and real-life interactions.

Keywords. Interdisciplinary Engineering, Transdisciplinary Collaboration, Asynchronous Collaboration, Extended Reality

Introduction

In the contemporary engineering where diverse disciplines converge working on shared spatial tasks, innovative approaches to enable and enhance collaboration despite different knowhow and understanding is required to bring all together. As projects become increasingly global and complex, the ability to work effectively across local boundaries has emerged as a critical competency. Extended Reality (XR) stands at the forefront of technological advancements poised to redefine the paradigms of team interaction, offering unprecedented opportunities for immersive and intuitive coworking. This paper explores XR's transformative potential in facilitating asynchronous collaboration on

[1] Corresponding Author, Email: anjela.mayer@kit.edu.

spatial tasks and demonstrates a use case between an expert and a non-expert, who are working at different times with an industrial robot.

Despite the rapid evolution of synchronous collaboration tools, asynchronous collaboration is still underrepresented in research, although it offers advantages like time-flexibility and the documentation of work and processes. In addressing this notable gap, our work introduces an Extended Reality (XR) methodology enabling asynchronous collaboration on spatial engineering tasks by capturing the work of the XR users and reproducing it using virtual replicas. XR serves as an umbrella term for the immersive technologies Virtual Reality (VR), Augmented Reality (AR) and Mixed Reality (MR). Its potential lies in its ability to render visible the otherwise invisible, such as digital metadata, virtual objects in the physical space, as well as virtual representations of non-present team members. Consequently, XR technologies significantly enhance spatial collaborative work, like in the engineering domain.

In the engineering field, digital twins of machines and systems are being well researched, however the role of human operators is often left undigitalized and undocumented. Our approach for asynchronous collaboration using XR and full-body avatars is especially suitable for the documentation of human operations in physical spaces. Human operations in their physical environment and cyber-physical machines can be captured and reconstructed using virtual replicas and avatars, fostering possibilities to collaborate with non-present coworkers.

Following the introduction, section 2 is presenting related work in capturing and reconstructing human operations. The concept of our method is presented in section 3, and in section 4 its implementation. Section 5 presents and discusses the preliminary results, while section 6 concludes our work and provides an outlook for future work.

1. Related Work

For the documentation of human operations, especially those involving 3D or spatial operations within physical spaces, like during manufacturing reconfiguration, immersive technologies are very suitable. Considering the XR continuum, most of the related work utilize VR, like Wang et al. who present a system which captures collaborative VR sessions where users are represented by abstract full-body avatars [1]. This opens the possibility to relive the VR group experience anytime. They find that users experience a higher level of immersion and social presence compared to 360° video playback both in VR or on 2D display. However, the playback of the experience is not interactable and users cannot act when watching the playback. Yokoyama et al. present VR2ML, a method to use the recorded 3D motion data of VR users and use them for machine learning (ML) [2]. ML can be utilized to enhance the VR experience for single users as well as in group work, and opens possibilities to make avatar paybacks more responsive and interactable.

A multimodal recording of VR users is explored by Chow et al. [3]. Their avatars are simple, consisting of an abstract head, upper body and hands. Besides the avatars, also verbal speech, user interactions with the VR environment, such as object manipulations and teleportation, and annotations are also recorded. During the playback users can still interact with their environment, as well as control the playback, with functions like pause and rewind. Their work is evaluated in a user study and they summarize several recommendations for the implementation of avatar playback in collaborative session. For instance, they recommend to use a minimap, with a top-down view on the VR environment, as well as an x-ray vision, to be able to follow the replayed avatars. Furthermore, since users are sensitive to proximity of the playedback avatar, they recommend to limit teleportation of VR characters so that they cannot teleport onto the same position (into each other).

In XR-Live, Thanyadit et al. play back a recorded chemistry instructor to students within a user study [4]. In their work they experiment with several visualizations, a duplicate of the workbench with the upper-body avatar demonstrating the experiment, as well as, a shadow of the instructor's hands and tools superimposed on the student's workspace. Furthermore, they replaced teleportation of the avatar by a walking animation (although their avatars do not have legs), to make following the instructor easier and added temporal tools like auto-pause function to create a step-by-step learning as well as a checklists for the tasks. In their validation study they could reduce the workload and increase the learning performance compared to the avatar only condition.

In the work of Bruza et al. also abstract upper-body avatars are utilized for content creation in education [5]. The avatars can be recorded and replayed including their interactions with a simple VR environment. Furthermore, they present a feature to create educational 2D videos in the VR environment, by defining trajectories for a virtual camera. The capture of the virtual camera was used to create videos for biochemistry and medicine education.

Besides VR, a few works enable user capture and avatar playback in AR and MR. Wang et al. present CAPturAR, which facilitates upper-body avatar capture and a linking feature of the recordings to context in AR [6]. For instance, user can select detected objects and link then with specific recordings. This method can be used to detect user actions and trigger linked events in AR, e.g. start a digital timer. They emphasize, their system can be used for creating demonstrations and tutorials by the avatar. A drawback is the hardware that the user has to wear in a backpack for using the system. Fender et al. present a stationary system within a work space, consisting of volumetric cameras, that capture a physical work space, and a VR headset to visualize and playback the captures [7].

A user study on social presence comparing avatar playback with 2D video and 180° volumetric playback of users is presented in the work of Cho et al. [8]. They find out that volumetric user representation is best regarding social presence in dynamic spaces, and together with video also best for static task, where the users do not leave their workbench. In their work they scanned an actor to create a photorealistic avatar, which did not pass through uncanny valley. Furthermore, the played back user recording was stationary and not moving through the 3D space, which limits the eligibility of their results for more dynamic use cases.

The preservation and recreation of human motions and interactions is a key methodology in asynchronous collaboration. Using XR technologies to capture and digitalize humans and their operations adds the human component towards holistic Digital Twins of processes and not just machines or systems. Furthermore, the preservation of these captures open the possibility to relive the operations at any time and adds a new possibility to document human tasks more easily and quickly. In this work, our process towards creating an XR prototype to capture and document human work within spatial physical tasks is presented. Our prototype represents the users by full-body avatars in Mixed Reality environments. Thus, the virtual objects are visualized within the real environment and the user can see and interact with both, the physical environment and with the virtual avatars.

2. Concept

The concept of this work builds upon our preliminary efforts, introducing significant enhancements to address the challenges identified in earlier studies. In the following subsections our preliminary work, as well as our novel contributions are described.

2.1. Preliminary Work

In our preliminary efforts, we established a foundational system for capturing and replaying human operations using XR technologies, as detailed in our prior publications [10, 11]. Utilizing the Oculus Quest 1 VR Head Mounted Display (HMD) and the HoloLens 2 AR HMD, we captured the head and hand poses of users, along with their interactions with both virtual and physical objects. These operations were then reproduced, allowing for the replay of captured actions using virtual representations of the user (avatar) and objects (3D meshes).

In the VR context, users' HMD positions and orientations were captured per frame alongside the poses of the Oculus Touch Controllers, which were visually represented as virtual hands. This enabled users to interact with and manipulate virtual assembly parts in an intuitive manner. For AR, without the use of physical controllers, Hololens 2's hand tracking capabilities were employed to capture natural hand movements and interactions within the user's actual environment.

Recordings were encouraged to be brief and focused on individual assembly steps rather than extensive processes, enhancing usability and collaborative efficiency. These recorded steps, once saved into text-based files, could be imported and replayed by others, offering immersive, step-by-step instructional content. Furthermore, to facilitate user interaction during the replay phase, immersive buttons for recording control and step navigation were implemented within the virtual workspace, allowing users to intuitively manage their experience through direct hand interactions.

2.2. Novel Contribution

In advancing our work on the XR avatar record and replay system, we aim to introduce several significant enhancements to improve collaborative experiences in asynchronous engineering settings. This iteration is focused on extending the system with full-body avatars, utilizing inverse kinematic models and body posture approximations alongside tracked head and hand poses. This development is designed to enrich the comprehension of body movements and augment the sense of presence for users not physically present, all while ensuring the system's compatibility with portable HMD devices to maintain ease of use and accessibility.

Further enhancements include the integration of individual, personalized avatars, moving beyond the use of a single abstract avatar to offer a more tailored and immersive user experience. This approach aims to increase user engagement and identification within the virtual environment. To complement the visual enhancements, we also plan to incorporate verbal speech recording, enabling a multimodal replay, as recommended in [3], that encompasses both visual and auditory information. This feature is intended to facilitate a more comprehensive understanding of spatial tasks, making instructions clearer and more accessible to users.

To improve the user interaction with the system, we propose the implementation of a self-haptic feedback mechanism, as was presented in [9], for controlling replay functions. By integrating interactive buttons directly onto the user's body, such as on the arm, users will experience tactile feedback, enhancing the intuitiveness and responsiveness of the system's interface. These planned enhancements collectively contribute to a more immersive, accessible, and effective system for documenting and replaying human operations in XR, addressing key challenges in asynchronous collaboration across disciplines. The specifics of these implementations and their impact on the user experience are detailed in the following sections.

3. Implementation

In this section the implementation extension of the preliminary work is summarized. While in the previous publications the system was implemented in Unreal Engine (UE) 4.27, the recent versions are implemented in UE 5.3, which is required to use the most recent plugins and features of the engine.

3.1. Full-body Avatar Extension

To implement the full-body avatar the MetaXR plugin was utilized which enables the utilization of Meta's Movement SDK[2] and allows to access the full-body tracking of the device, along with body, gaze, finger tracking as well as facial expressions. The UE Mannequin was used as the 3D model of the full-body avatar. For the animation of the avatar an Animation Blueprint was created in the UE editor which receives tracking information from the HMD and Movement SDK and utilizes LiveLink[3] along with the MetaXR animation retarget assets to animate the avatar. Since the skeletal structure and coordinate system of the UE Mannequin, depicted in Figure 1, and the tracked MetaXR system, depicted in Figure 2, are different, the received tracking data has to be mapped to corresponding bones and transformed to fit the coordinate system.

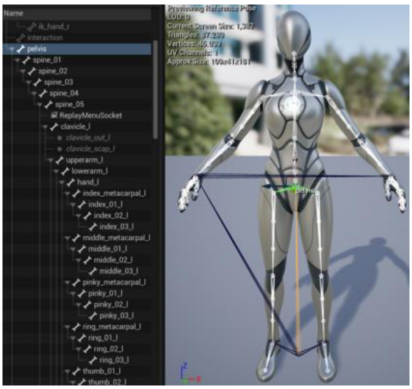

Figure 1. Skeletal structure of the Unreal Engine Mannequin characters.

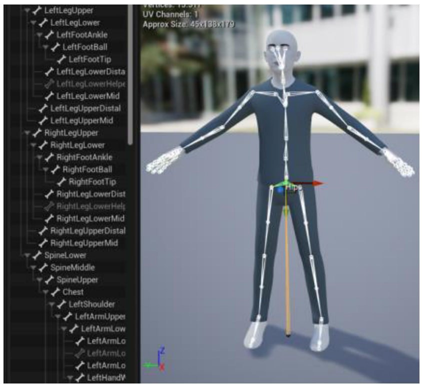

Figure 2. Skeletal structure of MetaXR Movement SDK character "Proteus".

[2] Movement SDK: https://developer.oculus.com/documentation/unreal/unreal-movement-implementation
[3] Unreal Engine LiveLink Plugin: https://dev.epicgames.com/documentation/en-us/unreal-engine/live-link-in-unreal-engine?application_version=5.3

The required adaptations are defined in the animation retarget assets when using the LiveLink method. In the used MetaXR plugin v60.0 the animation retarget of the hands was not precise, resulting in a visible offset between the real hands and the avatar hands in Passthrough AR mode. To improve accuracy, hand tracking poses were additionally passed to the inverse kinematic solver in the avatar Animation Blueprint.

3.2. Multimodal Record and Replay of Human Motions

To be able to record and replay the full-body motions, the record data protocol was extended to include all body-joint poses, instead of just the HMD pose for the head and the finger-joint poses for the hands. Since, the Movement SDK also provides parametrized facial expression information, the data protocol was extended by the facial expression struct which stores 50 floats values animating the characters face. The file exporter/importer was extended accordingly to parse the new information.

The Replay Avatar, which is the virtual replica of the user's avatar used during replay of the recorded user, was also extended by a full-body skeletal mesh and an Animation Blueprint handling the skeletal animation during playback. The body-joint poses are read from the record file and passed to the animation Blueprint where they are applied to the corresponding body-joints animating the character.

For the audio record and replay, of the user's verbal input, the Runtime Audio Importer[4] plugin was used. It implements features to record audio, as well as, mp3 and wav export and import during application runtime.

3.3. Personalized Avatars

In our project, the creation of personalized avatars was achieved through a blend of technologies including Polycam[5], Blender, and the UE Metahuman Plugin[6], facilitating the development of personalized Metahuman avatars. The process began with capturing 80 to 100 images of the subject using a smartphone equipped with Polycam in a well-lit setting to ensure a comprehensive capture of the subject's features. These images were processed by Polycam to create a detailed 3D surface mesh, as depicted in Figure 3, which was then refined in Blender by removing noise before exporting it in .FBX format. Within the Metahuman Creator, the facial mesh was integrated onto an adjustable 3D character, and the resulting Metahuman, as depicted in Figure 4, was imported into our UE project via the UE's Quixel Bridge. Given that Metahumans share the skeletal structure of the UE Mannequin, animation adaptation was seamless, allowing for a highly detailed and lifelike digital representation of the user. However, for applications on standalone devices like the Meta Quest Pro, we simplified the hair representation using cards to maintain performance, and recommended disabling shadow casting for the grooms to enhance performance. Moving forward, the intention is to incorporate the replay of facial expressions for even more realistic avatar interaction.

---

[4] Runtime Audio Importer Sourcecode: https://github.com/gtreshchev/RuntimeAudioImporter
[5] Polycam 3D Scan App: https://poly.cam
[6] Unreal Engine Metahuman Creator: https://www.unrealengine.com/en-US/metahuman

Figure 3. 3D reconstruction of a person with the Polycam App.

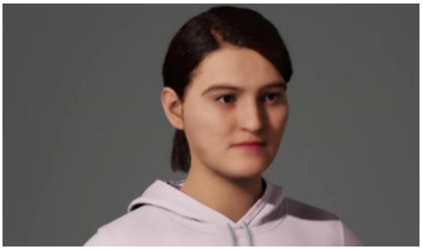

Figure 4. Applied face mesh of a person on a 3D character using Epic Game's Metahuman Creator.

4. Preliminary Results and Discussion

This section describes a test application of the software in the environment of an articulated robot cell in the machine hall of the Albstadt-Sigmaringen University of Applied Sciences and then discusses the overall project.

4.1. Robot Maintenance Use Case

In the machine hall of Albstadt-Sigmaringen University, robots are operated within a safety cell, utilized for both student internships and research projects. During operation, various problems occasionally arise unpredictably, for which suitable solutions are typically known. These are usually documented in the form of written instructions, possibly supplemented with images or graphics, and can be consulted and implemented step-by-step when needed.

To simplify maintenance documentation and increase efficiency in both creating the documentation and carrying out troubleshooting according to maintenance instructions, the application presented in this paper was tested in the robot safety cell use-case. Specifically, the application was installed as standalone software on a portable HMD (Oculus Quest 2), and the six necessary steps to re-enable the gripper on the robot after a special problem were recorded. Figure 5 illustrates the robot expert engaged in the procedure, while Figure 6 captures the resulting avatar in AR replay (the background appears black when recording passthrough AR with the Oculus Quest).

Figure 5. Robot expert recording a maintenance step in the physical environment.

Figure 6. Avatar replay of the recorded maintenance step.

The six recorded steps each had a duration of no more than 30 seconds. Even though, in these tracks there were slight but noticeable deviations between the recorded avatar positions and their corresponding audio. Subsequently, the scenario of a defective gripper was simulated, and a user without robot experience was tasked using the HMD and the created avatar replays to perform the maintenance, as depicted in Figure 7. Once again, difficulties with the user interface, illustrated in Figure 8, were encountered, but the user was able to execute the steps independently and without further assistance. The accuracy of the recorded avatar proved sufficient for rough mechanical steps, although it is assumed that the software may reach its limits for very precise applications. For example, on the Meta Quest 2 headset used, distinguishing between two adjacent cables in the avatar replay would likely be challenging.

Figure 7. User maintaining a robot cell using a HMD with assistance software.

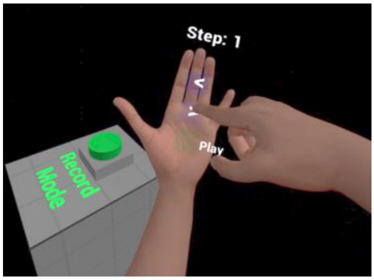

Figure 8. User Interface in Virtual Reality perspective.

4.2. Discussion

Building on our foundational work in VR and AR for capturing head and hand poses, our current research significantly advances the XR system's capabilities for asynchronous collaboration. This advancement introduces full-body avatars using inverse kinematics and body posture approximations, significantly progressing beyond the limitations of partial avatars documented in related studies work [1, 4, 5]. Our approach not only enhances the sense of presence for non-present users but also facilitates a deeper understanding of complex body movements, addressing a gap in the immersive experience previously noted in [1, 4].

Furthermore, we have integrated personalized avatars into our system, moving beyond the generic representations found in related studies [1, 5]. This change provides a more immersive and engaging user experience, enhancing the relatability and interaction among users within the virtual environment. The incorporation of verbal speech recording enables a richer, multimodal replay of spatial actions like proposed in [3]. This makes the instructions of complex tasks more accessible and easier to follow.

To refine user interaction, we have implemented a self-haptic feedback mechanism for UI controls like proposed in [9], that enhances user engagement through tactile interaction. This development simplifies the user interface and distinguishes our work from the broader field, where such tactile feedback mechanisms are not extensively explored.

Our focus on portability and accessibility ensures that the system remains practical and applicable across a wide range of interdisciplinary settings, running on portable HMD devices without the need for external systems and additional computing resources.

This approach addresses the limitations presented by more complex setups like in [6, 7], reinforcing our commitment to accessibility and ease of use.

However, from our tests, we also want to share more practical implications. The system's capacity for capture and replay functioned well, underscoring its potential for real-world application. Yet, a divergence in long audio and avatar playback synchronization was noted, suggesting future iterations should incorporate timestamps for enhanced synchronization. Our recommendation for users to make short, step-wise recordings rather than longer ones for ease of collaboration and modification was validated in our testing. While the self-haptic UI design aimed to be both accessible and unobtrusive, challenges in button interaction indicated room for improvement, particularly in providing clear visual feedback for activation states, which can be simply improved by adjusting colours or transparency.

Passthrough AR testing showcased limitations in text readability, which is depicted in Figure 9, and colour accuracy, shown in Figure 10. These observations highlight the need for better passthrough AR solutions or the integration of digital twins to enhance interaction fidelity, since it is essential for interaction with both analogue and digital interfaces.

Figure 9. Passthrough AR view on a display containing written text and widgets.

Figure 10. Color lagging resulting in a part of the user’s arm appear in greyscale.

Looking forward, the integration of digital twins of the robot and virtual replicas of objects within the work environment would enable a more comprehensive documentation of human operations and interactions. This future direction aims to extend the system's utility beyond current capabilities.

In registration, challenges arose due to the HMD's limitations in object and marker detection, leading to innovative workarounds like highlighting room characteristics for 3D model creation. This feature, while effective for the tests, suggests the exploration of external cameras and detection algorithms for use cases relying on registration.

5. Conclusion and Outlook

This study advances the application of Extended Reality (XR) in transdisciplinary asynchronous engineering by introducing full-body avatars, personalized interactions, and multimodal communication. Through enhancing the virtual presence and facilitating detailed understanding of complex tasks across disciplines, our work addresses key challenges in engineering collaboration, such as bridging the gap between diverse expertise and aligning project goals without the need for co-presence. Preliminary testing highlights the system's potential to significantly improve collaborative workflows and knowledge sharing among interdisciplinary teams. Future developments will focus on refining the system's integration of verbal and non-verbal communication cues and expanding the use of digital twins for a more comprehensive virtual collaboration environment. By pushing the boundaries of current XR technology, our research aims to foster more intuitive, effective, and human-centric collaboration in transdisciplinary engineering fields, paving the way for innovative solutions to complex engineering challenges.

References

[1] C. Y. Wang, M. Sakashita, U. Ehsan, J. Li, and A. S. Won, Again, Together: Socially Reliving Virtual Reality Experiences When Separated, *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, April 2020, pp. 1–12, doi: 10.1145/3313831.3376642.

[2] Y. Yokoyama and K. Nagao, VR2ML: A Universal Recording and Machine Learning System for Improving Virtual Reality Experiences, *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2020, pp. 756–757, doi: 10.1109/VRW50115.2020.00229.

[3] K. Chow, C. Coyiuto, C. Nguyen, and D. Yoon, Challenges and Design Considerations for Multimodal Asynchronous Collaboration in VR, *Proc. ACM Hum.-Comput. Interact.*, vol. 3, CSCW, 2019, doi: 10.1145/3359142.

[4] S. Thanyadit, P. Punpongsanon, T. Piumsomboon, and T.-C. Pong, XR-LIVE: Enhancing Asynchronous Shared-Space Demonstrations with Spatial-temporal Assistive Toolsets for Effective Learning in Immersive Virtual Laboratories, *Proc. ACM Hum.-Comput. Interact.*, vol. 6, CSCW1, 2022, doi: 10.1145/3512983.

[5] Vojtěch Brůža, Jan Byška, Jan Mičan, and Barbora Kozlíková, VRdeo: Creating engaging educational material for asynchronous student-teacher exchange using virtual reality, *Computers & Graphics*, vol. 98, pp. 280–292, 2021. doi: 10.1016/j.cag.2021.06.009.

[6] T. Wang *et al.,* CAPturAR: An Augmented Reality Tool for Authoring Human-Involved Context-Aware Applications, *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 2020, pp. 328–341, doi: 10.1145/3379337.3415815.

[7] A. R. Fender and C. Holz, Causality-preserving Asynchronous Reality, *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, doi: 10.1145/3491102.3501836.

[8] S. Cho, S. Kim, J. Lee, J. Ahn, and J. Han, Effects of volumetric capture avatars on social presence in immersive virtual environments, *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020, pp. 26–34, doi: 10.1109/VR46266.2020.00020.

[9] C. M. Fang and C. Harrison, Retargeted Self-Haptics for Increased Immersion in VR without Instrumentation, in *The 34th Annual ACM Symposium on User Interface Software and Technology*, 2021, pp. 1109–1121, doi: 10.1145/3472749.3474810.

[10] A. Mayer, A. Rungeard, J. -R. Chardonnet, P. Häfner and J. Ovtcharova, Immersive Hand Instructions in AR – Insights for Asynchronous Remote Collaboration, *2023 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 2023, pp. 1-6, doi: 10.1109/CIVEMSA57781.2023.10231018.

[11] A. Mayer, T. Combe, J.-R. Chardonnet and J. Ovtcharova, Asynchronous Manual Work in Mixed Reality Remote Collaboration, *Extended Reality, Lecture Notes in Computer Science*, vol. 13446, Springer, Cham, 2022, doi: 10.1007/978-3-031-15553-6_2