En Cotutelle entre
l'Université de Lille (École graduée MADIS) et McGill University

Thèse

pour obtenir le grade de :

Docteur

dans les spécialités

« Informatique » et « Music Technology »

par

Pierrick Uro

Interaction, Space, and Copresence in Co-Located Mixed Reality

Thèse soutenue le 18 Novembre 2025 devant le jury composé de :

M.	Laurent Grisoni	Professeur, Université de Lille,	(Co-directeur)
M.	Marcelo Wanderley	Professeur, McGill University,	(Co-directeur)
M.	Florent Berthaut	Maître de Conférence, Université de Lille,	(Co-encadrant)
M.	Thomas Pietrzak	Professeur, Université de Lille,	(Co-encadrant)
M.	Cédric Fleury	Maître de Conférence, IMT Atlantique,	(Rapporteur)
M.	Nick Bryan-Kinns	Professeur, University of the Arts London,	(Rapporteur)
Mme.	Caroline Appert	Directrice de Recherche, Université Paris Saclay,	(Présidente)
Mme.	Tifanie Bouchara	Maîtresse de Conférence, Université Paris Saclay,	(Examinatrice)

Région Hauts-de-France — Université de Lille — McGill

2026/01/07

Résumé

La Réalité Mixte ouvre de nouveaux moyens de fusionner les espaces physiques et virtuels, mais son usage dans le domaine du spectacle soulève des défis uniques autour des environnements partagés. Cette thèse se concentre sur la Réalité Mixte colocalisée, où les participants partagent à la fois un espace physique et virtuel, et examine comment les choix de conception influencent la collaboration et la présence.

La recherche s'articule autour de trois axes. D'abord, sur la base de ma participation à la performance théâtrale en Réalité Mixte Animate, une analyse thématique réflexive a été menée sur plusieurs productions afin d'identifier les défis récurrents et les stratégies employées par les créateurs. Ensuite, la conception et l'évaluation de gRAinyCloud, un instrument de musique collaboratif en Réalité Mixte, ont permis d'étudier les conséquences du découplage des espaces physiques et virtuels sur l'interaction collaborative. Enfin, une étude de la coprésence en Réalité Mixte colocalisée a proposé un protocole combinant auto-évaluation et mesures comportementales, offrant une manière plus nuancée d'évaluer la coprésence.

Ce travail allie la recherche basée sur la performance, des entretiens, l'analyse thématique réflexive, le prototypage, l'auto-étude et des expériences contrôlées, illustrant la valeur de la diversité méthodologique dans la recherche sur la Réalité Mixte. Ses contributions comprennent un vocabulaire de stratégies de conception, une démonstration du découplage spatial comme principe de conception et de nouveaux outils pour évaluer la coprésence. Ensemble, elles étendent l'espace de conception de la Réalité Mixte et procurent des ressources pour créer des expériences plus interactives, créatives, et inspirantes.

Abstract

Mixed Reality offers new ways to blend physical and virtual environments, but its use in performance raises unique challenges around shared environments. This thesis focuses on co-located Mixed Reality, where participants share both a physical and a virtual space, and examines how design choices shape collaboration and presence.

The research is structured along three axes. First, grounded in my participation in the Mixed Reality theater performance Animate, a reflexive thematic analysis was conducted across several productions to identify recurring challenges and strategies employed by creators. Second, the design and evaluation of gRAinyCloud, a collaborative Mixed Reality musical instrument, investigated the consequences of decoupling physical and virtual spaces on collaborative interaction. Third, a study of copresence in co-located Mixed Reality proposed a protocol that combines self-report and behavioral measures, offering a more nuanced way to assess copresence.

This work combines performance-led research, interviews, reflexive thematic analysis, prototyping, self-study, and controlled experiments, illustrating the value of methodological diversity in Mixed Reality research. Its contributions include a vocabulary of design strategies, a demonstration of spatial decoupling as a design principle, and new tools for evaluating copresence. Together, they extend the design space of Mixed Reality and provide resources for creating more interactive, creative, and inspiring experiences.

Acknowledgements

The work presented in this thesis was funded in part by the region Hauts de France and the Front de Recherche du Québec.

My thesis is marked by the fairly complex endeavor of making two academic systems collaborate across an ocean, and I want to thank everyone who made this cotutelle between the universities of Lille and McGill both possible and fruitful, in particular my supervisors. Thanks to the jury as well for the time and effort they put into helping me improve this work. I thank my colleagues from IDMIL and MINT, who provided me with their attention, advice, and participation in my research, as well as everyone I worked with on *Animate*, which opened my eyes to the potential that art and science bring to each other.

Thank you to everyone in my personal life, friends and family, who assisted me in one way or another in enduring these four years and completing my thesis despite my insistent overthinking. A special thank you to my roommate in Montréal, Cléo, who helped me settle on an entirely new continent at the very beginning of my thesis, and still feel like I was home.

Thanks to conference organizers and reviewers for the exchanges they enable, to conference goers for the work they present and discussions we had, and to everyone that listened to my presentations with keen interest and inspiring questions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Mixed Reality (MR) is increasingly used in artistic practices [92], offering new possibilities for interaction by blending physical and virtual spaces, reaching beyond the traditional stage. This potential, however, comes with unique challenges. Designing a Mixed Reality Performance (MRP) demands a constant balancing act between artistic intent and technological constraints, addressing issues of immersion, usability, and performance limitations. While designers develop strategies to mitigate these conditions, they remain largely undocumented, making it difficult to build a broader understanding of design practices in this field. To address these gaps, this thesis draws on a deliberately diverse methodological palette, spanning performance-led research, reflexive thematic analysis of interviews and literature, and controlled experiments, in order to capture both the theoretical context and the practical experience around the design of performances and collaborative interaction in MR.

This thesis focuses specifically on Co-located Mixed Reality (CMR), which is where multiple participants share both a physical and virtual space. The context of CMR raises questions of spatial alignment, collaboration, and the distribution of attention between real and virtual elements. Current systems often rely on a tight coupling between physical and virtual environments, which ensures spatial consistency, but also limits creative possibilities [33]. In collaborative musical performances, for instance, this alignment constrains movement and interaction, restricting how performers and audiences can engage with the work. Decoupling physical and virtual spaces emerges as a promising alternative, but its implications for collaboration and design remain underexplored.

Another key concern of CMR is the experience of *copresence*, or the sense of being together in a shared virtual space [20]. Copresence is crucial for collaboration and performance [78], yet its measurement remains challenging. Most existing approaches rely on subjective questionnaires, which provide valuable insight but risk overlooking the subtleties of real-time interactions [52]. Behavioral measures, such as movement patterns and spatial adjustments, could offer complementary perspectives, but their application in CMR remains limited.

This thesis addresses these challenges along three complementary axes, all inspired by my deep implication in the MRP *Animate*[1]. As part of the production team from the beginning of my thesis, I was initially given the role of technical assistant but quickly took on more responsibilities and weighed in on more than strict implementation decisions, before outright co-directing performances in Weimar in September 2022. This irreplaceable firsthand experience allowed me to witness and take part in the design and implementation processes of a live MRP production, along with the challenges that arise, and the strategies deployed in response. Building on this experience, I sought to find whether these strategies had also been used in other MRPs despite the stark difference in technological and artistic contexts in this field. Thus, the first axis of this thesis examines similar productions as *Animate* to identify the recurring challenges MRPs encounter and the strategies creators use to overcome them within their specific contexts.

The second axis revolves around the decoupling of physical and virtual spaces in collaborative musical performances. It applies a strategy identified in the previous axis to the adjacent context of musical performance through the design and implementation of gRAinyCloud as a case study. This collaborative CMR musical instrument serves as an illustration of how allowing multiple collaborators to individually manipulate their own viewpoints in relation to a virtual interface can enable new and exclusive creative behaviors and strategies. The design and implementation are described, along with a self-study to identify emerging collaborative practices and design guidelines.

The third axis is a direct consequence of a design choice in *Animate*, where the audience wore MR headsets and interacted with performers that wore no headset. Because of a learned choreography and

[1] https://timothythomasson.com/Animate

their improvisation skills, the two performers were able to act as if they could perceive the same virtual elements that the participants saw through their headsets. This raises a question: to what extent did these participants feel like the performers were sharing the virtual environment with them, and what could influence this feeling? However, a first necessary step before being able to answer is to define a reliable and objective way to measure this feeling of copresence. This is way methods for evaluating copresence in CMR are explored, combining a questionnaire with behavioral measures to develop protocols that can better capture the dynamics of shared virtual environments.

Taken together, these three axes form a trajectory that moves from analysis to design to evaluation (see Fig. 1.1). By first identifying strategies across existing MRPs, then experimenting with new possibilities through the design of gRAinyCloud, and finally developing ways to assess copresence, this thesis builds a multi-layered perspective on how CMR can support collaboration and performance. The following chapters develop this trajectory step by step.

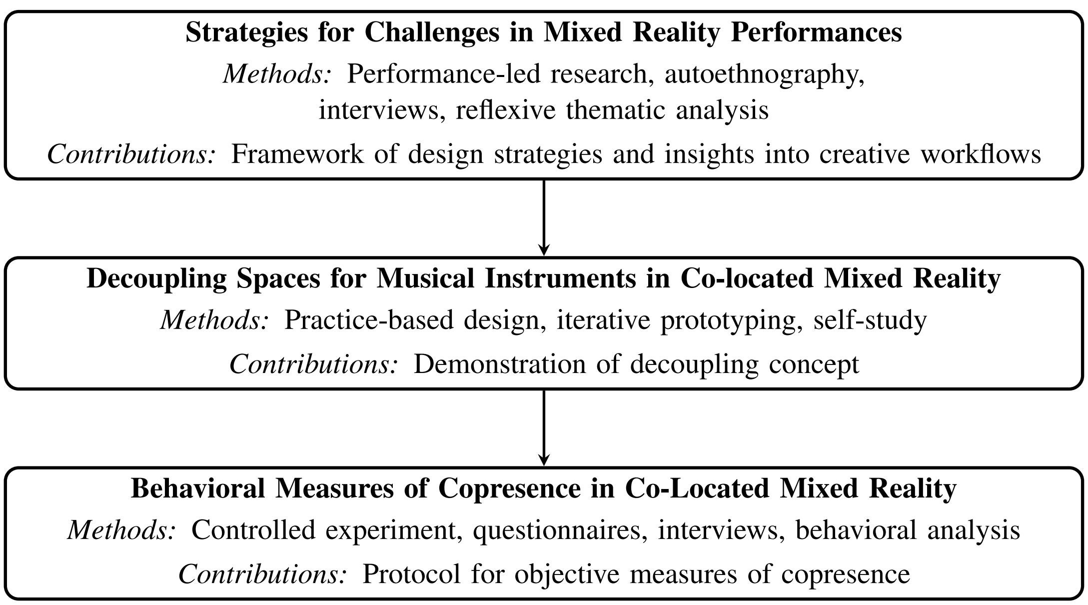

Fig. 1.1 Contributions and methods of the thesis.

Structure of the Thesis

The remainder of this document is organized as follows.

- **Chapter 2** presents the State of the Art. It defines the overarching concepts of Mixed Reality, Collaboration, and Presence and Copresence, and situates the work in relation to existing literature.
- **Chapter 3** details the methodology and results of a reflexive thematic analysis on the design strategies in MRP productions by leveraging personal experience, interviews, and case studies.
- **Chapter 4** introduces *gRAinyCloud*, a collaborative MR musical instrument designed to investigate the consequences of decoupling physical and virtual spaces. The system's design, implementation, and evaluation are described.
- **Chapter 5** turns to the question of measuring copresence in CMR. It examines existing measures and proposes a combined protocol that integrates self-report and behavioral data, illustrating its application through an experimental study.
- **Chapter 6** concludes the thesis by summarizing the contributions, reflecting on their implications for the design of MRPs and collaborative MR, and outlining perspectives for future research.

Chapter 2

State of the Art

This chapter will define the overarching concepts building the context in which the rest of this thesis is set. They are Mixed Reality, Collaboration, and finally Presence and Copresence. The following chapters will rely on the definitions and works presented here, but will also provide their own related works that are more specific to their respective frames.

2.1 Mixed Reality

Defining Mixed Reality (MR) is not as straightforward as can be initially thought. When considering the scientific literature as a whole, the definitions for it vary with more or less overlap. MR is even often referenced along with adjacent notions such as Virtual Reality (VR) or Augmented Reality (AR) . This group of denominations is sometimes used interchangeably, and at other times as a whole, avoiding the hassle of a thorough definition that would demand a clear stance. A good clarification of the ways MR has been used can be found in a study by Speicher *et al.* [112]. From interviews with 10 experts and a review of 68 papers, they were able to identify six notions designated as MR:

- MR as Milgram and Kishino's continuum [72], depicting a spectrum between AR and Augmented Virtuality (AV);
- MR as a synonym for AR, which can be used interchangeably;

- MR as collaboration, specifically between users immersed respectively in AR and VR;
- MR as a combination of AR and VR, where a system can feature AR parts and VR parts that do not necessarily have deep interactions;
- MR as an alignment of environments, which is the synchronization or alignment of virtual and physical worlds;
- MR as a "*stronger*" version of AR, principally through a better contextualization of the physical environment.

As a way to address this diversion, Speicher *et al.* provide a framework to help characterize these definitions and experiences based on the number of environments and users, the levels of immersion and virtuality, the implicitness of the interaction, and the existence of input and output.

In the rest of this thesis, the most commonly found definition will be followed, where MR describes the continuum ranging from AR to AV.

2.1.1 Reality-Virtuality Continuum

Milgram and Kishino initially introduced their "*virtuality continuum*" (see Fig. 2.1) as part of a taxonomy of MR visual displays [72], which became a grounding definition for the field. The extrema of this continuum are the Real and Virtual Environments. In the first, only real objects exist, which are defined as "*any objects that have an actual objective existence*". In the second, only virtual objects exist, in contrast defined as "*objects that exist in essence or effect, but not formally or actually*". MR finds its definition in the realm that lies between these environments, where both real and virtual objects can be perceived as coexisting. Within this case, AR and AV refer to situations with a clear predominance of respectively either the real or the virtual on the other.

A notable critique and improvement proposal of this definition was made by Skarbez *et al.* [104]. They consider the possibility of a "*Matrix-like*" display, a form of "*Ultimate Display*" that was able to not only replicate with perfect fidelity a visual stimulus, but also those of all senses. This would lead to the addition

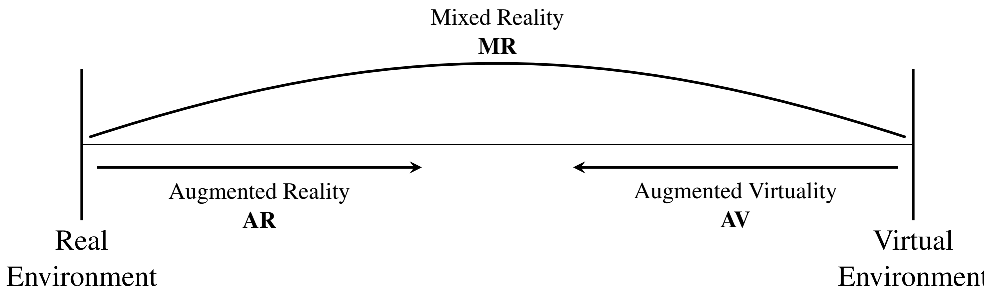

Fig. 2.1 Milgram and Kishino's Reality-Virtuality Continuum (adapted from [72] and [104]).

of a new extremum further than VR, which would become part of MR. This objection also serves to note that the inclusion of all senses, rather than the visual modality by itself, is crucial when considering MR or the Reality-Virtuality Continuum.

2.1.2 Immersion Modalities and Interfaces

A sense of immersion in MR experiences can be attained through a variety of sensory modalities and interface designs. While early definitions, such as Milgram and Kishino's, centered on visual displays, later research emphasized that MR is inherently multisensory [58, 104]. Visual immersion remains a cornerstone, relying on head-mounted displays (HMDs), projection-based systems, or handheld devices to blend physical and virtual elements. Yet this sensation also depends on other sensory modalities, such as hearing, haptic, or even taste [108, 100].

In terms of visual interfaces, stereoscopic rendering and head tracking form the basis for depth perception and spatial coherence. MR headsets such as the commercially available Microsoft HoloLens or Meta Quest Pro use optical or video see-through approaches, each with their own trade-offs in terms of field of view, occlusion handling, and latency [72, 90]. Projection mapping [35] and CAVE-like environments [79] can provide shared MR experiences without personal headsets, though they may impose constraints on interaction fidelity.

Auditory immersion is often achieved through spatial audio rendering, where sound sources are positioned and perceived in correspondence with the visual environment [100]. Spatialized audio can enhance realism as well as enable interaction, by offering cues for navigation, collaboration, or task completion.

Haptic feedback in MR can be implemented through vibrotactile actuators in handheld controllers [89], or through more specialized equipment with force-feedback devices [62], skin deformation [95], actuated surfaces [74], or ultrasonic tactile displays [71]. These technologies enable a tangible connection to virtual elements, reinforcing the perception of an environment sharing both real and virtual elements [62, 125].

Beyond single modalities, the integration of multiple channels is crucial. Studies have shown that congruent combinations of visual, auditory, and haptic feedback can increase presence, improve task performance, and enhance user satisfaction [108].

In the context of this thesis, immersion modalities are central because they influence not only the user's perceptual experience, but also how collaborators interact and coordinate within MR environments. The degree to which physical and virtual environments are perceptually correlated can determine whether collaborators share a coherent frame of reference.

2.2 Collaboration in Mixed Reality

Collaboration is referred to here as a situation where multiple users share an environment and have a varying ability to perceive or interact with each other. It is inherently a socio-technical phenomenon, as it is shaped both by interpersonal dynamics and by the tools, systems, and environments through which participants interact. A way to consider collaborative contexts is to look specifically at cooperative manipulation, defined by Margery *et al.* as the concurrent but cooperative interaction of several users on a same object, and how they classified it into three levels [68].

2.2.1 Levels of Collaboration

Margery *et al.* distinguished collaboration contexts based on the perception of the collaborators and the potential intersection of their interactions. Three main levels were identified, with the first only consisting of mutual perception between the users, along with a way to communicate. The second level allows manipulation of the virtual environment and can be further defined into two sublevels. One puts constraints on the possible changes, only allowing limited actions such as triggering animations and enabling or disabling objects. The other removes these constraints and enables a broader variety of interaction, most notably the ability to move objects to any position in the virtual environment. However, on this level, only one person is able to manipulate an object at a time, usually through a system of lock or mutex. An object can only be manipulated by several people at once on the third level, which can again be split in two cases. In the first case, though co-actors can manipulate the same object, they can only do so in independent ways. For instance, one can change the object's color, while the other changes its position, and a third triggers a scripted behavior, like emitting particles. In the second case, the actions are codependent, which means that the result becomes a combination of the inputs of every co-actor. This is the level at which cooperative manipulation happens, although concurrent manipulation will enable conflicts which need to be solved.

2.2.2 Conflict Resolution Strategies

Broll lists four options for conflict resolution strategies during real time cooperation in VR [26]. The first is to grant priority levels to the requests. This allows to only process the request with the highest priority while ignoring or delaying the requests with lower priorities. The requests can also be merged according to a weight obtained from their priority. The second option is a specific application of the priority approach where it is determined through earliness. The time and priority of the action can be measured either on departure or arrival of the request. The third option proposes to use constraints. It consists in splitting a high level parameter into several lower level parameters and distributing the new collection of parameters among the actors. In practice, if the control of an object is seen as a high level parameter, it can be split into the control of its color, position, rotation for instance, which can be controlled by different actors. The control of the position or rotation can themselves in turn be split into three degrees of freedom that can be controlled individually. This method is thus recommended for interactions with high initial levels of freedom, such as object manipulation, and amounts in a way in relegating the level of collaboration as defined previously from cooperative manipulation to its preceding sublevel. Even then, not all constraints can be respected, like when two co-actors try to control the speed and position of an object at the same time. The fourth and last option in certain cases is to merge the requests together in order to determine an overall outcome. For example, the speed of an object can be determined by adding together the speed that each co-actor is currently giving it. This calculation relies heavily on the kind of interaction, and can become complex, like when pushing on different points of an object, potentially rotating it.

It should be noted that prioritizing certain requests, such as in the first two methods, amounts to sequencing them and does not truly solve concurrent requests. On the other hand, the last two methods could be truly considered as solving concurrent requests, but they need to be adapted to each specific interaction due to their inherent complexity. In particular, the method of dividing the request into its degrees of freedom is only possible for when the parameter has several of them, which might not always be the same when two or more performers try to make a change at the same time.

These options and considerations are important to keep in mind in the design and evaluation of collaborative MR experiences since they have the potential to directly impact the experience of users, in particular their sense of copresence.

2.3 Presence and Copresence

Much like Mixed Reality, Presence and Copresence are key notions with a variety of conflicting definitions based in various research domains and contexts. The main definition of these concepts, which will be followed in this work, follows the one provided by Biocca *et al.* in their framework around social presence.

2.3.1 Definition

Presence is a staple in the evaluation of virtual experiences and has been extensively studied and defined in contexts beyond the strict scope of MR. A way to unify these varying definitions is to see presence as two dimensions: *spatial presence* and *social presence* [44]. On the one hand, spatial presence, also sometimes referred to as *telepresence* [17], corresponds to the feeling that the user may have of being located within a specific environment [66]. It is commonly described as the feeling of "*being there*" [17, 94, 96], and is individual, leaving no room for the consideration of another person. On the other hand, social presence refers to the feeling of sharing a space with other people [78]. Investigations around this concept mostly focus on how a person determines whether an entity is perceived as *real* or not. In the context of CMR however, no doubt can remain on whether a physically present person is an actual person, and the question becomes instead whether they are perceiving the same environment. Moreover, a review including perception and cognition studies of presence between 2013 and 2018 reported that most of them consisted of remote collaborative systems [8]. Hence, studying presence in a co-located context must rely on more appropriate tools than those designed exclusively for spatial or social presence.

Still, the question fits as a sub-dimension of social presence, further divided into *copresence*, *psychological involvement*, and *behavioral engagement* [19], which can be seen as three levels of social presence [18]. On a perceptual level, copresence is defined as the sensory awareness of an embodied other [77], and can even be extended to include a sense of mutual awareness [50]. On a subjective level, psychological involvement consists of how accessible the other seems, and in particular how aware one can be of the other's attentional engagement, emotional state, and comprehension [18]. And on an intersubjective level, behavioral engagement relates to someone's belief that their actions are interdependent with the ones of another. While copresence remains critical in this thesis, other dimensions of this definition of social presence appear in part relevant to it, namely attentional and behavioral engagement, as well as perceived comprehension.

2.3.2 Subjective Measures

Questionnaires attempting to evaluate presence are plenty and varied [103, 52]. The four most commonly found in the literature are Witmer and Singer's Presence Questionnaire (PQ) [127], the Slater-Usoh-Steed Presence Questionnaire (SUS) [119], the MEC Spatial Presence Questionnaire (MEC-SPQ) [121], and the IGroup Presence Questionnaire (IPQ) [97]. Yet none of these questionnaires were explicitly designed for MR, which leads to researchers adding custom single-item questions to address aspects specific to MR [42]. This necessary approach coupled with the already large variety of questionnaires [103] furthers the difficulty in comparing publication results [42]. A questionnaire built specifically for MR is the Mixed Reality Experience Questionnaire (MREQ) [86], which probes the perceived relationships among the user, and the virtual or real objects, environment, or agents. Like the previously described questionnaires scoped around presence, the MREQ does not cover the nuances that make up social presence. Namely, it inquires about the existence of the various elements, and whether they felt as belonging together, but falls short on notions of behavioral interdependence or perceived comprehension with other agents.

These considerations are part of Biocca *et al.*'s definition of social presence, which served for the design of a dedicated questionnaire, the Networked Minds Measure of Social Presence (NMMSP) [20]. It has notably been applied for MR with a subset of its questions to focus on specific dimensions [51]. The NMMSP was also adapted to the context of gaming, where communication is not the focus, in order to create the Social Presence in Gaming Questionnaire (SPGQ) with the notable addition of a mutual intention understanding question [40]. A similar approach can also be seen in order to fit presence questionnaires to the context of shared environments [30].

It should be noted that the overwhelming use of questionnaires is often described and criticized [103, 110, 44, 47, 52]. Post-experience questionnaires can easily fall victim to limitations such as recall errors [120, 44, 106], intrusion [103], forcing the participants into a meta-view [106, 105], the influence of prior VR immersion [46], or the complexity of the concept either relying on the participant's interpretation or outright imposing the researcher's conceptual framework on the participant [106, 105]. Participants also need a frame of reference to answer Likert scale-based questionnaires, which is an issue with between-subjects experiment designs [88]. While relying on questionnaires presented during the experience may alleviate some of these limitations [23, 98], completing it with an objective measure such as physiological or behavioral analysis is recommended [44, 106, 52].

2.3.3 Behavioral Measures

Much like Botvinick and Cohen's rubber hand illusion task for measuring immersion [22], many behavioral measures rely on an often stressful stimulus to elicit a reaction which will be measured. For instance, participants have been made to evolve close to a virtual pitfall, which allows recording instances of behaviors, like unprompted comments or where a reaction is incoherent, like stepping directly onto the pit [48]. In this scenario, more quantitative data can also be gathered through gait metrics such as stride length, width, and speed, along with the physiological measures of heart rate and galvanic skin response [82]. A similar scenario relies on flight phobia as a stimulus, which then allows measures through a subjective questionnaire, physiological data, and a behavior analysis [113]. This analysis is accomplished by a condition-blind observer, who reviews recordings and subjectively scores the participant's walk before and after the scenario based on their speed, behavior, reliability, and closeness to "*natural walking*" on a three point Likert scale. A unified scale was proposed for this class of scenarios, the Behavioral Presence Test in Threatening Virtual Environments (BPTT) [67]. It also relies on an external observer assessing the behavior of participants before and after they are exposed to the threat, like a fire or a pit. The scale consists of a list of anticipated behaviors, such as change in breathing pattern, or looking at the source of danger. While behavioral measures have also been used in MR by assessing risky behavior [67], all of these approaches are limited by their reliance on stressful stimuli, which constrains the possible tasks and whose applicability to copresence has not been verified.

There are also possible measures that do not rely on stressful situations [103], such as socially conditioned behaviors [101], pointing to an ambiguous object [107], comeback rate [115], or postural response [45]. However, none of these measures, which focus on individual actions or reactions, can be directly applied to a context of copresence.

Overall, when considering either questionnaires or behavioral measures, there is no established protocol that can be directly applied in the context of CMR, where several participants interact physically or virtually, as some questions may not cover its specificities, and protocols may not be designed with several participants in mind. Moreover, both options come with their own intrinsic limitations, like needing a frame of reference for interpreting questions, or being very specific to certain use cases. This reveals the need for better ways to measure presence and copresence in the context of CMR that do not rely solely on either questionnaires or behavioral measures, and remain aware of their limitations.

Chapter 3

Strategies for Challenges in Mixed Reality Performances

This chapter is adapted from [117]:

P. Uro, F. Berthaut, T. Pietrzak, M. Wanderley, and L. Grisoni. *Strategies for the Reconciliation of Artistic Intent and Technical Constraints in Mixed Reality Performances*. In ACM Designing Interactive Systems Conference, Madeira, Portugal, 2025.

3.1 Reconciling Artistic Intent and Technical Constraints

As an ever-increasing number of artistic productions employ immersive displays in Mixed Reality (MR) or Virtual Reality (VR), a form that can be referred to as Mixed Reality Performance (MRP), they often run into the issue of compromising between what the technology enables and what the production team envisions. Each project faces its own set of constraints, and while some productions succeed in crafting satisfying compromises, these solutions rarely translate directly to other contexts, even when the technological and artistic contexts appear similar.

For instance, a performance may choose to have a large audience sit down for a narrative experience to avoid tracking or collision issues [41], and reduce the amount of data needed to represent every participant as avatars [57]. While effective in that case, this approach would not fit in other contexts requiring more freedom of movement for the participants. Similarly, a low-quality black and white passthrough camera can, in certain cases, become an asset rather than a hindrance if its aesthetic aligns with the tone of the piece and does not interrupt the audience's immersion [59]. Yet again, not just any production can justify such a choice, and some may be forced to find an alternative camera.

These examples highlight a central challenge: the same issue can elicit radically different design responses depending on the constraints and intentions of the production. Designers must find ways to address occurring limitations while respecting both the technical and artistic constraints of the production.

This chapter examines how designers navigate these tensions by identifying overarching strategies used to re-conciliate immersive technology constraints and artistic intents. Through a reflexive thematic analysis, which draws on a performance-led case study, two interviews, and independent case studies, five strategies forming a design space are identified. This space is organized according to how much of the limitation is revealed to the audience. Its is shown how this design space allows designers to consider a range of alternative strategies when facing conflicts between limitations and ambitions during the creation of MRPs with varying constraints.

3.2 Mixed Reality Performances

The term Mixed Reality Performance (MRP) is defined by Benford and Giannachi as a theatrical genre that mixes "*real and virtual worlds*", and combines "*live performance of actors or participants with interactive digital media*" [12]. Numerous examples of MRPs stem from the long-term collaboration of the University of Nottingham's Mixed Reality Lab[1] and the artist collective Blast Theory in the form of *Desert Rain*, *Can You See Me Now?*, *Day of the Figurines*, or *Ulrike and Eamon Compliant*, dating from 1997 to 2009, though some of these experiences have also been qualified as "*extended theatrical performance*" [11]. These productions are extremely varied in their form, ranging from a 45-minute performance with projections on a water screen in *Desert Rain*, to a mobile phone test-messaging adventure game in *Day of the Figurines*. It is important to note that Benford and Giannachi's definition of MRP does not restrict

[1] https://www.nottingham.ac.uk/research/groups/mixedrealitylab/

its scope to headset-centric MR and remains flexible as to the involvement or not of actors as part of the performance.

A different group of researchers, New York University's Future Reality Lab[2], was also able to develop a number of narrative experiences, in this case more focused around immersive technologies. Among them, *Holojam in Wonderland* features 2 actors and 4 spectators sharing a virtual environment as part of a narrative experience categorized as Immersive Mixed Reality Theater [49]. Deciding to focus on scale, they went on to produce *CAVE* with the *CAVRN* system to immerse up to 30 participants [64, 57]. Their later production, *Mary and the Monster*, intended for participants to share a narrative experience while immersed either in AR or VR [56]. Again here, even within the same group of creators focusing on immersion through headsets, the variety in technological and artistic contexts is great. The three productions used different headsets, and while the first involved actors and a small audience freely roaming in a shared space, the others featured no live actors, and large seated audiences.

A similar variety is to be expected with more independent productions. The diversity of their creative stakes can be quickly seen through how they were designated. For example, *Debussy3.0* is described as either an augmented performance, ballet, or show and presents an interaction between stereoscopic 3D and the performance of two dancers [35]. With a similar screen-based immersion, Marner *et al.* presented *Half Real* as "*a live action interactive theater show employing spatial augmented reality*" [69]. The show immerses the actors into a projected virtual environment and allows its audience to select directions of the narrative via a voting system.

While some productions such as the previous two define themselves around the artistic discipline they are built around, some can be described with more technological details. This is the case of *Wake*, which is presented as an "*asymmetric, co-located, co-present, mixed reality social experience*". It lets a headset-wearing participant interact with the volumetric projection of a performer inside a virtual environment [55].

The productions described here range from a single participant to an audience occupying an entire theater room, with or without live performers, or using different means of immersion. Yet, they all arguably

[2]https://frl.nyu.edu/

fall under the concept of MRP, and as such have the potential to draw lessons from each other.

A recurring framework in the conception and analysis of MRP is Benford *et al.*'s trajectories, which were derived from the projects created at the Mixed Reality Lab [11]. This analytical framework describes the experience of the participants through their journey across space, time, roles, and interfaces. One of its uses is the identification of *transitions*, key points in the experience where a participant changes their state in one of their trajectories. These moments are important since they are where the risk of breaking the immersion of the participant is highest. Transitions include the beginning and end of the experience, changes in roles and interfaces, shifts from physical to virtual worlds, or the presence of *seams*.

Seams are what Chalmers and Galani call the noticeable boundaries and differences in a system [32], which in Benford *et al.*'s framework could cause a problematic transition. This does not however mean that any seam must be eliminated, since Chalmers and Galani argue for a seamful design, a concept that aims to recognize the existence of these seams and to exploit them.

In one of the works that defined MRPs, *Can You See Me Now?*, Benford *et al.* were able to draw from this idea [10]. In the piece, GPS and Wi-Fi signals proved to be unreliable and brought a level of uncertainty to the experience. They were able to identify the various states of connectivity and proposed five strategies for coping with them: Remove, Hide, Manage, Reveal, and Exploit. While these strategies may be informative for the design of other experiences, they have not yet been applied to contexts broader than the connectivity issues encountered in *Can You See Me Now?*.

In a case study of two MRPs, Rostami and McMillan [91] document challenges and strategies artists and performers encountered and adopted through video analysis. They regroup them under four categories: *Stage Managing VR Performances*, *Choreographing for Cables, Consistency, and Charging*, *Improvising Interventions*, and *Priming Participants*. While the detail provided by the video analysis of the incidents during the performances is valuable, such a method may not capture issues and answers that occurred earlier in the design process.

Weijdom built on top of Benford and Giannachi's definition of MRP to define a Performative Mixed Reality Experience (PMRE) [126], which draws more focus on the body becoming part of the aesthetic of the piece. This serves as a basis to study the design processes of two productions in order to propose a performative prototyping method, which uses a combination of bodystorming, Wizard of Oz, and puppeteering techniques along with the development of an interdisciplinary vocabulary as a way to achieve a common ground for designers of PMREs.

MRPs can also be considered through the adjacent scope of immersive musical performances. In this regard, Zappi *et al.* proposed dimensions to analyze the scenography of Immersive Virtual Musical Instruments (IVMI) [129]. Though this framework focuses on IVMIs and may not fit all MRPs, it introduces dimensions that aim to describe the family of IVMIs, which is also subject to vast diversity in artistic and technological contexts.

The scope of MRP can also be broadened to non-performance MR applications and the challenges faced by their designers. Eight key challenges were identified by Ashtari *et al.* [2] for authoring AR/VR applications along with some approaches taken by creators. Similarly, Krauß *et al.* [61] revealed three key challenges for AR/VR designers, and potential solutions for them. They also identify roles taken by the designers and give guidelines for the design of authoring tools. These findings relate to a scope encompassing that of MRPs, and as such may be applied by MRP designers, yet they might miss challenges caused by the particular context of a performance, or strategies afforded by it.

The work presented in this chapter draws from these frameworks and attempts to adapt their models to the specific context of MRPs.

3.3 The Case of *Animate*

The motivation for the study presented in this chapter, and what was employed as the foundation for the following analysis process, is the work I accomplished as part of the production team of *Animate* [93]. I took on the role of Lead VR Developer from the beginning of my thesis in September 2021 through the first installment of the show in early September 2022 where I supervised the showings.

3.3.1 Description

Animate can be associated to a MRP at the crossroads of performance, radio play and installation, focused on a near-future Canada radically transformed by climate change. The show premiered in August 2022 at the Kunstfest in Weimar, Germany and was experienced by over 400 people in 10 days.

Artistic Intent

The plot of the story *Animate* focuses on two characters, Daniel and Laurie, as they escape a near-future marked by climate disasters. Both are emotionally burdened by their own unresolved pasts, and their journey takes them across the wild landscape of Newfoundland, Canada. While news of global climate catastrophes is broadcast on their car radio, they drive towards the Tablelands, a geologically unique region in Gros Morne National Park where the Earth's mantle is exposed, resulting in a Mars-like landscape.

Over time, the climate-change transformed landscape exerts a force on them as if it was a conscious, living, and breathing entity. It produces eerie sounds and affects the characters' personal relationship. But it is only in the dramatic conclusion, where the earth comes alive in an apocalyptic scene of rocks rising from the ground and attacking the characters, that the intricate interconnectedness of humans and a natural world under the throes of radical environmental change is symbolically and viscerally manifested.

The dramaturgical strategy that evolved in the early phases of the project involved collaborating with the author Kate Story. The original short story was adapted into four dramatic scenes, written in the form of a theater text. These were then recorded by actors to produce a radio play format, which served as a sonic accompaniment for the piece. This layered audio was later integrated with MR visual components and live performance, crafting a hybrid theatrical installation that situates the audience within both the narrative and its immersive environment.

Technical Constraints

The set for the premiere of the show was the KET-Halle, a massive decommissioned agriculture machinery factory whose main area spans over 144 by 120 meters with a ceiling of 10 meters high at its lowest, though the actual area used both as a stage and a wandering space for the audience still spanned 16 by 32 meters. Inside this space, the Berlin-based MONOM studio had set up its 4DSound system: an array of 48 omnidirectional speakers that allow the physical placement of sounds within a 16 x 16 x 5 meters volume together with subwoofers placed on the periphery of the space. The audience wore Meta Quest 2 headsets, chosen because of their inside-out tracking and accessible pass-through features, and for their affordable cost allowing easier up-scaling to 12 people.

The audience went through three immersion phases. First, they were not virtually immersed at all, sitting on wooden logs for a dramatical reading by the two performers of the beginning of the novel on which the plot is based. Then, they entered a second space and were equipped with the headsets. There, the two actors pulled a rope to guide the audience, who was holding onto it, and who could only see a completely virtual scene. The virtual environment then opened up into AR, revealing the physical stage, the rocks, and the performers, who were not wearing headsets themselves. As the scene proceeds, the rocks started moving, rushed after the audience, formed choreographies around the room, and gained in intensity before falling together to the ground.

Although *Animate* was not initially identified or designed as a MRP, its artistic and technological contexts place it firmly within that scope, as an immersive experience placed firmly at the intersection of virtual and physical environments and allowing the interaction of audience members and performers.

3.3.2 Addressing the technical challenges

Implementing the show meant considering each of the technical challenges in light of the artistic intentions with the practical constraints caused by the very specific context of the show. Not every issue could be completely solved, and some had to be mitigated or avoided with the final objective to limit as much as possible the negative impact it could have on the audience's experience while staying true to the underlying artistic intent. It would often come down to what was deemed acceptable for the experience of the audience and would not threaten their immersion.

Occlusion

The issue of occlusion only occurred for the virtual rocks in the last part of the show where they began floating and moving around, since it was only then that the audience left the completely virtual environment. In this context, there were two categories: the stage and the setting were static objects, and the audience and actors moving around were dynamic objects. Preserving immersion would require implementing occlusion for both of those categories.

The use of models was possible for any of the static objects in the room and its walls as long as the spatial alignment worked correctly, *i.e.* if the headset was accurately aware of where it was in the real space. This had the added benefit of enabling collisions between the rocks and the modelled objects, which served the intended merging of both environments.

The question of which elements of the room should be modelled and thus occlude the objects could however be raised. In fact, objects could be part of the natural decor, or be completely practical, such as fences or speakers, and should be considered carefully. Each object could thus provide occlusion based on factors like how conveniently they could be modelled and positioned, or how much their eventual occlusion would impact the overall experience. Occlusion could indeed take away from the show, hence it should not be applied blindly to everything. For instance, providing occlusion to fences or speakers would accentuate them by bringing them into the performance, when they should have been ignored.

Fig. 3.1 First-person view from a participant in *Animate* attempting to protect themselves with their hands, which are displayed in front of the rocks.

In the case of dynamic objects, the hand tracking provided by the headsets was used to dynamically model a person's own hands (cf. Fig. 3.1). Attempting to model the other dynamic objects, *i.e.* the rest of the audience and the actors, with non-immersion-threatening quality would have required at least partial body outside-in tracking, which would have quickly become too complex when scaled to a group of people. Depth estimations would have not only risked low accuracy with the higher distances encountered, but also would have indiscriminately occluded any detected shape, without consideration for its value for the performance.

Spatial alignment

The large scale at which the tracking needed to operate excluded outside-in tracking, leaving its inside-out alternative which is already handled by the Meta Quest 2. However, the specific conditions of the space brought issues of instability to the system, namely the high ceiling, varying natural light, and the feature-rich but spatially periodic environment, in particular the highly similar pillars that the headsets could mistake for each other. These unfavorable conditions could be alleviated to an extent. For instance, varying light could be tackled with more consistent electric light. Environment-related tracking issues could be helped by removing parts of the set that appeared problematic, like mulch that had been initially spread on the ground. It was however not enough to guarantee perfect tracking at this scale, as it was found to be unreliably lost and reset in the middle of a performance, and even walking from one side of the stage to the other could introduce a drift from the initial alignment by over 10 centimeters.

Since the tracking system itself could not be trusted, the solution was to refrain from using a highly repeatable spatial alignment protocols, which initially relied on repeatedly positioning three orthogonal planes using the ring-like part of the Quest 2 controller and approximating a position and rotation. Instead, the fact that issues could happen at any point during the performance was acknowledged, and the focus was put on either shrinking their impact on the overall experience or making them fixable with little to no interruption of the person's experience. The positional offset caused by the drift was not experience-breaking in either scene, so only the radical change in alignment caused by an autonomous reset was a major issue.

The alignment protocol was downgraded to only a single point, the headset itself, resetting both orientation and position of the virtual environment's origin. This allowed technicians to make immediate decisions on the balance between speed and quality of alignment, by moving the person to the optimal spot, simply orienting them in place, or timing the reset to when the spectator looks in the right direction.

The choice between these options was left to the operators on a case-by-case basis, depending on how responsive the person seemed to be, and whether it was more important to have the proper position at the cost of moving the person or if the orientation sufficed in that part of the scene.

Shared experience of Mixed Reality

One of the goals of the last scene was to virtually drown the audience in a massive amount of flying rocks while giving them the option to protect themselves with their hands. Synchronization was an essential aspect of the audience experience here, as the rocks' visual behavior should match what the external sound system plays for all users, while maintaining an illusion of copresence, with everyone seeing the same objects at the same places.

Environment Behavior Spatial synchronization could rely on the loosely similarly calibrated headsets. Coordinating the time for all headsets could be easily done by sending them a start message, which triggered a static timeline. When interaction came into play, however, the state of the objects needed to be synchronized, which became tricky with how complex the environment was, as hundreds of independent rocks moved, collided, and were interacted with in a massive physical simulation.

Because of limited network resources, the chosen solution was simply not to synchronize the state of the rocks and instead rely on the illusion and audience's assumption that the environment was actually synchronized.

Not synchronizing interactable objects was in fact possible by designing the behavior of the rocks to be flexible around a fixed timeline. The rocks were pulled towards scripted targets which were the same for everyone using time and space synchronization, yet they could still be pushed off course by the audience.

In order for the rocks to keep an acceptable similarity between headsets, the audience could not be

Fig. 3.2 A participant appearing to be interacting with rocks as seen by another member of the audience.

allowed to unsettle them too much. To achieve this, instead of using realistic simulated collisions on their hands, which would allow slapping the rocks away with high bursts of energy, low-intensity force fields centered around each hand continuously pushed away the rocks that were in a range of one meter. The energy that can be given to a rock was thus more spread out in time, and less disruptive of its intended movements, which also tied in with the artistic intention of allowing the audience to connect with the rocks without letting them take the objects' agency away. Interacting from a distance rather than in contact also made it less obvious to other members of the audience that what they each saw was slightly different (cf. Fig 3.2).

Even though every audience member saw a marginally different environment based on how the rocks were pushed for each of them, the behavior of the environment on a macroscopic scale was the same. It would have been possible to also synchronize everyone's hands in order to still trigger the interaction with rocks floating close by in every instance of the environment. However, after weighing the instability of the spatial calibration against the limited impact of that synchronization, this was not implemented, since test audiences appeared easily overwhelmed by the chaotic swarms of rocks and paid small attention to the other people in the room.

Audiovisual Coherence In *Animate*, rocks are audiovisual objects, as they have audio and visual representations, and both needed be taken into account in order to maintain consistency and therefore immersion. The most important aspect of this coherence was that audio and visuals matched both in position and nature. Because the behavior of the rocks was determined on a macroscopic scale, the sounds they made could afford to be entirely scripted, which drastically simplified the sound design work as it could focus on a single optimal case instead of a large spectrum of potential states.

The early approach of triggering sounds based on collisions of rocks made the difference between the static audio and the dynamic visuals too obvious since a collision could happen in one representation of the rocks and not the other, thus breaking the feeling of immersion. Instead, the nature of the sounds matches their scripted macroscopic behavior, like gathering into a tight cluster, forming a tornado in the center of the stage, or dropping all at once. The position of these sounds can also acceptably match the visual state of the environment by making them occupy a larger area, which makes their exact location harder to identify, and thus their slight incoherence harder to notice.

3.4 Iterative Process

The challenges encountered in the production of *Animate* are highly specific to its context, both artistic and technical. This does not mean that nothing can be learned from it for MRPs in different contexts. This is also the case for other productions, and the goal of the process described here is to identify overarching elements in the ways MRP designers approach critical challenges they encounter in their productions. In this context, *challenges* are considered as situations where the available immersive technology is not capable of accomplishing a desired artistic outcome, resulting in a compromise between what is wanted and what is possible. Designers thus need to come up with an *answer*, which is how they decide to approach and work around the compromise imposed by the *challenge*. Since the *challenge* relies on the technology being used, and the desired result, so does the *answer* given.

The aim is to model the *answers* brought by various designers into *strategies*, which describe potential approaches regardless of their artistic or technological context.

This will be undertaken through a three-phase reflexive thematic analysis attempting to associate the approaches in various MRP productions as described by their designers. The phases consist of a performance-led case study, two interviews, and independent case studies (see Figure 3.3). During each step, an analysis is performed following Byrne's example [28] of Braun and Clarke's reflexive thematic analysis [24, 25]. As such, underlying theoretical assumptions within this study should be addressed. Since this analysis is of qualitative essence, it follows constructionist epistemological considerations, implying that both recurrence and meaningfulness in the data are taken into account in the generation of themes. The approach adopted leans on both critical and experiential orientations, since the identification of challenges relies on a description of the experience of the subjects, while the identification of strategies is related to the finding of an underlying structure within the testimonies. Coding was accomplished both in a semantic and latent manner, as the data in the entire corpus already comes in analytical form as communicated by experts, which can be interpreted explicitly. Still, as the goal is to find an overarching model in the approaches of the subjects, latent coding also found value where deeper meaning could be constructed by the researcher's interpretation. The first phase adopts a predominantly inductive approach, while the second and third reflect previous iterations and thus present a stronger level of deductive process.

Each of the three steps completes the six phases of the process of reflexive thematic analysis, which

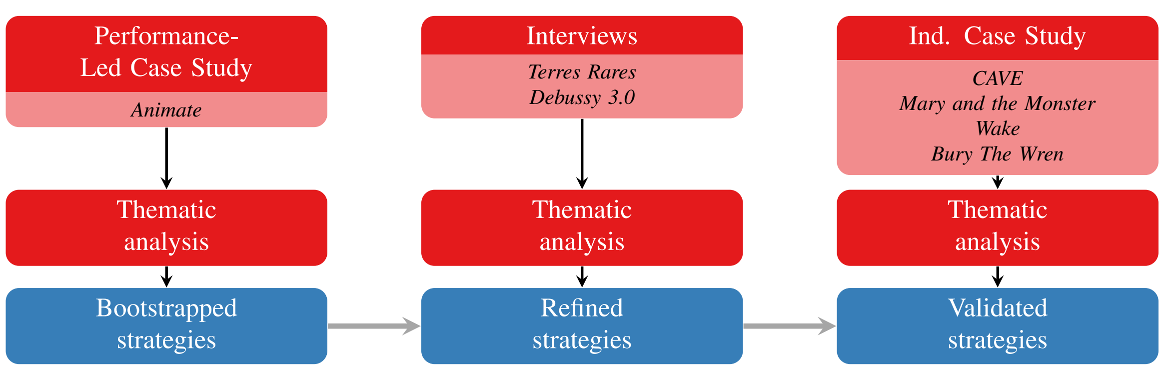

Fig. 3.3 Method used for the iterative process.

are data familiarization, initial code generation, theme generation, potential themes review, theme naming and defining, and report production.

The first phase of the analysis focuses on my experience as the Lead VR Developer on the immersive theater piece *Animate* through the analysis of a presentation that was given shortly after the premiere of the show. This helps gain an initial understanding of potential answers, which is used in the next phase, and reveals bootstrap strategies.

The second phase consists in the analysis of two interviews with designers of MRPs, in which they were asked about the various challenges they encountered and how they answered them. The second interviewee is a co-supervisor of this thesis and did not participate in the preparation or analysis of the second phase. The insight gained during the previous phase allows finding challenges that were not explicitly mentioned in the documentation of the pieces. The analysis of the interviews also draws from the initial case study when devising its themes, resulting in refined strategies.

Finally, several independent case studies are accomplished by completing an analysis of published material extensively detailing the design process of various recent MRPs. This serves as a validation step to the previous results by challenging them with independent perspectives.

The perspectives that were selected were chosen with the goal to represent the diversity of what can be included in a MRP, as can be seen through their representation in Figure 3.4. This includes various immersive technologies, various sizes of audience, and either live or pre-recorded performances. The focus was also put on picking projects in a relatively close chronological context in order to keep a similar technological landscape throughout the various productions.

The involvement of a co-supervisor and myself in two of the studied projects grants a familiar point of view on the decisions that were taken throughout their respective design. That same involvement is also what motivated this study. Moreover, the auto-ethnographic aspect of the initial case study, and the potential bias of the interviews being led by the same person were addressed through the cross-validation step of the independent case studies, which bring a larger variety of points of view on the implementation of MRPs.

Throughout each step of this process, challenges that could be identified are listed in Table 3.1 in order

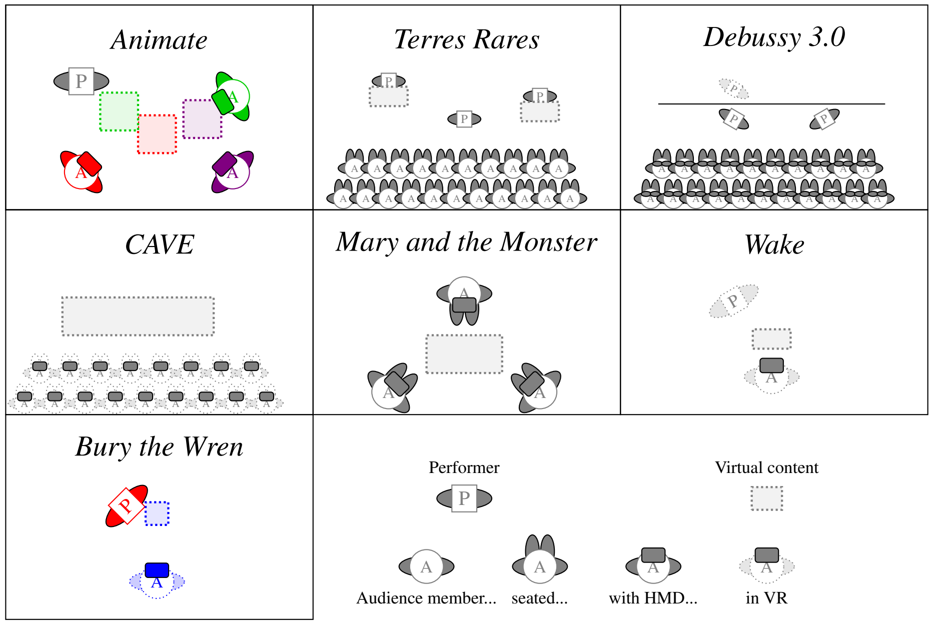

Fig. 3.4 Top-view representations of the mixed-reality setups for parts of the studied performances : P correspond to performers, A to audience members. Solid lines correspond to physical persons/objects while dotted lines represent virtual avatars/objects. Audience members only see gray elements and elements of the same color as themselves. Audience members in dotted lines with headsets do not see the physical space, while those in solid lines see both virtual and physical spaces.

to keep track of the variety of challenges found within the corpus. For each project, one such challenge will be detailed along with the approach taken by the authors in response.

3.4.1 Performance-Led Case Study

The first step of the iteration is a case study of *Animate* through my own experience on its production as part of a team of artists and developers with a research-creation approach.

Methodology

The data set for this iteration of the analysis is a fifteen-minute speech that was given by myself in October 2022, a month after the show premiered. The presentation was given at a workshop focused on audio and MR as an early recollection and nascent analysis of my experience of the production process. It was structured around the challenges that were encountered, and how the resulting *imperfections* were

Table 3.1 Specific challenges that were identified in each project.

Projects	Identified Challenges
Animate	Low reproduction fidelity
	No occlusion from physical elements
	Occasional loss of alignment
	Complexity of large scale shared interaction
	Complexity of shared audio
	Un-immersed performers
	Encumbrance of emotion sensing equipment
	Discreet haptic feedback
Terres Rares	Partially immersed performers
	Restricted effective area for display/capture
	Respective occlusion of performers
	Capture-display latency
	Inaccurate intersection with a prop
	Inaccurate mechanized control of a prop
Debussy3.0	Disturbance of capture by quick movements
	Capture-display latency
	Impact of point of view
	Occultation of physical and virtual elements
	Single body-tracking suit
	Restricted effective area for capture
CAVE	Complexity of large scale tracking
	Complexity of large scale shared interaction
Mary and the Monster	High bandwidth for large scale hand tracking
	Occasional loss of alignment
	Performance-costly avatars
	Restricted field of view
	Multiple viewing paradigms
	Restricted exploration while sitting
	Fatigue-inducing interaction
	Complex interaction
Wake	Restricted field of view
	Restricted movements of tethered headset
	Capture-display latency
	Imperfect positioning of dancer projection
	Interference between trackers and infrared depth camera
Bury The Wren	Low reproduction fidelity
	Uncanny avatars
	Uncanny animation
	Weightless virtual objects
	Restricted effective area for capture
	Disruptive VR to AR transition
	Unstable VR to AR transition
	Imperfect Photogrammetry
	Abrupt transition between virtual objects
	Distraction of the object tracker before immersion
	Limited rendering budget
	Inconsistent shape of physical handle and virtual objects

addressed to achieve a working show. For this reason, analyzing it under the larger scope of strategies can reveal underlying patterns that were not obvious at the time of the presentation. This means that in this analysis, while the approach is generally inductive, attention is put towards identifying strategies, which brings deductive components.

Animate

Animate has been described thoroughly in Section 3.3.1 as a MRP production which notably features two non-immersed performers and up to twelve headset-wearing participants.

A key challenge in *Animate* was the sharing of several hundred flying rocks among a dozen headsets. Each rock followed a physical model as well as a predefined choreography, while the audience could push them. Such a large amount of independent objects proved to be too much to be shared over the network, and an approach was taken to let each headset have their own instance of the scene. Each participant thus saw their own rocks, which meant they ran into the risk of noticing their neighbor interacting with rocks they could not see themselves. The answer to this limitation was to highly reduce the amount of impact a participant could have on a rock. The interaction went from being collision based, to relying on a force field around each hand that could only slightly push the rocks away. This ensured that the macroscopic behavior of the rocks would remain close to the scripted choreography, and that the audience would thus be unlikely to notice the difference. This choice was also made possible by a parallel decision to let go of live spatial audio generated from the collisions of the rocks as it was deemed too unreliable. Instead, a pre-mixed audio was created in order to guarantee the effect desired by the designers.

Resulting Themes

The analysis accounts for two themes as seen in Fig. 3.5 : *Strategic Management of Limitations* and *Complementary Considerations*.

Strategic Management of Limitations The first theme reveals three potential approaches in its sub-themes, starting with the use the audience's awareness. The first entails that designers keep in mind what

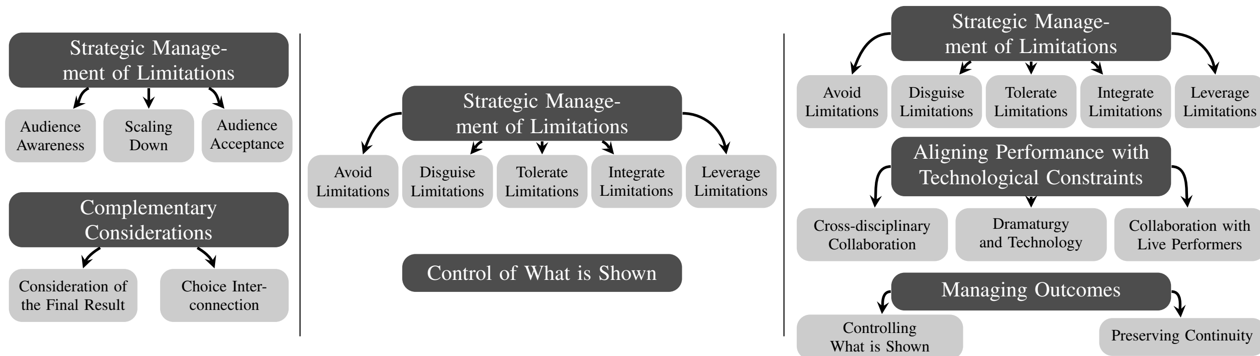

Fig. 3.5 Theme map after each iteration: case study (left), interviews (center), and focused literature review (right).

their audience will notice or not, in order to make the limitations of the piece less discernible and make use of channels they do not have access to for orchestration. The challenge of sharing a large amount of objects is answered this way by altering the interaction of the rocks in order to make their inconsistent position less perceivable.

A second approach is to scale down the ambitions of the project. This can be done by considering the trade-off between complexity in the implementation or running effort and the effect granted on the audience. A compromise can then be found by reducing or removing certain features, in particular the complexity of the interaction or of the virtual environment. An example of this is the foregoing of haptic feedback in the production, as the acquired technology added too much complexity in relation to the effect it provided.

The third option is to encourage the audience to accept the limitation as part of the story. This way, issues deemed unsolvable can be addressed by only altering them enough to make them part of the aesthetics of the piece. For instance, the passthrough cameras used for the AR sequence were in black and white, with poor resolution. The decided approach was to shift the entire aesthetic of the virtual environment towards how the camera displayed the physical world, supporting the visual merge of both environments.

Complementary Considerations The second theme involves additional considerations that can inform the choice of an approach. First, designers should consider the finality of the piece during the implementation. This means that technological decisions should be driven by dramaturgical decisions based on the intended narrative experience. Moreover, breaking points should be preemptively assessed and safeguards should be implemented based on their likeliness to happen. For example, a tendency for the headsets to lose their alignment with the physical world caused by the light and the repetitive nature of the setting was identified. In response, the alignment protocol was simplified, sacrificing accuracy for speed and ease of use in order to let orchestrators realign headsets during the performance with limited interruption for the participant.

Secondly, decisions are rarely independent, whether they happen at the choice of a technology, or during implementation. As such, deciding to limit an aspect of a performance may also solve or cause other issues in the project, much like the decision to instantiate each headset, which impacted the audio design.

3.4.2 Interviews

For the second phase of the iterative process, two interviews were conducted with researchers that worked on an augmented ballet, and a cyber opera. In both cases, the interviewees were assigned to the implementation of the immersive technology within the performance and remained in close collaboration with the directors of the pieces. As such they can provide insight on the technological details, the decisions that were made and the motivations behind them, as well as their potential outcomes.

Methodology

The interviews were tailor-made for each interviewee and thus had different questions, but followed a common pattern. Using available literature on each production, implementation details that revealed explicit or implicit challenges were identified. The first questions were about the challenges that were *explicitly* presented in the written documentation and requested more details on how answers were devised and compromises were found. The next questions focused on *implicit* challenges, which are potential issues that were quickly mentioned or could be hypothesized from the written documentation, with little to no details on how they were addressed. These questions served to open up the scope of the interview for the final part, where the interviewee were asked to share any other challenges that may have come up during the production, and what the answer of their team was. While a plan was prepared for both interviews, the order of the questions was not strictly upheld, and some questions were asked when they were more appropriate with the topic of the conversation. Both interviews were conducted in French, any following citation was thus translated.

The following analysis did not begin using the same codebook as the first iteration in order to allow different expressions to come out of the independent discourse of the interviewees. However, during the analytical phases starting from theme generation, codes from the previous iteration were included and potentially merged with the new codes. This allows the iterations to build on top of each other while keeping the resulting themes and their structure flexible when facing new data.

Terres Rares

Terres Rares is a cyber-opera featuring four on-stage performers and four musicians [3], which premiered in 2022 [1]. One of its authors, a researcher who designed the immersive part of the performance and is also a co-supervisor of this thesis, was interviewed for 40 minutes. The interviewee for this production will be further referred to as P1. The interview was structured around four challenges that were explicitly stated, and one implicit.

The first two acts of *Terres Rares* consist of traditional operatic settings with acrobatic sequences and limited use of technology on stage. What pushes the performance into the realm of MRPs is its third act, which relies on the use of a Spatial Augmented Reality (SAR) display across four sequences. This system consists in the combination of a projector and depth camera to reveal slices of physically located volumetric content [31, 53, 16]. Using either their own bodies or props as a screen, actors can intersect the virtual shapes in order to reveal the resulting slice on the corresponding physical surface. They use this system to finely control playback of video recordings by positioning props or themselves inside the virtual shapes as they perform.

An example of the limitations that the production came across is the latency between the detection of a

[3] https://terrev.univ-lille.fr/terres-rares/

surface and the projection of the revealed content on the surface. This caused an issue with fast movements not being followed correctly, but coincidentally, the performers were meant to act as memories or ghosts, making them move slowly and avoiding the issue.

Debussy3.0

Debussy3.0 is described in a dedicated article as an augmented ballet and consists in the performance of three people on and behind the stage for over 400 spectators [35]. The first author of that article was interviewed for a total of an hour and will be further referred to as P2. The interview was structured around the challenges that were identified within the paper, three of which were explicit, and two were implicit. P2 was also able to mention a sixth challenge, which was the limited range in which the body-tracking suits could be used, and which was not found from the paper alone.

Across three acts, two dancers perform a choreography in front of stereoscopic 3D screens, one of which is being tracked using a XSens MVN motion capture suit. This tracking is used to control virtual elements that are being displayed, such as an avatar mirroring the movements of the dancer, or a trail following their hand depending on the act. An "*Augmented Reality Engineer*" directly controls the timing and visual filters applied to various virtual elements in order to match the performance of the dancers.

One of the challenges met by the production involved the use of stereoscopic 3D with a static virtual camera, which meant that the point of view of a spectator impacted the position of the virtual objects they saw. For instance, a person close to the stage could see an object in its center, while someone further away could see the object above the audience. This prevented the designers from establishing a spatial relationship between the dancers and the virtual components. In response, the scenography was adapted so that most virtual elements appeared as behind the screen, where this impact was limited, and some on-stage virtual elements were removed, such as sculptures.

Resulting Themes

The second iteration resulted in two themes (see Fig. 3.5): *Strategic Management of Limitations* and *Controlling What is Shown*.

Strategic Management of Limitations The first theme inherits from the previous step, expanding it with new considerations found in the interviews, which leads to its five sub-themes, corresponding to five potential strategies.

A common strategy is the *avoidance of the issue*, for instance, P2 expressed about the magnetic sensitivity of the sensors on the tracking suit "*So on a stage with an all-metal structure, in the end, it was fine, we had a few places where we could not go, but we avoided them, and that's it*". Another instance of this strategy is in *Terres Rares*, where a large inflatable ball was supposed to fall from above and was meant to be grabbed and manipulated by the performers as a screen. This was deemed dangerous in the dark, and too impractical with an inaccurate timing, since only one of the performers could see the projection on the prop, and was subsequently abandoned.

The second strategy has the same aim of preventing the audience from perceiving a limitation, and relies on *disguising the limitation*. This can be achieved by relying on the limitation not being disturbing for the experience of the audience, like P2 who commented "*But that's not very bothersome because it's really a small thing*" about the inconsistent positioning of virtual objects. Similarly, P1 shared "*Anyway, the background of the stage is black, precisely to absorb any potential mismatches between the captured area and the projection*" about how the issue of lag was answered.

In some cases, there is no choice but to *tolerate the limitation*. This case may happen due to cost, logistic or technology maturity causes. For example, P2 explained "*We did not have enough time at all, since we had only 3 months and received the retro-projected screens the day before the premiere.*" when asked about the slight lag in *Debussy3.0* between the capture and display which was reported in their paper.

The next strategy is the *integration of the limitation* within the context of the performance. For example, flags were used as props in *Terres Rares*, though they were a slow and less than ideal projection surface. Still they were kept, as they provided an ethereal look to the sequence. P1 commented "*it worked well with the kind of content we wanted to display*".

The next step is to *bring the limitation forward* and make it part of the purpose of the piece. Though it was not used for the issue of tracking during jumps in *Debussy3.0*, P2 mentions the possibility of having the choreography feature jumps meant to trigger glitches. In that case, the glitches would become part of the performance by revealing the limitations of the technology.

Controlling What is Shown The second theme relates to the projected effect of the piece on the audience. It is the result of the grouping of codes attributed to the *Consideration of the final result* sub-theme and codes found in the current iteration. The second sub-theme, *Choice interconnection* was trimmed from the final theme map since it found little resonance with the new data set and was deemed less representative of the overarching narrative within it.

As discussed in the previous iteration, designers should carefully consider what they mean to show, which includes the limitations. Similarly as public interfaces in the taxonomy of Reeves *et al.* [85], strategies can be put on an axis based on the resulting effect of the limitation (Fig. 3.6). As such, a limitation that must not be seen will be avoided. If it is deemed less problematic, it can be hidden, or even tolerated. It then crosses into being more constructive, and can be revealed through an integration into the context of the piece. A limitation can even become part of the message itself by being amplified and completely becoming a feature to play with, with the complete awareness of the audience.

Considering immersive experiences through the scope of how elements are seen also finds echo in the literature. The necessity to make concrete dramaturgical decisions as to what is shown or hidden to the audience can be linked to the dimensions of visibility and immersion in the scenography of immersive virtual musical instruments [129]. Choices in the design of the piece that can be placed among these dimensions can have consequences on the entire experience, and designers should stay keenly aware of how they intend their production to fit within this frame.

3.4.3 Independent Case Studies

The third iteration aims to validate and complete the previous findings, in particular the strategies, by broadening the sources with material from the literature.

Methodology

The data set is the result of a selection among a corpus built through queries in the form of combinations of the keywords *Mixed / Augmented / Virtual + Reality + Performance / Experience / Theater*, and papers they cite or are cited by. A strict manual selection was then made based on several criteria. The data set represents the diversity of authors, technology, and artistic context of the field, while being recent enough that a majority of the technical challenges would remain relevant to current MRP designers. Each piece of material needs to provide sufficient detail and justification of the implementation process and design choices to enable a thematic analysis. Coincidentally, this selection process resulted in only Ph.D. and M.S. theses being part of the final data set, which can be seen as a symptom that this nature of report tends to be more narrative and to include more details on processes.

Using case studies unrelated to the person performing the analysis as the validation step also helps to break out of the previous framing of the strategies. Still, as part of the reflexive approach of the analysis, it is important to note at this point that while an effort is made in taking various perspectives into account, the results remain a reflection of an interpretation when conducting the analysis.

Like the previous iteration, the analysis began with a new codebook, but integrated the previous results in its later stages.

CAVE

The shared VR narrative *CAVE* premiered in 2018 and presents the opportunity to be studied through two points of view. Both DeFanti and Herscher have worked on it and describe aspects of its implementation in their respective Ph.D. theses [41, 56]. In DeFanti's thesis, the part dedicated to *CAVE* (pages 33 to 36) was selected, and in Herscher's thesis, for which the show occupies a larger portion, the focus was put on chapter 2 covering the constraints, implementation, and results of the experience (pages 6 to 34).

Though *CAVE* features no live performer, it gathers up to 30 users in a shared environment as a co-present audience. Each spectator wears a stand-alone headset which communicates with a server in order to share their position among the entire group. The audience can thus be represented as avatars in the VR environment during the narration.

An example of a challenge answered in the piece is related to large scale shared immersion. A large number of users immersed with headsets while in physical proximity would normally lead to collisions and tracking issues. This could however be circumvented by having the audience sit down. This approach is also able to draw from the similarity of the experience to live theater, which allowed the audience to better accept this restriction.

Mary and the Monster

Following the production of *CAVE*, Herscher *et al.* began working on *Mary and the Monster* with two objectives: allowing for more interaction for the audience without encroaching on the experience of the group, and designing an experience for both VR and AR. This production is analyzed through the third chapter of Herscher's Ph.D. thesis [56] (pages 35 to 64), which is dedicated to this multi-user experience that premiered in 2019.

Much like *CAVE*, the show is a recorded narrative experience with no live performer for 25 spectators in VR and 8 in AR. VR participants can see each other as customizable avatars arranged into three half-circular rows and are immersed in the content in full scale. Alternatively, the AR audience has no avatar since they can see each other, and sit in a single row around a round table on top of which the content is displayed at a 1/6th scale.

An example of challenge encountered by this production is what caused this smaller scale. The AR headset that was selected for the show featured a limited field of view for virtual elements. For this reason, spectators would be unable to see the entire virtual stage at life scale, which is why the designers chose to shrink the stage.

Wake

Wake is described as an asymmetric, co-located, co-present, mixed reality social experience. It is included in the analysis through the parts of Henson's Ph.D. thesis [55] that describe the preliminary experiments, the experience itself, and its design (pages 24 to 29, and 43 to 60).

The experience consists in the interaction of a participant and a dancer, along with a facilitator. The participant is immersed in a VR headset and is eventually able to see the non-immersed dancer as a volumetric projection on a plane using a depth camera. The hands of the participant and dancer are tracked and represented by rocks tethered through ropes either to other objects, their own hands, or the hands of the other person.

An example of challenge in *Wake* is what is suggested to be an interference between the infrared light used by the trackers and the camera. This resulted in a poor performance of the trackers when they were raised in front of the camera. While Henson notes that switching to a different kind of camera could solve the issue, she opted to deal with the technological limitation. The participants were warned that the interaction may be "*sticky*" so they could expect a "*weird*" behavior and not have their immersion broken. She also notes the helpfulness of a performer being present, as they can smoothen the overall experience.

Bury the Wren

Bury The Wren is designated as an extended reality project and theatrical performance, and was co-created by Kates and Christensen. The data set that was selected to study this project is a portion of the fourth chapter in Kates' M.S. thesis [59] (pages 167 to 283). This chapter describes in detail the design and technical choices that were made for the production of the project through Kates' *practice as research* method.

The show is described as a "*one-on-one experience*", where a spectator and a performer interact along a narrative structure. Initially, the spectator is fully immersed in VR and can not see the performer, who manipulates a tracker on which virtual objects are anchored. These objects, captured and rendered through photogrammetry, can also be handed to the spectator. In the final sequence of the performance, the cameras of the headset are used to display the physical environment along with the performer, bringing the spectator into AR.

An example of challenge in *Bury The Wren* is directly caused by the constraints of the choice of technology. Kates and Christensen chose to use a headset that featured external trackers, but were exposed to some inherent limitations brought by it, namely the passthrough camera. Its latency, resolution, and black and white image could have prevented any AR feature in the performance. However, they chose to look for a way to make use of the limitation, and decided to embrace the daguerreotype aesthetic it brought. As a result, they filtered and shaded the image during the AR sequence so it leaned even more towards their newly intended look.

Resulting Themes

As seen in Fig. 3.5, the third iteration of reflexive thematic analysis resulted in three themes: *Strategic Management of Limitations*, *Aligning Performance with Technological Constraints*, and *Managing Outcomes*. The final state of the themes, sub-themes, and retained codes can be found in Table 3.2.

Strategic Management of Limitations The same sub-themes as in the previous iteration were found, with one modification on the fifth sub-theme that was changed from *Play with the Limitation* to *Leverage Limitations*. This change stems from new codes found in the independent case studies pointing towards using a limitation as a tool for generating ideas, especially in Kates' work as she expressed "*I wonder if the disruptive seams might have provided us with some unknown usefulness*" or "*Working resolutely against technical common sense created the happy incident of the digitally-tarnished candlestick*".

Aligning Performance with Technological Constraints The second theme is the result of the reintroduction of codes and sub-themes that were discarded in the previous iteration. They were then not considered representative enough to remain within the final themes, but the addition of new themes from the independent case studies showed their legitimacy. The theme depicts the inherent challenges in projects involving both performances and technologies.

Working with collaborators from other disciplines is a recurrent matter in MRP, as it usually merges at least the domains of performing arts and technological design. During their interview, P2 recalled "*After some time we manage to understand what the other expects or how they view this relationship, thus moving beyond the initial perspective of the other discipline*". This reflects how cross-disciplinary collaboration may not be trivial and how efforts should be put towards reaching a common understanding of the objectives in the project. These efforts can be made either as part of the creative process or outside of it. For example, both Henson and P2 integrated their collaborators' communities by taking part in their classes and workshops on respectively Physical Integration and Dance. Kates undertakes steps promoting the collaboration as part of her process, reporting "*we ran the risk of [the developer] becoming disconnected from the primary objective of* Bury The Wren*: to tell a compelling story*", and making sure in response that the developer also fully takes part in the rehearsals. Following this approach then allows both the creation and implementation processes to inform each other throughout the life cycle of the piece.

Even with all collaborators on the same page, the production needs to accord dramaturgical and technological processes of conception and implementation. Comparisons between theater and MRP were made at several points, noting the potential of traditional techniques, with for instance Herscher describing how *Mary and the Monster* drew from the traditional practice of a "*blackout*" to adapt it in VR, with the improvement of making physically impractical changes to the set. Kates also comments on what technology added to her practice of scenography, writing "*This is a significant departure from designing in the traditional theatrical space where modifications on a large scale are usually impossible due to time and resources*". Here, she refers to the design of specific software tools that permitted great flexibility during the creation of the show. The design and use of such tools can thus become key in establishing an efficient workflow in the design and implementation of a MRP.

In the case of productions involving live performers, the role they play involves important specificities. First, they may not always be immersed in the same conditions as the audience, and might not be able to completely perceive the virtual environment. This is the case of *Terres Rares*, on which P1 described the various elements they put in place to help the performers interact with objects of limited visibility to them, such as marks on the floor. The involvement of live performers can however also bring initially unsuspected value. For instance, Kates shares "*In an effort to overcome uncanny programmatic movement we chose to map our object to a tracker so that when they were manipulated by a human holding the tracker they reproduced that believable movement in VR*". Henson also discusses how the presence of live performers could help mitigate the issue of infrared interference between the depth camera and the trackers through orchestration. This use of a performer shows that bringing live performance into a production, despite creating some challenges, can also enable potential answers to previously complex issues.

Managing Outcomes The last theme encompasses both the control of what is shown, which was present in the previous iteration, and the preserving of continuity. Continuity relates to Benford and Giannachi's trajectory framework [13], which helped Kates identify limitations related to transitions and their potential ruptures during the experience of the spectator. This consideration led to the modification of the transitional space between VR and AR into a version inducing a smoother visual transition. This same care about continuity throughout the show and the identification of problematic transitions also helped them identify potentially overlooked limitations. By considering the arrival of the participant as part of the experience, Kates was able to notice that the presence of the object tracker before immersion could be distracting and hid it in the darkness by altering the lighting. Preserving continuity also involves preparing for failures, such as in *Mary and the Monster* for which a monitoring feature was implemented in order to quickly identify any issue with the spatial alignment of the headsets. The same commitment to continuity also informed design choices in the production of *Animate*. In the face of the high risk of a headset losing its alignment with the physical space, the procedure to realign it was made simpler and thus quicker at the expense of precision.

3.5 Discussion

These results are built upon to provide a set of strategies for reconciling technical constraints and artistic intent, which can be used to analyze choices made during the design of existing MRPs but also to envision new solutions for challenges when designing novel MRPs. Limitations of this analysis are also discussed and how future work may build from these results.

3.5.1 MRP strategies design space

The main result of the analysis is the five strategies under which answers to challenges can be organized. These strategies build a spectrum based on the level of awareness of the audience, as seen in Fig. 3.6.

1. AVOID aims to side step the issue, or scale down the desired effect, for instance by reducing the complexity of either the environment or the interaction. Therefore, it is associated with the lowest

Table 3.2 The final codebook after the third iteration.

Themes	Sub-themes	Codes
Strategic Management of Limitations	Avoid Limitations	Simplify Interaction
		Restrict Physical and Virtual Space
		Restrict Environmental Complexity
		Omit Features
	Disguise Limitations	Rely on Limitations Remaining Unnoticed
		Actively Conceal Limitations
		Manage Attention to Prevent Noticing Limitations
	Tolerate Limitations	Accept Temporal Constraints
		Account for Technological Maturity
	Integrate Limitations	Align Limitations with Purpose
		Incorporate Limitations into Aesthetic Design
	Leverage Limitations	Embrace Limitations as Creative Catalysts
		Push Against Limitations for Artistic Discovery
		Allow Limitations to Shape Project Goals and Features
Aligning Performance with Technological Constraints	Cross-disciplinary Collaboration	Foster Understanding Among Collaborators
		Include Developers in the Creative Process
	Dramaturgy and Technology	Unintentional Integration and Creative Solutions
		Adapting Theater Techniques to Mixed Reality
	Collaboration with Live Performers	Support and Flexibility of Live Performers
Managing Outcomes	Controlling What is Shown	Simplifying the Audience Experience
		Optimizing Visual and Technical Impact
		Managing Complexity
	Preserving Continuity	Maintaining Seamlessness and Minimizing Disruptions
		Handling Discontinuities and Failures

amount of awareness.

2. DISGUISE is the reliance on the limited capacity of the audience to notice a limitation. It corresponds to a slight acceptance that the audience could notice the issue. This can be accomplished with more or less effort, since some limitations may already be hardly perceivable by participants.

3. TOLERATE corresponds to a neutral stance where no action is undertaken in reaction to the limitation. This is usually the marker of constraints preventing a different answer to a limitation, such as time or maturity of a selected technology.

4. INTEGRATE aims to alter either the limitation or the piece in order to match them together. This allows the limitation to be perceived by the audience as more of a feature than an unsolved issue.

5. LEVERAGE makes the limitation a key part of the show. Where the previous strategy simply merges the limitation with the rest, this one brings the limitation forward, and may even amplify it to convey a message.

It is important to note that these strategies do not define discrete categories, as some solutions may be seen as in-between two strategies. One should also consider that the interconnection of challenges and answers can lead to a challenge being answered in several ways, or to a decision answering several challenges.

Interestingly, this spectrum aligns with Reeves *et al.*'s taxonomy of public interfaces, which organizes them around two axes depicting how much effects and manipulations are revealed to observers [85].

Parallels can be drawn between these strategies and those proposed by Benford *et al.* as potential reactions to the connectivity uncertainty in *Can You See Me Now?* [10]. *Remove*, *Hide*, and *Exploit* are highly similar to AVOID, DISGUISE, and LEVERAGE. *Manage* is close to TOLERATE and DISGUISE, but focuses on establishing an alternative experience when the issue threatens to break it. This follows one of the recommendations in the sub-theme *Preserving Continuity*. *Reveal* has a key difference to INTEGRATE in how it is used, which may stem from the difference in context of both strategies. Instead of matching the limitation to the theme of the experience, it pushes to inform the user on the uncertainties of the system in order for them to make better decisions. The approach of revealing limitations and imperfections rather than avoiding or hiding them also aligns with the concept of seamful design [32].

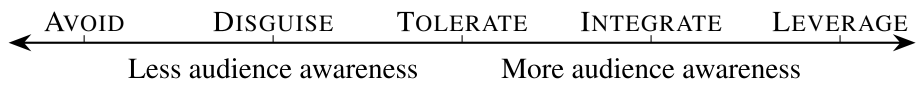

Fig. 3.6 Five strategies arranged in relation to the projected awareness of the audience towards the limitation.

3.5.2 Using the design space

Challenges that were previously identified can now be examined again under the scope of these strategies. Formalizing these strategies also amounts to defining a generative design space. As designers become aware of how their first approach may be placed on the spectrum, they can imagine other potential answers that would be placed elsewhere. This is not an attempt to point out better alternatives to choices in the productions, as some constraints may not have been considered, and some changes may have repercussions on the purpose of the pieces. Instead, we try to show the other possible approaches that can be taken in the face of such challenges.

The approach in *Animate* to let go of a complete synchronization of the virtual objects and to reduce the interaction in order to mask it can be placed between AVOID and DISGUISE. This can be argued by the issue of synchronization being avoided, while inducing another limitation. The audience can expect to see the same environment as their neighbor, yet can tell from their behavior that there are differences. This resulting secondary limitation thus finds a complementary answer in the form of reducing the interactivity of the scene, effectively hiding the difference in the environments from the audience. There was an opportunity to embrace the asymmetry instead, and leaning towards LEVERAGE by having the performers also seemingly interact with unseen rocks, while maybe even commenting that they can not see what the other sees. This would bring the focus of the audience precisely on the asymmetry, and allow them to experiment with their neighbors and discuss how their environments differ.

In *Terres Rares*, the latency between capture and projection could cause issues, had the performers moved too quickly. P1 commented on this issue that the answer was coincidental since the conception of the show already meant for the performers to move slowly, necessitating no further action in response to the limitation. Still, the back wall was made dark in order to limit the visibility of any potential light not hitting the intended screens. This leads to this combination of approaches being placed between DISGUISE and TOLERATE, since the designers relied on the limitation not being noticeable. A challenge like this could also have been met with an approach of amplifying the latency in order to use it, leaning more towards LEVERAGE. It is thus possible to imagine a performer interacting with a highly delayed projection in creative ways.

In *Debussy3.0*, audience members positioned differently would not see the virtual objects at the same position, breaking any potential spatial relationship between the dancers and the virtual objects. The answer was to remove any element from the stage, keeping only elements behind the stage, and some elements which did not depend on position close to the audience, in the form of water particles. This approach can be placed between AVOID and DISGUISE, since the designers limited the placement of the virtual elements in an attempt to reduce the impact a misplaced virtual element could have on the experience. An alternative to this approach may have been to work around the spatial relationship between the virtual content and the spectator, rather than the dancers. This way, the content acts as if it was personalized for each person in the audience, which is close to the INTEGRATE strategy.

In *CAVE*, the issues of collision and tracking with a large co-located audience are addressed by having them remain seated throughout the piece. While this can be seen as close to AVOID, it also induces the limitation of restricting the freedom of the audience. This secondary limitation finds its own answer in the form of the implementation of a theater-like setting. This helps the audience accept the piece as one adjacent to theater, with similar codes, such as having to stay seated, and matches the INTEGRATE strategy. Removing the issue entirely by scaling down the amount of spectators and thus allowing them to move around is a potential alternative that embraces AVOID, although it would require reconsidering the artistic ambitions of the show.

In *Mary and the Monster*, the limited field of view offered by the AR headset prevented participants from seeing the entire stage at once. In reaction, the virtual environment was shrunk down to a 1/6th scale. While the limited field of view is still present, it does not take away from the show anymore, which can be seen as an approach close to DISGUISE. Alternatively, the inability to see the whole stage could be used as a way to force the audience into choosing which of multiple happening actions to focus on. This alternative would then be placed towards LEVERAGE.

In *Wake*, the camera could interfere with the trackers, which led to issues with the tracking of the hands of the participants. They were warned of the "*stickiness*" of the interaction by the performers, which meant that they expected the issue and had better chances of an overall smooth experience. This answer can be placed between TOLERATE and INTEGRATE, since an effort was made towards letting the participant accept the limitation. A different possibility would be to use the fact that the interference happened when the trackers were raised close to the camera. By instructing the participants against raising their hands, designing the experience around the requirement to keep their hands low, or disabling the tracking when the hands are raised, the issue could probably be mostly avoided. This response would thus be placed between AVOID and DISGUISE.

In *Bury The Wren*, the aesthetic brought by a camera with poor resolution, latency, and in black and white was embraced. By adding shaders to the image, its imperfection became part of the feel of the scene, corresponding to an INTEGRATE strategy. Alternatively, the use of an external camera would have allowed a feed of better quality, side-stepping the issue, and corresponding to an AVOID strategy.

3.5.3 Other dimensions in the design of MRP

In the analytic process, some elements were dismissed as themes or sub-themes, but may be interesting to consider in the description of strategies. The appearance of a challenge and application of its answer has appeared to be possible at various stages of the life cycle of the production. For instance, challenges result from the choice of technology, or the budget available like in *Debussy3.0*, which typically happen early on. Other challenges emerge very late, like a lag that appeared when setting up on stage in the same production. Answers can also vary in timing, from early in the implementation process with the design of adapted authoring tools [102, 61] or sketching [27], to the preparation of countermeasures such as in *Animate*, or *Can You See Me Now?* [10], or even improvised interventions [91].

Drawing from Benford *et al.*'s trajectories [13], other potential dimensions to consider are the spaces, interfaces, and roles involved in the challenges and answers. Challenges may relate to either the physical, the virtual world, or both, and so do their potential answers. For example, the glitches caused by jumps in *Debussy3.0* relate to the physical world, while the problematic positioning of the performers in the virtual environment of *Wake* shares both spaces. More specifically, how participants are immersed and perceive each other can be considered, for instance by adapting Zappi *et al.*'s dimensions for immersive virtual musical instrument scenography [129]. Challenges can also involve different parts of the interface of the immersive technology and be related to tracking (*CAVE*), capture (*Debussy3.0*), display (*Mary and the Monster*), or networking (*Mary and the Monster*). Finally, strategies can be discussed through the roles involved in the performance, as people occupying different roles may encounter different challenges and be able to answer them differently. In this context, various roles can be taken by spectators [13], actors [91], or even designers [61].

The themes also bring up the existence of non-technological challenges stemming from the interdisciplinary collaboration of art and science. The need for an establishment of a common vocabulary, as pointed out by P2 and Kates is a well known issue both in the global context of art/science collaboration [38], and the specific context of MR applications, with strategies adapted to it [61].

3.5.4 Limitations and Future Work

The testimonies that were collected for this analysis are single perspectives on their respective projects, except for *CAVE* which was represented by two designers. Since the limitations and how they were addressed may not always be reported for reasons of size limits or perceived value for publishing, their identification often relies on details and anecdotes spread out through the documentation. A richer analysis would heavily rely on interviews in order to draw undisclosed ordeals that the productions went through. It could also greatly benefit from looking at different perspectives of the same project, particularly from designers who played different roles in the production. Furthermore, a greater reliance on in-depth interviews would allow one to include in the analysis productions that were not described in publicly available literature as thoroughly as those in this study. Most notably, this would enable the inclusion of productions with a more commercial outlook than the ones analyzed here.

A notable limitation of this study is that the iterative analysis stems partially from productions involving myself or one of my co-supervisors. This is important to keep in mind as it had a strong opportunity to influence the continuation of the process in light of the authors' my own practice, with a focus on specific aspects. Still, the analysis was conducted with this bias in mind and includes the third step as a way to address it.

The data corpus included works discussing productions between 2019 and 2021 in order to have a coherent technological background for the challenges they encountered. Future work can find value in looking at the potential evolution in challenges and strategies before and after this period.

Additionally, value can be found by investigating the generative power of the strategies presented in this paper. This could be accomplished in future work by accompanying the production of a MRP and observing how using this framework can guide the design process, and conversely if and how the framework needs to be expanded.

3.6 Conclusion

Through three steps of an iterative reflexive thematic analysis on a performance-led case study, two interviews, and independent case studies, five potential strategies were identified when approaching challenges in MRPs involving the friction between a desired outcome and the constraints of immersive technologies. These strategies are organized depending on the extent to which the audience should be aware of the limitations, and include: AVOID, DISGUISE, TOLERATE, INTEGRATE, and LEVERAGE.

The application of this design space for the analysis of previous MRPs was shown. When faced with such compromises, designers of novel MRPs can draw from these results to envision new approaches to their challenges. Though these strategies were found in the context of MRPs, they may also have broader applications. For instance, in the context of an educative environment DISGUISE might be used to maintain engagement or LEVERAGE may turn a constraint into a learning moment. Outlining the similarities within the field of MRPs, will also help designers draw inspiration from other works that would seem at first sight too different, and extend this exchange of ideas to and from fields beyond MRPs, like the field of musical expression.

Chapter 4

Decoupling Spaces for Musical Instruments in Co-located Mixed Reality

This chapter is adapted from [116]:

P. Uro, F. Berthaut, T. Pietrzak, and M. Wanderley. *Decoupling Physical and Virtual Spaces in Co-Located Collaborative Mixed Reality Instruments with gRAinyCloud.* In Proc. NIME, Canberra, Australia, 2025.

4.1 Application of a LEVERAGE Strategy to Spatial Issues

Among the many limitations that stem from the co-located aspect of some MRPs, one is the spatial relationship between physical and virtual spaces. By altering it, the virtual viewpoint of others no longer matches their physical body, which can introduce some unwanted perceptual confusion [34]. This issue is not specific to MRPs and can also be found in the adjacent context of CMR musical instruments.

There, the commonly found approach to this limitation is to harmoniously align the virtual environments of all users with the physical environment [70, 81, 54]. This amounts to removing any control over the relationship between physical and virtual spaces, and can be associated to the AVOID strategy identified in the previous chapter. This guarantees a consistency in position for all virtual objects, and a direct correspondence between any potential virtual representation of other users and their physical body. Although this approach provides collision prevention and is highly adequate for closely coupled collaboration, it also puts a tight restriction on what MR can enable for musical interaction. Additionally, since physical and virtual spaces are coupled, they become a resource to share, further restricting potential applications as conflict over space becomes a possibility.

Beyond the upsides and downsides of this straightforward approach, and with the guidance of the strategies framework, one can wonder what kind of CMR musical instrument could implement an approach closer to the LEVERAGE strategy. Exploring this idea would require to implement the decoupling of the physical and virtual spaces by allowing users to manipulate their viewpoint through navigation metaphors. Various navigation metaphors have been proposed in the context of MR, such as teleportation, flying, or path-planning [14]. A "grabbing-the-air" metaphor was implemented for a distributed VR collaborative instrument [9], but such metaphors have not yet been applied to CMR instruments. In the context of visual MR, where the physical other is visible in addition to the virtual space, decoupling spaces involves an avatar in addition to the physical body of the other as a representation of their viewpoint in relation to the virtual space [33] (see Fig. 4.1). Although this potential "dual-presence" may be a drawback for task performance [34], it should be less impactful in a creative context such as musical instruments, and may even be beneficial if part of a consciously designed LEVERAGE strategy. Overall, this approach enables countless collaborative scenarios that would not be possible with tightly coupled spaces, and has the potential for new collaborative strategies.

This chapter will present gRAinyCloud, a CMR musical instrument that applies this decoupling of spaces. It describes its theoretical context and implementation, before reporting on a self-study based on improvisation sessions in order to investigate emerging strategies and guidelines for future applications of this approach.

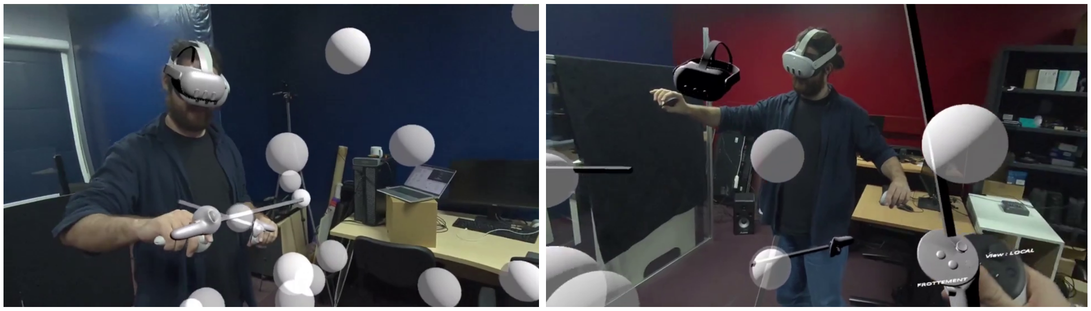

Fig. 4.1 First-person view of a user manipulating their viewpoint, either coupled (left) or decoupled (right), the other's point of view is represented by a virtual headset.

4.2 Musical Instruments in Co-located Mixed Reality

4.2.1 Collaborative MR Instruments

Collaborative instruments exist in extremely varied forms, occupying the broad spectrum of MR and even further into VR [21]. They can be characterized based on whether the collaborators are co-located, and if so, how they are embodied from each other's point of view. For instance, Bell implemented a networked instrument that allowed participant, located across Europe and represented as avatars, to explore a sound corpus arranged by timbre similarity [9]. In this context, participants are free to move around in relationship to the environment through a world manipulation or "point-tugging" [37] metaphor where grabbing the air allows pulling or pushing their viewpoint. With co-located collaborators, virtual viewpoints tend to be strictly tied to the physical movement of the user's head, without a way for them to be manipulated freely. This is the case for the performance *Trois Machins de la Grâce Aimante* [54], which features four instruments independently manipulated by four users who can see each other as avatars. The same applies to LeMo [70], where two participants could operate on the same virtual sequencing interface while also appearing as avatars, although the sound itself was not necessarily shared. This hard coupling of virtual and physical reference frames is also present in co-located instruments without avatar-related representation. It appears in ARLooper [81], with which several participants can record, place and edit audio samples as shapes in the virtual world. These shapes are shared between participants, but can not be edited simultaneously.

The lack of navigation options that appears in these co-located examples can be seen as an avoidance of the conflicts that could arise if the control of the spatial relationship between the physical and virtual environments was accessible. On one hand, this allows a guarantee that at any point of the interaction, the relationship between virtual and physical elements is maintained, which in these examples applies mainly to the participants themselves. If a participant could shift aside the entire virtual environment during a performance with LeMo [70], the avatars of either user would not appear where their physical body is, notably raising collision issues. On the other hand, this approach drastically limits the potential for interactions in co-located instruments. Enabling this decoupling of physical and virtual spaces relies on users accepting a dual representation of the other in relation to both the physical and virtual environments [34], while addressing the challenge of concurrently controlling the spatial relationship between both environments.

4.2.2 Concurrent Manipulation

When considering the shared control of an element, in this context the spatial relationship between virtual and physical environments, different sharing modalities exist. These have been classified as cooperation levels [68], described in the first chapter of this thesis (Sec. 2.2.1). An approach to address concurrent manipulation is constraint based interaction [26], with users having control over independent parameters of the same object, resultin in a form of degree of freedom separation. In the context of spatial manipulation, this would be akin to a person controlling the position of the virtual origin in relation to the physical world, while another could control its scale. For truly dependent interaction, a combination of the inputs needs to be computed. However, in the context of musical interaction, collaborators may want to each have complete control over the environment, which can not be accomplished through these strategies.

By introducing an individual mode, collaborators are able to independently manipulate their own viewpoint, effectively decoupling their physical and virtual spaces. This enables the alteration of viewpoints in terms of position, scale, and permits to quickly share someone else's view [33]. When co-located in VR, this mode runs the risk of collision between users [63] and may require complementary information or secondary avatars corresponding to the physical position of the others. Specifically in MR with a direct view on the physical other, this mode results in a "dual-presence", introducing a perceptual conflict that

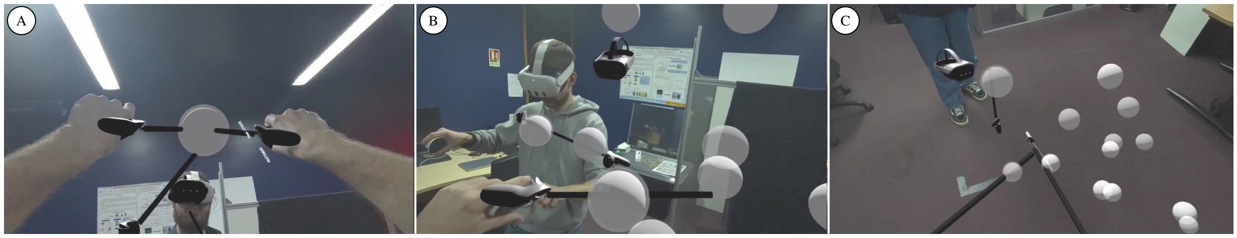

Fig. 4.2 Various first-person views of gRAinyCloud. A) : Two musicians play a virtual sound in a FIXED viewpoint. B) : One musician translated the virtual space to avoid physical conflicts with the other while playing the sounds, the virtual headset shows their actual viewpoint within the virtual space. C) : One musician zoomed out to be able to play fast sequences of sounds, the other musician's activity is therefore displayed as a small avatar within the virtual space.

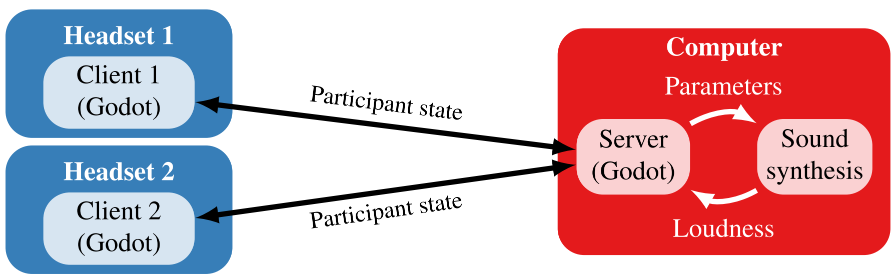

Fig. 4.3 Overview of gRAinyCloud's architecture.

may impact task performance [34], but may not necessarily hinder more creative activities. Applying this concept to musical interaction may bring forth novel strategies, which is why a musical instrument in MR that can leverage this independent decoupling of spaces was implemented.

4.3 Implementation of gRAinyCloud

gRAinyCloud is a MR musical instrument that allows two musicians to explore a set of sounds associated with virtual shapes placed in the physical space (see Figure 4.2), triggering them using virtual rods attached to the controllers of a MR headset.

The instrument was implemented using the Godot game engine, with a client on each headset communicating through network messages using OSC with a server running on an independent computer, which exchanges information with a PureData patch for sound synthesis (see Fig. 4.3).

4.3.1 Viewpoint decoupling

The main purpose of gRAinyCloud is to open novel opportunities of musical collaboration in MR by allowing musicians to manipulate the viewpoint on a shared virtual instrument while preserving the advantages of physical communication and cooperation. To do so, inspiration was drawn from previous research in XR, notably on multiscale collaboration and viewpoint decoupling. Multiscale collaboration, as presented by Le Chénéchal *et al.* or Nguyen *et al.* [65, 75], enables users to interact within a shared virtual environment from a usual viewpoint (at a 1:1 scale) and from a zoomed-out viewpoint (*e.g.* looking at a miniature of the virtual environment). In the most common setting, one user acts as a conductor, placing guides and cues for other users. Another inspiration is the decoupling of viewpoint which, as demonstrated by Sol Roo and Hachet [109], enables transitioning between various views of the same scene, each affording specific interaction techniques, such as coarse interaction of tangible artifacts in spatial augmented reality and fine-grained interaction of the virtual copies of the artifacts.

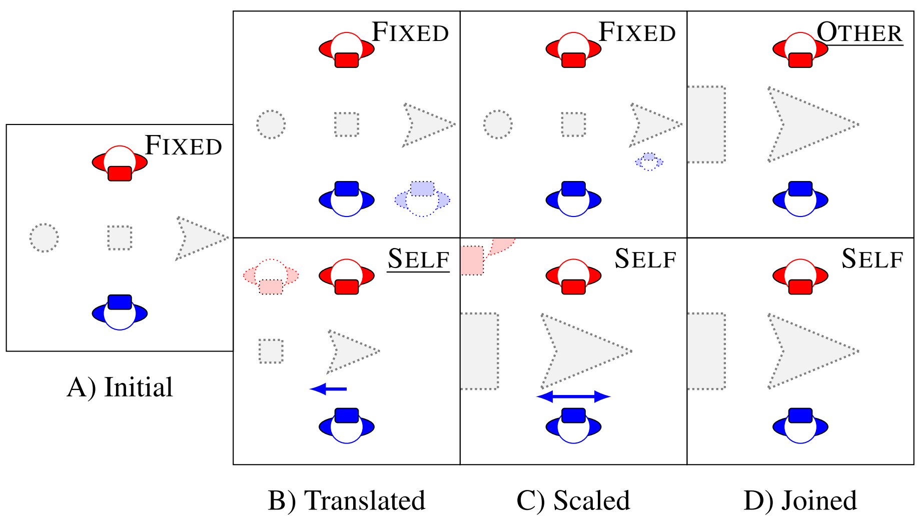

Fig. 4.4 Example sequence of changes in points of view (from Left to Right) : From an initial FIXED and shared view of the virtual space (A), the blue musician translates the virtual space (B), implicitly switching to their SELF view with avatars indicating the decoupled points of view, (C) then they zoom-in on one of the shapes, with corresponding changes in scale of the avatars. (D) Finally, the red musician switches to the OTHER musician's view, restoring a shared view of the virtual objects.

In gRAinyCloud, advantage is taken of these mechanisms for the design of MR musical instruments.

As shown in Figure 4.4, starting from a FIXED view of the virtual environment shared by both musicians, this viewpoint can be translated, rotated and scaled, therefore decoupling the virtual and physical spaces. This manipulation is performed using a "grabbing-the-air" metaphor [37], by pressing and holding the trigger on one controller for translations and two for rotation/scale. As depicted in Figure 4.2.B and Figure 4.4.B, virtual avatars of the musicians representing their headset and controllers consequently appear in the physical space to visualize the other's viewpoint. Depending on the transformation, the avatars may be translated, rotated and scaled (see Figure 4.4 B and C).

These manipulations result in three views : FIXED which provides the original alignment of virtual and physical spaces for all, SELF which corresponds to the musician's own viewpoint transformation, OTHER which allows one to see the virtual space positioned the same way the other has positioned it, thus realigning the two avatars and physical musicians (see Figure 4.4.C). As depicted in Figure 4.5, changes between SELF, OTHER and FIXED are performed either explicitly using a joystick (Left/Right to access the FIXED view, Top to access the OTHER view and Bottom to access the SELF view) or implicitly whenever the space is manipulated. Note than when one is in the OTHER mode, changes in viewpoint made by the other are dynamically applied, potentially resulting in changes in virtual objects positions while interacting, but also opening opportunities for collaboration.

Figure 4.2 shows how these changes are seen by the musicians and how they might enable different ways of interacting and collaborating, such as from moving the viewpoint in order to gain access to virtual shapes while avoiding physical conflicts (see Figure 4.2.B) or zooming-out in order to more easily access all virtual shapes for more dynamic controls (see Figure 4.2.C).

4.3.2 Sound synthesis and Mappings

The decoupling approach presented above could be applied to any virtual space shared using individual MR displays, like headsets or mobile screens. In order to test it, a first implementation in an instrument inspired by sound corpus exploration is proposed.

gRAinyCloud provides two modes of playing the virtual shapes, both activated when intersecting the shapes : 1) the *percussive* mode corresponds to the sound being triggered at a controllable tempo (and

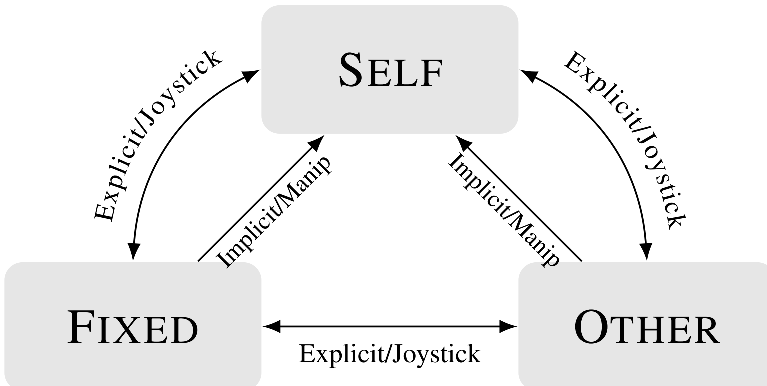

Fig. 4.5 Transitions between FIXED, SELF and OTHER views can be performed explicitly using the joystick to select the views or implicitly, since any manipulation of the virtual space will result in a switch to the SELF view.

multiple of a shared global clock) and with a controllable low-pass filter cutoff; 2) the *rubbing* mode corresponds to the sound being played as a texture (using granular synthesis) with controllable pitch and Low-Frequency Oscillator amplitude modulation. Users are free to switch between the two modes by pressing a button on the controller.

Parameters of these two modes are controlled through the rotation angles (pitch and yaw) of the controller relative to the performer-shape direction. In addition, the distance from the controller to the shape is mapped to the gain, this distance is normalized to the maximum reachable distance, which is the length of the instrument added to the scaled radius of the shape. Each controller of each musician can generate a voice for each shape, allowing for polyphonic interaction with the sounds.

After the synthesis, the loudness is transmitted to the server which displays it on the shapes themselves as a visual feedback for the musical interaction.

4.4 Evaluation

An evaluation was conducted to examine the impact of viewpoint decoupling on strategies for and experiences of collective music making. The approach was inspired by a self-use and improvisation method previously applied in the context of an Augmented Reality NIME [123].

4.4.1 Procedure

In addition to playing sessions and discussions during the design phase of gRAinyCloud, a supervisor and I conducted a more formalized set of three sessions of short (around ten minutes) improvisations interleaved with discussions. The playing sessions were not guided by rules or constraints, but instead built upon the discussions, creating new perspectives and the desire to test various strategies. The playing sessions were recorded from each headset to provide the musicians' point of view, along with a global recording of the physical space, and the generated audio, allowing for detailed examination.

4.4.2 Observed strategies

First, the collaboration strategies that emerged and were discussed can be described.

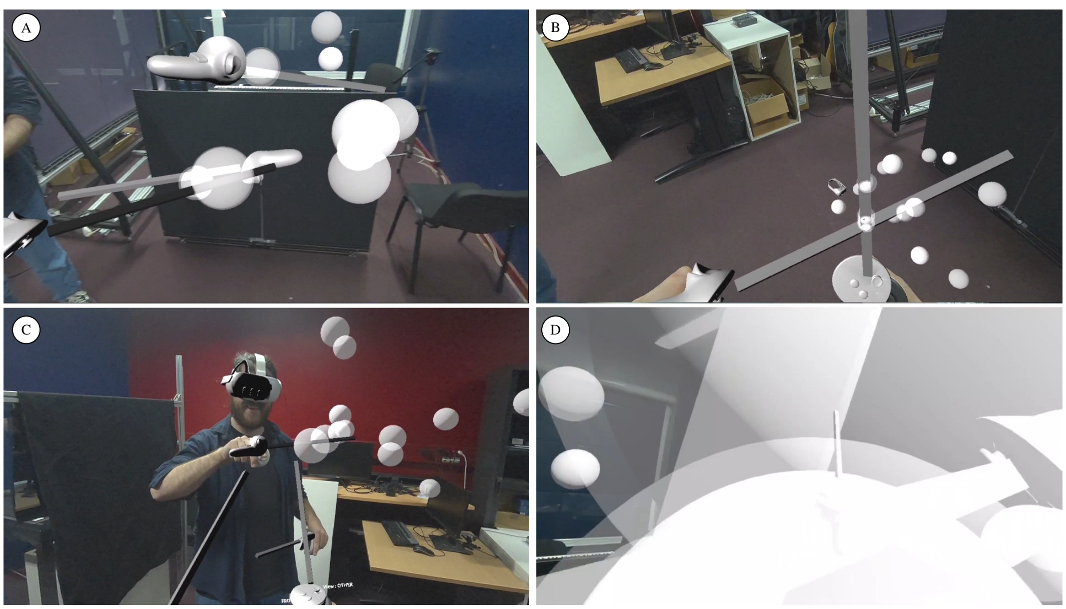

Fig. 4.6 Observed collaboration and playing strategies : A) Moving view in order to avoid physical conflicts while maintaining a visual feedback on the other's activity. B) Zooming-out in order to quickly access all the virtual space. C) Moving the space while the other is in one's view, therefore disrupting the interaction. D) Zooming-in to finely manipulate one specific sound, with the other's controllers and instrument appearing extremely large.

No use of the FIXED viewpoint

Throughout the three sessions, the FIXED viewpoint found no real use in contrast to the other views. Our hypothesis is that this is due to the size of the ensemble. Being only two musicians meant that reaching a shared viewpoint only required switching to the OTHER view. The FIXED viewpoint may have been more relied on with a larger number of collaborators, since the solution to needing to quickly synchronize their point of view would be more complex. This hypothesis still requires validation.

Avoiding physical conflicts

All potential contest over physical space was avoided easily as users could grab the environment and bring it with them while staying physically away from the other. In a way, this strategy results in recreating private spaces for each musician. This can be seen in Figure 4.6.A, where the user sees the other's controllers and interact with the same virtual objects while avoiding physical conflict.

Stealing from / Steering the other

This behavior also eventually caused someone to "steal" the objects away from the other when they were in OTHER view, as depicted in Figure 4.6.C. This is not necessarily negative since, when done on purpose, entangling the interaction of multiple musicians can lead to alternative creative strategies, such as with the Tooka [43] or the Perceptron [114]. This could also be used as a way to conduct a performance, if one of the musicians was given a dedicated role of navigating through the virtual space.

Rejoining / Leaving the other's viewpoint

The option to join or leave the other participant's view has been mainly used for two purposes. The first is to jump to their view and back in order to quickly swap between two setups, which is especially relevant when there is a large scale difference, and was for example employed to switch between viewpoints in Figure 4.6.B and Figure 4.6.D. The second is to meet the other physically on an object, usually to engage in a more closely coupled collaboration, which has also been performed manually, by progressively shifting the space in order to match the view of the other by placing their avatar on their physical body. This strategy emerged as an answer to the discrete change disrupting an ongoing interaction, making it possible to meet the other physically without interrupting the current interaction with an object.

Zooming in and out

Along the sessions, large changes in scales were used to access other modes of interaction. Zooming-out, *i.e.* reducing the scale of the virtual space, was for instance used for fast selection among a set of small shapes. Zooming-in, *i.e.* increasing the scale of the virtual space, allowed for fine-grained control over a single very large shape. In both cases, the changes resulted in modification of the other's avatar, appearing either as a very small character interacting on a small part of the space (see Figure 4.6.B, or as a large overlooking figure (see Figure 4.6.D with the giant controller and rod on the right). Both had an impact on the collective music making experience and behavior, for example resulting in playful interactions with the small scale avatar.

4.4.3 Discussion and guidelines

From the formal playing sessions and the overall design process, a number of guidelines for future investigation of decoupling in CMR instruments can be provided.

Favor awareness of the others and of the virtual space

The first relate to awareness, *i.e.* the perception and knowledge of others' activity and status within the instrument. The current view should be visible at all times in order to avoid the scenario of being "stolen from" when in the view OTHER. Displaying the view as a small label next to the controller was not sufficient in these sessions, and a more immediately available channel should be used, like for instance changing the aspect of the virtual elements. Complementarily, having the view of the other readily available, especially when they see one's own view, would help keeping the users aware of the potential for a "stealing from" scenario. Additionally, the virtual representation of the other is an important consideration for maintaining a sense of copresence, even in CMR [118]. In an exploratory instrument such as gRAinyCloud, helping users quickly ascertain the current transformation of the structure becomes especially important when they can jump from a view to a completely different one. In this case, changing the shapes or colors of the elements would help, but it could also be achieved through non-interactable virtual guides.

Increase expressiveness in decoupling

As seen in the strategies, a discrete transition between views is more appropriate without an ongoing interaction, and allowing a smoother or controlled transition between the viewpoints is desirable. Implementing a way to select other musicians either physically or through their avatar can be useful, especially when considering scaled up contexts with more collaborators. Finally, the approach of a personal and shared version can also be applied to the audio itself [70]. In this case, it is possible to imagine controlling whether our own audio output was heard only by ourselves, or if it was broadcast to all, with the potential for listening in on what others might be playing. This would allow a part of private experimenting, which is especially relevant to exploratory instruments such as gRAinyCloud.

4.4.4 Limitations

The strategies and guidelines reported in this study stem from a self-study approach of a supervisor and I on a specific instrument. They would gain from being completed through studies investigating the practice of a more diverse group of participants, and with different instruments leveraging the decoupling of physical and virtual spaces. gRAinyCloud remains very simple and lacks the complexity of a more mature instrument, which could see emerge more elaborate strategies. Studying this approach through a variety of instruments may also reveal different strategies that are not linked to the exploratory nature of the instrument, or to other design choices such as the length of the rods, which may have mitigated physical conflicts.

4.5 Conclusion

This chapter investigated how decoupling virtual and physical spaces in CMR instruments can shape musical collaboration. The instrument gRAinyCloud was introduced to enable musicians to transition between different viewpoints, providing alternative ways to structure interaction and performance Its implementation and evaluation revealed some collaborative strategies that emerged over a short period of practice.

A first perspective for future work is to refine visualization and interaction techniques to better support awareness and expressiveness in decoupled collaboration. Understanding how musicians interpret and respond to these shifting viewpoints could inform the design of more adaptable and intuitive MR instruments.

Another perspective concerns the effect of decoupling on collective music making. Fixed visualizations of others' activity, as explored in [15], have been shown to influence how performers engage with their collaborators. Here, musicians can dynamically switch between avatars for activity visualization and direct interaction with physical performers. Further study is needed to understand how this approach influences musicians' perception of collective expression and the ways they structure their performance.

The case of gRAinyCloud illustrated how decoupling physical and virtual spaces can reshape collaboration in CMR. Yet this alternative way of designing shared interaction also raises questions about how participants experience being together in such environments. If spatial coupling is no longer guaranteed, how do collaborators establish and maintain a sense of copresence? Addressing this issue requires methods capable of capturing not only what participants report, but also how they behave in real time. The next chapter therefore turns to the challenge of evaluating copresence in CMR, proposing and testing a protocol that combines self-report, behavioral measures, and interviews.

Chapter 5

Behavioral Measures of Copresence in Co-Located Mixed Reality

This chapter is adapted from [118]:

P. Uro, F. Berthaut, T. Pietrzak, and M. M. Wanderley. *Behavioral Measures of Copresence in Co-located Mixed Reality*. IEEE transactions on visualization and computer graphics, 31(5):2848–2858, May 2025.

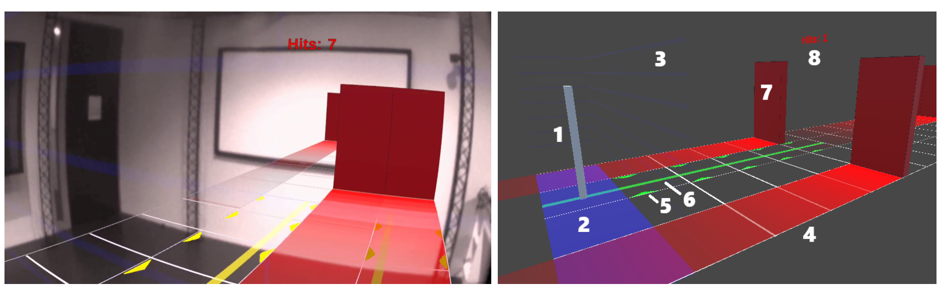

Fig. 5.1 First-person and annotated views of the virtual environment. Left: Point of view of a participant seeing several walls approaching them. Right: An annotated view of the virtual environment: The participant (1), task area (2), guard rails (3), lanes (4), lane feedback (5), position feedback (6), walls (7), and score (8)

5.1 Measuring Copresence

As seen in the previous chapter, allowing a different perception of a virtual environment brings forth new ways for collaborating individuals to interact. In this case, it was through an independent decoupling of virtual and physical spaces, but other works have presented contexts where co-located users might not be immersed in the same content, like with different views of a virtual object based on roles [76], or control over private view filters [84]. While in these examples users are meant to be aware that others are exposed to a slightly different environment, some situations can be ambiguous by design. A similar situation can be seen in an extreme case with *Animate*, which was presented earlier in Section 3.3. In it, an ambiguity is built around the performers who appear as though they perceive the virtual elements despite not wearing any immersive equipment, along with other participants not actually seeing the same floating virtual rocks. In the same show, a more moderate example can be identified as the virtual elements seen by the audience members behave similarly but are not the same for everyone, unbeknownst to them. Such cases raise a yet unanswered question: to what extent do users in CMR feel like co-located others are sharing the same experience?

The psychological phenomenon of *presence* is one of the main aspects of evaluating these experiences, and it has been extensively studied and defined in contexts beyond the strict scope of MR. Specifically, the focus is put on *copresence*, defined as "*The degree to which the observer believes he/she is not alone and secluded, their level of peripherally or focally awareness of the other, and their sense of the degree to which the other is peripherally or focally aware of them.*" [20]. It should not be conflated with *social presence*, its parent dimension, which past studies have mainly employed to investigate whether a virtual entity is perceived like a person [78]. This inquiry becomes hollow when the other in question can be directly perceived as a physically present human, such as in CMR. The difference in focus brought by a co-located context makes this distinction key, since it also means that measuring tools and methods introduced for social presence will not be blindly translatable. Adapting known tools also needs to account for the context of MR, which has been pointed out to introduce new challenges when compared to Virtual Reality (VR) [87, 122, 106]. Notably, MR presents the potential for various elements in the environment to be felt with varying degrees of presence, creating a form of mixed presence [44]. Thus, the aim is to determine how copresence applies to CMR, and how it can be measured in this context.

In the field of presence research, the most commonly found measuring mean consists of subjective measures such as self-report questionnaires. According to a recent review by Souza *et al.*, 86% out of 239 user studies on presence employed subjective measures while only 12% completed them with objective measures [110]. This is also the case for studies on social presence in MR [60, 128, 36, 7]. The issues of questionnaires are well known and often commented upon, pointing out their complexity with naive participants, that they can either only happen after the experience or risk breaking the sense of presence of the participant [52]. Another criticism is that a questionnaire is capable of eliciting in a participant the feeling that it was trying to measure in the first place [105].

The alternative is then to add complementary measures to the protocol, such as physiological or behavioral measures, each with their own advantages and disadvantages [52, 106]. Active tasks have been found to affect physiological data through engagement in the task, the stress from questionnaires, and artifacts caused by movements [48]. However, while behavioral analysis allows for an approach that can be both qualitative and quantitative, it is heavily dependent on the context or task and must thus often be adapted [44], which is the case for CMR. Still, its versatility and non-intrusiveness make it a prime candidate for investigating copresence in CMR. It does however bring the challenge of selecting a task that is both able to provide behavioral measures of copresence, and to leave enough room for this feeling to fluctuate. The task and the design of the study also need to account for the likely issues of fatigue caused by wearing a headset and a learning effect known to happen in behavioral studies [111, 3]. Most behavioral measures of presence found in the literature rely on stressful stimuli [48, 82, 113, 67], or lack a shared experience context that would allow the evaluation of a feeling of copresence [104, 101, 107, 115, 45]. This calls for the establishment of behavioral measures and an example of a task that would allow the evaluation of copresence in a CMR context.

In this chapter, a new protocol is presented to allow the evaluation of copresence in CMR. It is built around the assumption that a higher sense of copresence will induce changes in behavior in users during a collaborative task. We expect users to perform physical movements to help each other avoid virtual obstacles coming towards them in CMR. Positional data is analyzed in conjunction with answers from the participants to interviews and a questionnaire drawing from established copresence questionnaires and adapted to the context of CMR. This protocol is applied in a user study comparing two conditions eliciting low and high copresence through positional visual feedback to identify which measures are relevant for its future iterations.

5.2 Protocol

The approach to this protocol draws inspiration from social behavior in everyday life situations. It assumes that the feeling of sharing a space with someone will induce changes in behavior. More specifically, it recreates the situation where, when walking alongside another person and seeing that this person will encounter an obstacle in the physical environment, one will tend to step aside in order to leave room for the other. This behavior depends both on the perception of the other person, but also on the perception of a shared physical environment that affects the other person.

Here this situation is transposed to a CMR environment for two users, in which the obstacles are virtual but participants remains physically present. The hypothesis is that physical movements performed to help one another avoid virtual obstacles will reflect the level of copresence, since it will indicate that participants feel that they are together within the same physical and virtual spaces. If the copresence is lower, the expectation is that participants forget about each other and collectively fail to adapt their behavior and avoid obstacles.

In order to evaluate the effectiveness of this protocol, an experiment was conducted in which conditions are known to have an effect on copresence, namely added visual feedback that act as an avatar to reinforce the integration of physical users in the virtual environment [39, 77].

5.2.1 Task

The task environment consists of four virtual lanes with a width of one meter each, and an area at the edge of these lanes with a depth of one meter (see 5.1). This is the task area, colored in blue, and is where the participants are instructed to stay. The task consists of avoiding a sequence of walls appearing on the

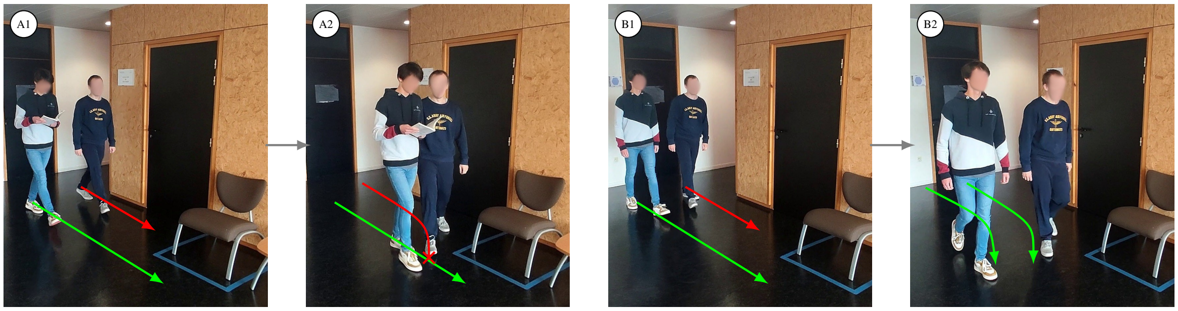

Fig. 5.2 The proposed protocol recreates the physical situation where one person adapts their behavior to help another avoid an obstacle. The expectation is that different behaviors should occur depending on if the person is not aware of the other ($A1 \to A2$) or if they are ($B1 \to B2$).

other edge of the lanes every four seconds and advancing toward the participants at a speed of 2 m/s. A transparent grid surrounding the area gains opacity when the participant approaches it so they do not exit it by inadvertence, serving as guard rails, and a message indicating to go back to the blue area is displayed to the participants whenever they exit it. Walls will appear on the lanes as red rectangular obstacles and need to be dodged.

If a participant is on the same lane as an obstacle when a wall hits the blue area, it is recorded as a hit. Similarly, if the two participants are on the same lane when a wall reaches the blue area, a hit is added whether there was an obstacle in the lane or not. The goal of the participants is to minimize the amount of hits by the end of the experiment. Finally, they are asked not to cross each other during the task for safety and analysis reasons.

Wall Configurations There are six different kinds of walls configurations, as two obstacles are placed among four lanes. Each of these configurations only leaves two spaces for the participant out of the four lanes. With this setup, a wall is equivalent to a pair of positions which the participants must reach. Then, two successive walls represent a movement required of the participants, with a starting wall and an ending wall. In order to cover all possible movements, each unique pair of wall configurations must be represented. It is however possible to reduce the amount of total walls necessary by considering that an ending wall can serve as a starting wall at the same time. We can thus build five sequences of walls that each covers all movements possible with 31 walls, including a first wall that serves only as a starting wall. Each pair experiences all five sequences in counter-balanced order.

Forcing Walls A particularity of the task is the existence of *forcing* situations, which is whenever a participant must move in order to leave room for their partner so they can avoid a hit (see Figure 5.3). Such situations require the participant to not only focus on their own movement but to also take into account the existence of the other person within the virtual environment. These situations are precisely the ones where the tendency to leave room to the other can be measured, and are the ones where measures and analysis should be focused. Flooding the pair of participants with *forcing* situations is however risking that its resolution by them becomes too conscious. This is why each wall sequence consists of all 30 possible situations instead of a repetition of the 10 *forcing* ones.

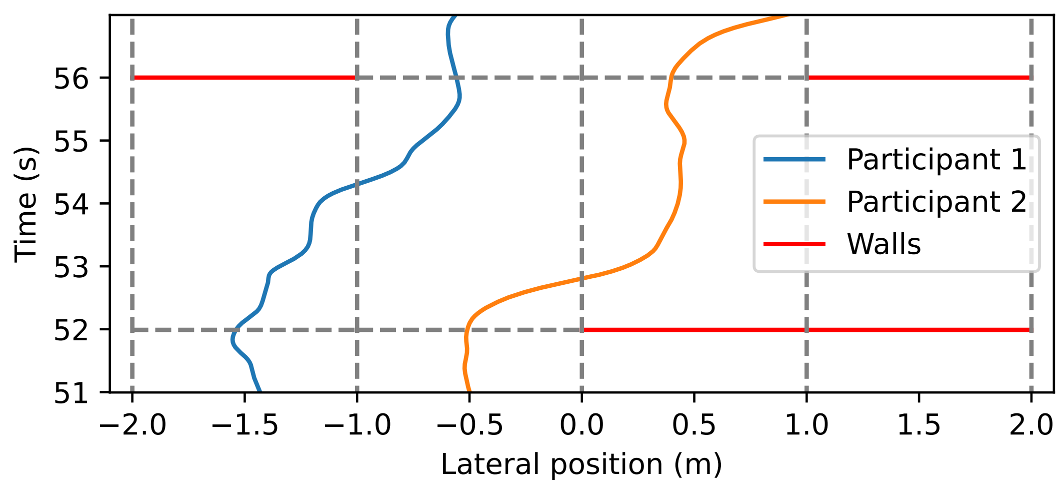

Fig. 5.3 Positions of two participants attempting through a forcing wall. They are positioned by the first wall at t=52, then the right participant could stay in the same lane without hitting a wall at t=56, but must step right in order to leave room to their neighbor.

Within or Between? Within-subjects design has been shown to detect existing embodiment effects that between-subjects design was not able to reveal [88], making it the favored design for this kind of protocol. It would allow less individual variance in the data, reduce the amount of participants necessary, and enable richer discussions during a post-experience interview. It does however lengthen experimental sessions, which should be avoided in contexts prone to cybersickness such as MR, and is vulnerable to transfer across conditions. This task is especially liable to learning effects across conditions, as participants may identify *forcing* situations over time and develop strategies, which makes within-subjects design inappropriate. Many other behavioral studies of presence or social presence in MR have gone with a between-subject design presumably because of this learning effect [111, 3, 73]. In order to alleviate this limitation, an extra *perspective block* is added after the participants fill in their questionnaires. Its purpose is to allow them to experience alternative conditions in order to enrich the discussion that follows in the semi-structured interview, and it should not be included in the quantitative analysis.

5.2.2 Metrics

The protocol integrates both subjective and objective measures through the use of a questionnaire, an interview, and behavioral measures.

Questionnaire

The post-task questionnaire provides a subjective measure of copresence by drawing mainly from the NMMSP questionnaire [20]. Since the NMMSP was designed in direct relation to the definitions of social presence and copresence used for this study, it serves as a perfect basis for a questionnaire focused on copresence in CMR. As the task is a game, it is logical to also consider de Kort *et al.*'s adaptation of it to gaming situations [40], in particular its behavioral engagement section, which is most relevant to copresence.

The self-report questionnaire consists of 24 questions (see Table 5.1) built out of four dimensions presented in random order:

Cognitive Level The separation between three Cognitive Levels Situation, Behavior, and Intention mirrors de Kort *et al.*'s addition of a question around mutual intention understanding to the NMMSP. It also matches the Psychological Involvement in part described by Biocca *et al.* as the belief of having insight into the intentions, motivations, and thoughts of the other [20].

Cognitive level	Dependence level	Target	Positive question
Situation	Awareness	Attributed	I was aware of the situation of the other person in relation to myself and the environment.
Situation	Awareness	Reported	My neighbor was aware of my situation in relation with themselves and the environment.
Situation	Dependence	Attributed	I was behaving according to the situation of the other person in relation to myself and the environment.
Situation	Dependence	Reported	The other participant was behaving according to my situation in relation to themselves and the environment.
Behavior	Awareness	Attributed	I was aware of the behavior of the other person.
Behavior	Awareness	Reported	The other participant was aware of my behavior.
Behavior	Dependence	Attributed	I was behaving according to the behavior of the other participant.
Behavior	Dependence	Reported	The behavior of the other person was according to mine.
Intention	Awareness	Attributed	I was aware of the intentions of the other person.
Intention	Awareness	Reported	The other participant was aware of my intentions.
Intention	Dependence	Attributed	My behavior was related to the intentions of the other person.
Intention	Dependence	Reported	The other participant was behaving according to my intentions.

Table 5.1 Positive versions of the questions in the questionnaire ordered by dimension. Questions were presented to the participants in shuffled order.

Dependence Level Decomposing Dependence Level into Awareness and Dependence allows including Awareness as a key part of the definition of copresence by Biocca *et al.* [20], that also relates to Attentional Allocation and Mutual Awareness, two factors indicating copresence. On the other hand, Dependence relates to the Behavioral Engagement dimension of social presence, with its factors being Behavioral Interdependence, Mutual Assistance, and Dependent Action.

Target The Target of the question can be either the participant themselves (Reported), or their partner (Attributed). This follows most copresence and social presence questionnaires which feature mirrored versions of their questions, one where the participant reports on their own feelings, and one where they conjecture on the feelings of the others.

Positive/Negative A redundancy is added by presenting both a positive and negative form of each question. In the analysis, the negative answer is inverted, and any pair of answers that suggests an inconsistency with a difference of more than two points is dismissed.

Interview

The interview is semi-structured in order to nurture discussion and gather comments and opinions from the participants themselves that can complete the interpretation of results. It consists of different parts aiming at various aspects of the experience of the participants.

The first questions are aimed directly at the condition that is being evaluated. Their answers should help support analytical results with the own opinions and feelings of the participants themselves.

Then, in case specific behaviors or errors committed by the participants were identified by either the interviewer or participants, they are pointed out and discussed. Namely, asking for the perceived cause of specific errors helps to understand the cognitive processes of the participants by having them confront their subjective and objective comprehension of their neighbor [29]. This serves as a stepping stone for a more general discussion on the potential strategies that both participants developed. Encouraging the participants to exchange on these strategies is helpful for exposing potential mismatches in their respective approaches of the task.

A few questions give an opportunity to the participants for expressing to what extent they felt copresence in relation with their partner. Examples of such questions are to ask them if they felt like they were playing by themselves, or if they stopped feeling that a person was standing next to them.

Finally, potential anticipated limitations are addressed, like by inquiring on the interface or the effort necessary for the task.

While some of these answers could be collected through a questionnaire, obtaining them through semi-open questions enables the emergence of issues or effects not anticipated by the experimenters.

Performance

A link between task performance has been pointed out by a meta-analysis of 80 studies that validated a model in which social presence positively impacts *flow*, a mindstate of high focus [124], which in turn positively impacts task performance [80]. It shows that task performance is also subject to task effort, which is impacted by trust, and in turn social presence. We can thus conjecture a potential link between copresence, as a subcomponent of social presence, and task performance.

In this protocol, participants are given an explicit goal, which is to avoid the walls by not being in the same lane as their partner or the wall when it reaches the task area. A performance measure is thus obtained by using the ratio of walls on which an error was committed. Since both participants share a score, and an error may be caused by either or both of them, the performance metric is also shared among the pair.

Trajectory

The position over time of the participants can be modeled as a trajectory in order to allow analysis. Here, a trajectory is defined as the sequence of lateral directions a participant moves towards. The transitions between these directions follow a hysteresis model. By default, a participant is still, and when their velocity reaches a threshold of 0.5 m/s, they are considered to be moving in that direction. Once they are moving, they need to go under a velocity of 0.25 m/s in order to be considered still again. The trajectory is further simplified by allowing a change in direction only 0.5 s after the last one. This model was established through trial and error until each block was adequately characterized while displaying stable trajectories based on visual confirmation, an example of it can be seen in 5.4. With it, the trajectory of each participant can be analyzed through its temporality and the distance kept between both.

Temporality Gait characteristics such as stride length, width, and speed have been used as behavioral measures in threatening environments [82, 67], and while there is little threat in this scenario, a link may be found between movement speed and copresence. A higher speed can also be interpreted as a form of task performance, or even a marker of flow, both previously linked to positive social presence [80].

For each trial two delays are computed around the starting and ending wall (see subsection 5.2.1). The Starting Delay is the time spent between when the starting wall reaches the task area and the moment the participant begins moving. If the participant started moving before the wall, the Starting Delay will be negative. Complementarily, the Ending Delay is the time spent between the end of the participant's

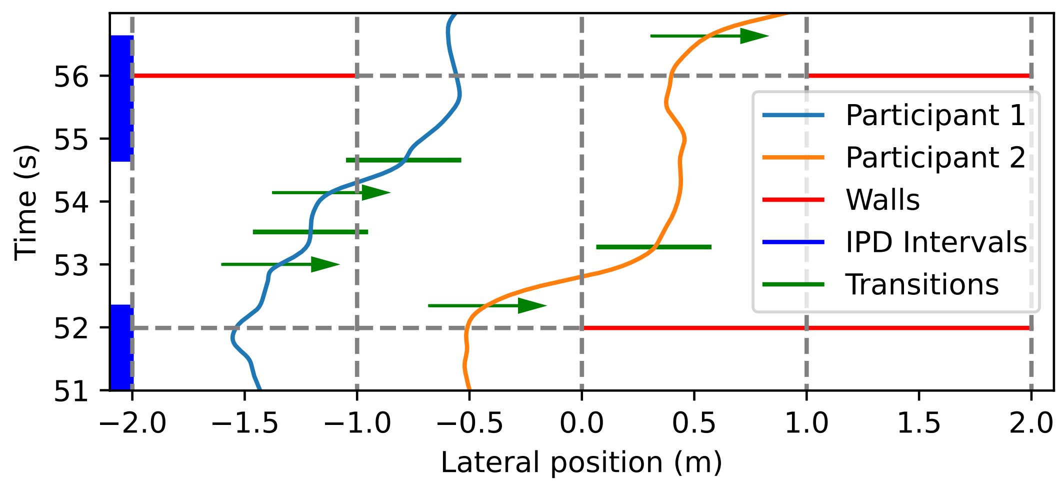

Fig. 5.4 Positions of two participants going through a forcing wall (the same as Figure 5.3). Trajectory transitions are represented as green lines or arrows, with a line representing transitioning to a still state, and an arrow being in movement towards the direction it points to. The blue rectangles to the left indicate the intervals on which the interpersonal distance is computed. Note that the short interval around t=54 is not computed as participants are not in adjacent lanes, and the interval lasts for less than one second.

movement and the arrival of the ending wall. Again, if the participant stops moving after the wall reaches the task area, the Ending Delay will be negative. Walls featuring an error or with an unexpected trajectory are excluded from the analysis as outliers.

Interpersonal distance The maintaining of an interpersonal distance can be seen as a mark of a greater sense of copresence or social presence [5, 6]. Notably, distance between collaborating people has been shown to be different with co-located and remote participants despite a similar self-report of copresence [83].

A measure of interpersonal distance can be extracted by averaging the distance between the participants during all periods where they are both considered still and in adjacent lanes. Such periods that lasted less than one second are excluded so that both participants are considered to be in resting position.

Gaze

Studies of gaze in social situations in virtual environments point towards an avoidance of another's gaze with higher social presence [4]. For a different reason, in the context of this protocol, looking towards the other may also be interpreted as a lack of mutual awareness. The reasoning behind this interpretation is that looking towards the other in a game where the rules dictate their position can be seen as a mark of the participant not being fully confident or aware of the potential position of their partner. In both cases, more time spent towards the other is expected to reveal a lower sense of copresence.

To measure the gaze of a participant, the direction towards which their headset faced is recorded during the entire task. Six areas towards which this direction can point are defined (see Figure 5.5): FORWARD-CENTER (FC), FORWARD-DOWN (FD), FORWARD-UP (FU), PARTNER (P), OPPOSITE-PARTNER (OP), and BACK (B). The FORWARD area (F) spans 120° horizontally. This value is based on the 60° a participant must turn before a person standing on a perpendicular line next to them enters their field of view, which gives an indication that the participant is looking far enough to be able to see their partner. This area is further divided into FC, spanning 100°, with FU, and FD, respectively above and below. The choice of the size of the division is also made with the field of view in mind, as a participant must look 50° downwards before their feet enter their vision. This distinction allows separating when a participant is looking at the walls ahead from when they are looking at their own feet to adjust their position. The B area also spans 120° and is located directly opposite F. The direction of the lateral areas, P and OP, depend on the side of the participant during the task. For instance if the participant is on the left of their partner, P is towards the right.

5.3 Experiment: Effect of visual feedback

In order to evaluate the effectiveness of the protocol, a first study is implemented on the effect of visual feedback informing of the others' activity. The experiment is designed on the basis that the existence of a visual feedback of the other person causes a higher level of copresence [39]. Having a condition with high copresence and one with lower copresence allows the evaluation of each metric selected in our protocol.

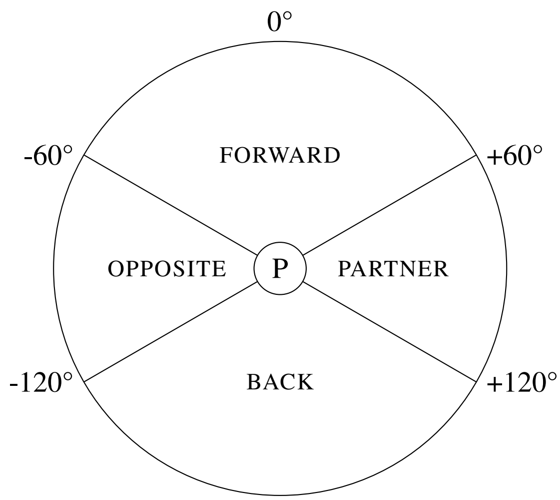

Fig. 5.5 Top-down view of the delimitation of the gaze areas assuming the partner is on the right. The FORWARD area is vertically split between CENTER, UP, and DOWN.

On the lanes, discreet and continuous feedback inform the participants of their position in relation to the virtual environment. Arrows are displayed inside the lane in which a participant is and move whenever they go into a different lane, providing discreet feedback. Simultaneously, a line extending from the blue area follows the lateral position of the participant, serving as continuous feedback. The two conditions relate to these feedback modalities: in the condition FEEDBACK, each participant is able to see their own position feedback as well as their partner's, while in the condition NO FEEDBACK, the participants can not see their partner's feedback and can only see their own.

With these two conditions, the experiment follows a between-subjects design with 2 Feedback × 5 blocks × 30 Walls (of which 10 are *forcing*).

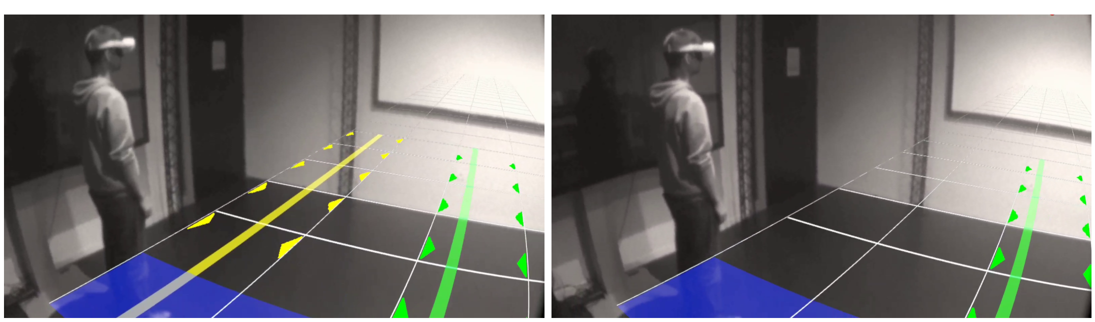

Fig. 5.6 Both feedback conditions (FEEDBACK left, NO FEEDBACK right), a participant can see their own feedback in both conditions.

5.3.1 Apparatus

Both participants wear a Meta Quest 2 headset in communication over OSC on a shared Wi-Fi network with a server hosted on a laptop and with each other. The headsets allow MR by providing access to the black and white external camera feed, which is displayed under the virtual environment.

The participants also wear Srhythm NC25 headphones that are playing pink noise throughout the task. The virtual environment was developed on Unity (2020.3.36f1) with the Oculus package for MR and the extOSC package for communication between the headsets and server.

Due to technical constraints, the physical room is smaller than the lanes on which the obstacles advance, meaning that they are displayed in front of a physical wall despite being positioned behind it. Orthogonal subdivisions were added to provide positional context for the virtual obstacles in order to limit this issue.

5.3.2 Procedure

After the participants are equipped with the headsets and headphones, the virtual environment is described to them, along with the task and its rules. The participants are instructed not to communicate verbally with each other during the task in order to allow potentially conflicting approaches among the pair. For this intent, white noise is played on the headphones worn by the participants in order to mask the sound of the footsteps of their partner and most attempts to communicate with sounds. They are asked to adjust the volume of the noise themselves, so it is loud enough that it covers footsteps as demonstrated by the experimenter, but not so loud that it becomes uncomfortable.

A short training phase then begins, where they avoid ten walls, each blocking one lane at a time. After an optional break, the participants go through all five blocks, also interleaved with optional breaks during which the participants may take off their headsets and headphones if they feel the need to. Each block consists of 31 walls, each covering any two of the four lanes, and lasts 135 s, for a total of 12 min including the training and without the flexible pauses. Once all blocks are completed, the participants take off their headsets and headphones, and each one completes individually the questionnaire described in Section 5.2.2. An additional block using the condition that was not selected is then completed in order to grant to the participants a perspective of the other condition. Finally, an interview between the experimenter and both participants is conducted.

5.3.3 Participants

In total, 15 pairs of participants were recruited. One pair was however excluded from behavioral analysis because they exhibited outlier behavior by playfully going into the lane of their partner without necessity. The data from a total of 28 participants was therefore analyzed, of which 7 female, and 21 male, aged between 19 and 68 with a median of 30 (std=12.3). Seven pairs completed the task in the NO FEEDBACK condition, and seven in the FEEDBACK condition. Of these pairs, 3 were with mixed gender, leaving 2 all female pairs, and 9 all male. Gender was balanced between conditions, with 4 female participants with feedback, and 3 without.

5.3.4 Results

For the following analysis, only data that occurred within a *forcing* situation are considered. After a Shapiro-Wilk normality test, a two-way mixed ANOVA was run with FEEDBACK as a between-subject factor with two levels, and BLOCK as a within-subject factor with five levels. Where necessary, a Greenhouse-Geisser sphericity correction was applied. Post-hoc analysis was run with Bonferroni correction. If a lack of normality could be found, a non-parametric Kruskal-Wallis test was instead conducted on the same factors. Results are reported as statistically significant when $p < 0.05$.

Performance

The error rate was computed for each pair in *forcing* situations.

With a Shapiro-Wilk value of 0.692 ($p < 0.001$), a significant difference from normal distribution was found, preventing the use of ANOVA. Instead, a Kruskal-Wallis test showed that performance was significantly affected by both FEEDBACK ($H(1) = 5.675$, $p < 0.017$) and BLOCK ($H(4) = 21.462$, $p = 0.001$), as seen in Figure 5.7.

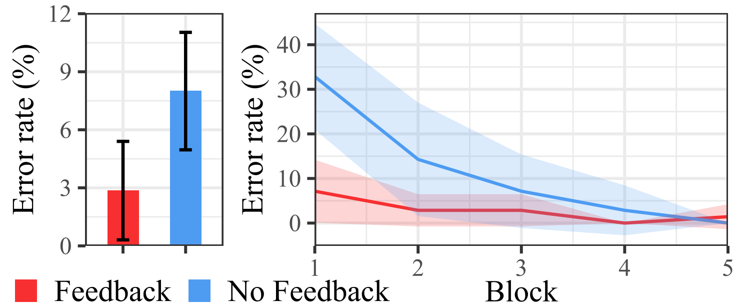

Fig. 5.7 Error rates of both conditions. Error bars and shaded areas are 95% confidence intervals.

Trajectories

Trajectories were computed using the model and values described in Section 5.2.2.

Temporality The Shapiro-Wilk test was not significant for either Starting or Ending Delays ($W=0.981$, $p=0.191$ and $W=0.221$, $p=0.622$ resp.). A main effect of BLOCK can be found both on Starting ($F_{2.6,67.8}=8.50$, $p<0.001$) and Ending Delays ($F_{2.9,75.0}=9.50$, $p<0.001$) (see Figure 5.8). No effect was found for FEEDBACK on Starting ($F_{1,26}=0.26$, $p=0.618$) or Ending Delays ($F_{1,26}=0.05$, $p=0.819$). There was also no interaction between FEEDBACK and BLOCK for either (resp. $F_{2.6,67.8}=0.78$, $p=0.493$ and $F_{2.9,75.0}=0.72$, $p=0.540$).

For Starting Delay, a post-hoc analysis shows a significant difference between the first block and blocks 3, 4, and 5 (resp. $p=0.002$, $p<0.001$, $p<0.001$), and between the second block and blocks 4 and 5 (resp. $p=0.041$ and $p=0.022$).

For Ending Delay, a post-hoc analysis shows a significant difference between the first block and each other block ($p=0.003$ for block 2, and $p<0.001$ for every other).

Interpersonal Distance The normality test for interpersonal distance was passed ($W=0.953$, $p=0.610$). No main effect was found for either FEEDBACK ($F_{1,12}=1.56$, $p<0.236$), BLOCK ($F_{2.7,31.5}=2.48$, $p<0.086$), or their interaction ($F_{2.7,31.5}=1.11$, $p<0.355$).

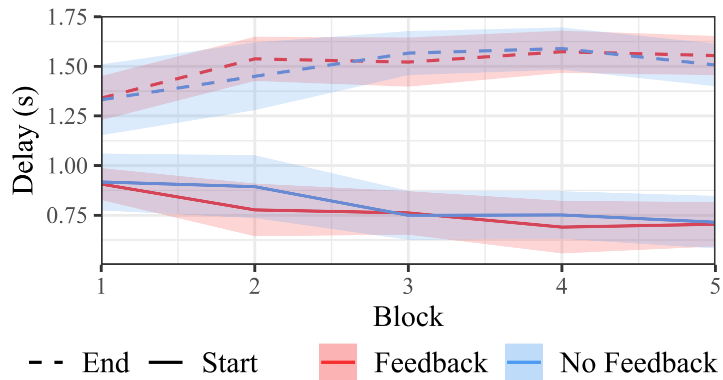

Fig. 5.8 Starting and Ending Delay of both conditions for each block with standard error. Shaded areas are 95% confidence intervals.

Gaze

The time spent looking towards each area during the 10 *forcing* walls each block was gathered, totaling 40 seconds per block. A pair was excluded for this metric as their recording failed on the last three walls, compromising their total time spent during forcing walls.

Focusing on the FORWARD-CENTER and PARTNER areas, a Shapiro-Wilk test showed an absence of normal distribution ($W = 0.884$, $p < 0.001$ and $W = 0.836$, $p < 0.001$ resp.), requiring a Kruskal-Wallis test for both areas on factors FEEDBACK and BLOCK. In the FORWARD-CENTER area, a significant effect was found for FEEDBACK ($H(1) = 35.679$, $p < 0.001$), and none for BLOCK ($H(4) = 0.658$, $p = 0.956$). In the PARTNER area too, a significant effect was found for FEEDBACK ($H(1) = 51.501$, $p < 0.001$), and none for BLOCK ($H(4) = 0.280$, $p = 0.991$).

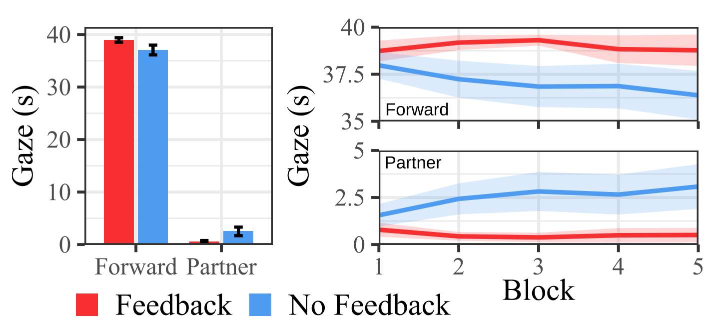

Fig. 5.9 Time spent facing towards the Partner and Forward areas in both conditions for forcing walls. The total duration of these trials is 40 s. Error bars and shaded areas are 95% confidence intervals.

Questionnaire

The mean global score in the FEEDBACK condition is 2.15 (std=0.71), and 1.85 (std=0.57) in the NO FEEDBACK condition (see Figure 5.10). A pairwise t-test comparison shows a statistically significant effect of feedback on the global score (s=3.077, p=0.002). A more detailed analysis of the dimensions in the questionnaire, that can be seen in Figure 5.11, suggests that there are more positive answers and a quasi-total absence of negative answers in the FEEDBACK condition.

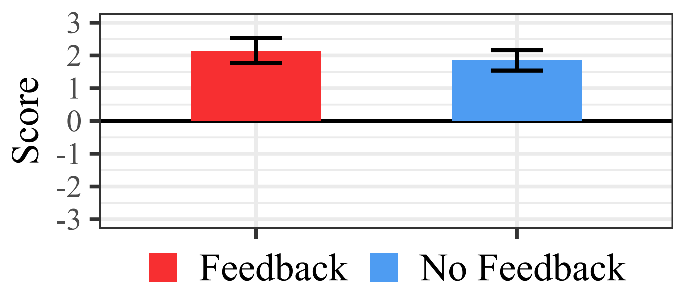

Fig. 5.10 Global Questionnaire Score for both conditions. Error bars are 95% confidence intervals.

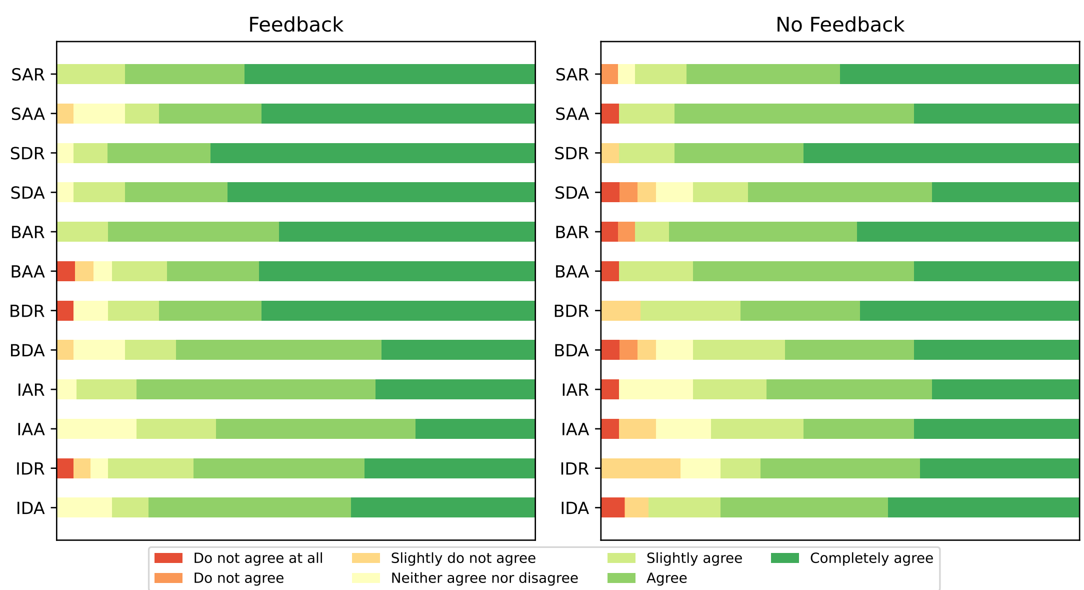

Fig. 5.11 Answer proportion for each question. Each letter represents a value from the three dimensions **S**ituation/**B**ehavior/**I**ntention, **A**wareness/**D**ependence, and **A**ttributed/**R**eported (see Section 5.2.2). For instance, **BAR** refers to **B**ehavior **A**wareness **R**eported.

Correlation between questionnaire and explicit errors

The correlation between the errors rate and the global questionnaire score was computed, and resulted in a moderate negative correlation (Pearson's r=-0.415, p=0.028), *i.e.*, the lower the error rate, the higher the self-reported copresence.

5.4 Discussion

In this section, the effectiveness and the limitations of behavioral measurements for quantifying copresence in CMR are discussed in light of the semi-structured interviews conducted with the participants. Insights are provided on the importance of territories and learning, with reflections on how to implement this protocol in other contexts.

5.4.1 Behavioral measures of Copresence in CMR

The goal of this protocol was to be able to measure copresence in CMR through behavioral analysis. The role of this first study is to validate the protocol by selecting a condition whose effect on copresence is expected, to confirm this effect through a questionnaire, and to evaluate the behavioral measures prepared for it. This first study demonstrates the effectiveness of behavioral measures and the presented approach, through the following evidence.

First, the questionnaire was able to show a difference in the global score between both conditions, with the FEEDBACK condition scoring better than NO FEEDBACK. This means that participants felt an overall stronger mutual awareness and interdependence with their partners when their positional information was made available to them within the virtual environment. This can be understood through statements from participants in the FEEDBACK condition stating "*The marker became more you than you*", or "*After a while, I considered the line to be an extension of my partner's position*". On the other hand, a participant in the NO FEEDBACK condition commented that they needed to "*leave the game to look at the other before entering it again*", confirming that the lack of feedback deteriorated the sense of sharing the environment, while its presence kept the other inside the virtual environment as an elementary unidimensional avatar.

These results suggest that the feeling of copresence was increased with the addition of visual feedback, which was the base hypothesis for testing the protocol. At the same time, the measures show an effect of this visual feedback on the behavioral errors. Error rates were higher in the absence of visual feedback, confirming the hypothesis that a reduced representation of the other users within the virtual environment would impact the overall performance in avoiding walls.

Finally, for both areas of interest in gaze, FORWARD-CENTER and PARTNER, a main effect of FEEDBACK was found, along with an interaction with BLOCK. Participants without feedback spent more time looking toward their partner and less time looking forward. They reported during the interview that looking towards their partner served to check for potential conflict, and that while some were able to guess where the other was, they would still look either for reassurance or entertainment.

This combination of results can be interpreted as follows: given a stronger sense of copresence (induced through added feedback), participants pay more attention to each other's behavior and adapt their trajectories to collectively avoid walls. This is confirmed through a statistically significant correlation between the errors from the behavioral measures and the global score from the shared experience questionnaire, which tends to reinforce the link between the behavior of the participants and their feeling of copresence. It is arguable that the added visual feedback alone might explain the observed effect. However, the fact that participants in the NO FEEDBACK group were able to perfectly complete the task at the end of block 5 leads to the interpretation that differences in performance are more related to thinking or remembering that one has to step aside, due to the reminder of the presence of the other participant. This also aligns with the testimony of the participants.

However, it can be noted that no effect of FEEDBACK was found for either delays or interpersonal distance. This can mean that neither of these measures was adequate for this task, or that the amount of data is too small to find an effect. What might cause the lack of results from these behavioral measures for this task is that it relies heavily on the participants being highly aware of their position and timing. This contrasts with a study finding an effect of co-location on mean and minimum clearance, whose task involved movement without making it such a key part of the goal since it involves the gathering of objects in the environment [83]. This suggests that either the avoidance task or the trajectory metrics should be adapted in order to better fit each other.

5.4.2 Errors and Copresence

This study is also already able to give insight on how the feeling of copresence can fluctuate. During the post-task interviews, participants were asked about the reasons for their errors and close calls. While some evoked fatigue, or a clear disagreement where the other was kept in mind, but their intention was unexpected, many errors stem from a false feeling of relief. A participant commented "*As far as I was concerned, I was through, but I did not think about my partner.*", a reoccurring attitude which the participants sometimes had enough time to correct without triggering a hit. This betrays occasional lapses in copresence, which are revealed through conflict-inducing *forcing* walls. Another said "*I see a free lane in front of me, so I figure I'm in the clear for four walls.*", which reveals a tendency by participants to look ahead and anticipate their movements, sometimes without considering their partner.

This supports a vision of a continuous copresence fluctuating throughout the task similarly to the "Breaks in Presence" (BIP) approach of measuring presence [106]. However in this case, instead of a high sense of presence dropping because of an event, a low sense of copresence can surge up. More specifically, the sense of copresence slowly degrades with fatigue or intense focus on the walls, and jumps back up when the other is brought back into consideration, either because of an error or a physical contact with the other, as has sometimes happened during trials. The appearance of visual feedback on the other person in this case would serve as a near-constant reminder of the other's presence, which limits the degradation of copresence over time.

The design of the study also pointed out the limitations of a simple interview for understanding which part of a complex decision-making process could lead to task errors. Despite the investigator noting down errors as they come up for discussing at the end of the trials, participants were often unsure of which specific errors were referenced. Although it prolongs the protocol, a solution could be to provide a replay so that participants can provide commentary on the evolution of their perception, decisions, and actions as the trial unfolds. This approach would both provide additional opportunities for participants to exchange and share their respective experience during the task, and enrich the understanding of how the task is approached, attempted, and sometimes failed.

5.4.3 Learning effect

The results show a strong learning effect across blocks, which needs to be taken into account for future development of the protocol. A learning effect was detected through BLOCK for errors, starting and ending delays. Over time, participants committed fewer errors, started and stopped moving earlier. This is especially visible for NO FEEDBACK groups, which start at a higher error rate than FEEDBACK ones, as can be seen in Figure 5.7. This effect could be because of the participants learning the task and developing a strategy, or because they build a stronger sense of the other over time allowing them to move without hesitation.

The creation of a strategy is commonly reported by participants during the interview. Some explain that they identified that the external lane to their side would always be theirs, and they would always stay there when it was open, hinting at the emergence of territories [99]. Others solved the puzzle by noticing that with only two available lanes and no right of crossing each other, there was only one option left, which was to always take the leftmost lane if they were on the left. Finally, a few improved this approach by considering time. They identified that there were cases where they had to move quickly to allow room for their neighbor as soon as possible, and cases where they should move slowly in order for their partner to have time to step aside.

It also seems like the sense of sharing an environment can be learned. A participant noted that "*At some point you realize that you play as two, then you do not need to think*". Another noticed that their understanding of their partner's behavior was stable, but that they could guess their intention as they learned the game.

This calls for an iteration of the protocol with too much complexity for a strategy to be defined during the task, for instance by allowing the participants to cross each other, increasing the number of lanes, or restricting them to certain participants. If the task can not be made complex enough to not be learned, behavioral measures may only be representative during its discovery. The observation of a cross-condition learning effect afforded by the perspective blocks also substantiates the decision of a between-subject design for the study. Future applications of the task should also consider this potential learning effect in their analysis.

Another important aspect involved in the learning process is the gaze. A statistically significant interaction between FEEDBACK and BLOCK for gaze and a post-hoc analysis reveal a form of learning for where participants in the NO FEEDBACK condition look during the task. The sharp increase in PARTNER gazing matching the sharp decrease in error rate between the first and second blocks in the NO FEEDBACK condition hints towards a correlation between the two. However, some participants reported during the interview that they did not understand before the second block that they could look directly at their partner, which can partially explain the difference in gazing. Other participants also reported a lack of understanding of the rules and did not realize that they should not stand in the same lane until their first error. These two possibilities were less likely to impact participants in the FEEDBACK condition, as the position of their partner was available in the virtual environment without having to turn their head. The feedback of both participants turning red when in the same lane could also have helped some understand the rules before committing an error.

5.4.4 Other contexts and copresence factors

The proposed behavioral measure can also be transposed to evaluate copresence in application contexts which do not rely on full body movements. In order to do so, the key components that need to be preserved are: 1) spatial territories that participants can appropriate and share; 2) a collaborative task that may provoke conflicts within these territories; 3) sufficient independence between participants within that task so that their levels of copresence can vary.

For example, in a context of mid-air manipulations, one could envision a line of virtual boxes between two participants facing each other and in which they have to reach from above to touch one target each as fast as possible. Moving the two targets between boxes can then be used to test if participants will leave the closest target to one another to ensure success. In the case of a low level of copresence, as in the wall dodging task, participants can be expected to forget about the other and fail to leave room for them.

Following the same principle, other contexts could also be envisioned, covering selection, manipulation and navigation within virtual content, and with more than two users.

Within this context or others, more potential factors that could be studied include environment asymmetry, inspired by actual events encountered during the experiment. A participant was observed stepping forward, out of the task area by 4 meters, without either them or their partner noticing because they were both too focused on the task. Leveraging the focus required by the avoidance task with the spacial asymmetry would thus reveal its potential effect on copresence in CMR. By exploiting the continuous nature of behavioral measures, it is possible to progressively or abruptly displace the virtual environments in relation to the physical area. By controlling the triggering of potential Breaks in Co-presence (BIC), as an equivalent of Breaks In Presence (BIP) [106] for shared environments, the behavior pre- and post-BIC could be compared, which would give a measure for the impact of the displacement causing the BIC. It would thus be possible to investigate the tolerance to spatial misalignment in shared CMR environments.

A different kind of asymmetry relying on the same ideas of BIC is one where participants punctually encounter different walls. This is based on the observation of a participant seeing their partner not taking any action at all to avoid a wall in their lane and immediately thinking that the system encountered an issue. The participants noticing the differences by themselves, as well as changes in their behavior could serve as measures of their awareness of their neighbor and understanding of their behavior for other factors.

5.4.5 Limitations

This protocol and its evaluation suffer from limitations which should be addressed in future versions. First the FEEDBACK and NO FEEDBACK conditions are still entangled with the measures of task performance (which could be improved by the visual feedback alone) and gaze (influenced by the visual feedback). The protocol therefore requires additional interviews and questionnaires to remove any doubt on what is exactly measured. This issue could be addressed by increasing the complexity of the environment, *e.g.* with more lanes and options, in order to enable a greater diversity of solutions, therefore reducing the learning effect and the influence of visual feedback.

Second, trajectories did not provide strong evidence of an effect of the two conditions. Other designs for the obstacles and movements, together with alternative methods for analyzing these trajectories, might help reinforce their usefulness in the estimation of the feeling of a shared mixed-reality experience.

Third, it is recognizable that the number of data points may seem limited. It is due to the attempt of mitigating the learning effect in the experiment by using a between-groups design. Here too, a more complex task with multiple solutions might reduce the learning effect and allow for within-subject design with more participants.

Finally, the role of the passthrough quality on the feeling of copresence can be argued. The impact of photographic realism on copresence is unclear [78], unlike behavioral realism, which would be similar for low or high quality camera passthrough or visual see-through. Still, this asks for further investigation.

5.5 Conclusion

In this chapter, a novel experimental protocol was proposed to evaluate the feeling of copresence in a CMR experience through both objective, *i.e.* behavioral measures, and subjective, *i.e.* questionnaires, aspects. An experimental task was designed in which users collectively interact with a virtual environment in a way that highlights their perception of others in the physical and virtual environment through positional visual feedback.

The results suggest that this protocol provides a complementary way of measuring the experience of users in CMR by confirming the effect of feedback via a questionnaire and measuring effects on performance and gaze.

This protocol is meant to be easily iterable and expandable in order to accommodate the evaluation of varied factors. As discussed above, it can also be transposed to other application contexts. Further research with this protocol should to strengthen its validity with new factors such as interface asymmetry.

Chapter 6

Conclusion

This thesis sought out to investigate how MR could support collaboration and performance in a co-located context. The work was motivated by the challenges and opportunities witnessed during my involvement in the creation of the MRP *Animate*, which served as a practical and conceptual foundation for this research. Through three complementary axes of inquiry, this thesis examined how creators devise strategies to overcome challenges in MRPs, how decoupling physical and virtual spaces can expand collaborative interaction, and how behavioral measures can be applied for measuring the feeling of copresence.

The reflexive thematic analysis of existing MRPs revealed a wide range of strategies employed by creators to navigate technological constraints while respecting artistic goals. These strategies, namely AVOID, DISGUISE, TOLERATE, INTEGRATE, and LEVERAGE, highlight the negotiated nature of performance design, and its constant balancing act between technical and artistic considerations. The design and evaluation of gRAinyCloud demonstrated how decoupling physical and virtual environments can open new opportunities for collaborative interaction in CMR. Freeing performers from the constraint of a strict spatial alignment, created a potential for new forms of interaction and coordination, which illustrates how interaction design directly impacts artistic practice. The exploration of copresence evaluation revealed the potential for behavioral data to provide insight unaccessible through questionnaires. By proposing a protocol that combines a questionnaire, behavioral measures, and interviews with a task that enables spatial conflict, this work contributes methodological tools for a more nuanced assessment of copresence in CMR.

A central feature of this thesis is its methodological diversity, which reflects the hybrid nature of MR as both a technical and artistic field. The overall methodology integrates performance-led research, reflexive thematic analysis, interviews with experts and participants, self-study, and controlled experiment. Each method was used to address a different layer of understanding. Firsthand creative involvement revealed challenges otherwise invisible to purely observational studies, building foundations for the following steps of interviews and wider reflexive thematic analysis. Expert interviews allowed a deep and educated insight on the design of MRPs while opening the scope of production contexts. The reflexive thematic analysis of more works within the literature allowed to further widen this scope on a different scale. A practice-based self-study permitted a deep dive into the subtle consequences a difference in shared interface design can have on collaborative musical interaction. Finally, completing a controlled experiment grounded the contribution of a novel protocol and made it repeatable and adaptable. This diversity did not come without challenges as it demanded a diligently reflexive decision process, while careful design choices balancing artistic practice and academic rigor were a recurring concern. However, this pluralistic approach illustrates the value of methodological flexibility in MR research and demonstrates how artistic creation can act as a legitimate context for research.

In doing so, this work connects design research in MR, the study of collaboration and copresence in Human-Computer Interaction, and practice-based approaches from the performing arts. Few studies have systematically documented how MRPs are conceived from within the creative process, which is a symptom of journal and conference formats having a limited amount of words or pages per submission. When reporting on a production, what is usually seen as a valuable contribution is the final result shown to the audience, and not the journey consisting of the trials and errors, which are at least as valuable to other designers. Similarly, positioning the thesis within a co-located context singles it out from the current trend around telepresence, which employs similar notions but can encounter very different challenges, like physical conflict when co-located or avatar representation when distant. This shift is understandable through the progress of tracking and display technologies, as well as the covid crisis, but co-location remains an important context to study with unique challenges and applications. Despite the similarity in notions, like presence and copresence, the difference in context can drastically change the effects of some variables, and how to measure them, justifying a first step towards behavioral measures of copresence in CMR in order to open the door to further studies of copresence.

While this work makes several contributions, it also opens opportunities for improvement and further research. The thematic analysis focused on a relatively small sample due to the limited availability of detailed and critical documentation of MRPs, but could have its scale widened through more interviews with designers in order to access the information that did not fit in their papers. Informing creators of the reported strategies and documenting the unfolding design process would also be highly valuable as a mean to validate and build upon the framework, as well as to provide an additional detailed case study on which future designers can draw from. *gRAinyCloud* provided valuable initial insights, but could gain from further experiments, either with a similar design in a longer-term study, or with alternative questions, investigating interactions within larger groups of collaborating participants, or how to represent the performance to an unimmersed audience. Similarly, the behavioral measures of copresence stand to be improved upon, with further iterations of the presented task for validation, or adaptations of it while retaining the basis of a collaborative yet independent task involving appropriable spatial territories.

This research marks a step toward a better understanding of the perception and behaviors of people sharing virtual spaces. In the future, I would like to see this work contribute to explaining how individuals notice, interpret, and act upon differences of various natures and scales between their perception and the one of their peers. In the same vein, I believe that such an understanding could enable a more deliberate leveraging of differences in perspectives and experiences, with significant potential not just for artistic productions like *Animate*, but also for educational contexts as in museums, or even for industrial applications.

The essence of this thesis is multidisciplinary, letting art inform science and science inform art. Ultimately, it does not claim to offer definitive solutions, but rather to contribute perspectives and tools that may help others navigate the challenges of CMR experiences. By foregrounding strategies used by designers, by showcasing the potential of spatial decoupling, and by providing an alternative method for measuring copresence, it offers a set of resources that can inform both creative practice and research on collaborative interaction more broadly. My hope is that these contributions will support designers, performers, and researchers alike in developing MR experiences that are not only technically sound, but also socially and creatively meaningful.

References

[1] C. Arslan and F. Berthaut. Examining the Design Process for 3D Interactions in Performing Arts : A Spatial Augmented Reality Cyber-Opera Case Study. *IEEE Trans. Vis. Comput. Graph.*, 31(5):1–10, Mar. 2025.

[2] N. Ashtari, A. Bunt, J. McGrenere, M. Nebeling, and P. K. Chilana. Creating Augmented and Virtual Reality Applications: Current Practices, Challenges, and Opportunities. In *Proc. CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–13, Honolulu, HI, USA, 2020. ACM.

[3] J. Bailenson, R. Guadagno, E. Aharoni, A. Dimov, A. Beall, and J. Blascovich. Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. *Proc. 7th Annual International Workshop on PRESENCE*, Jan. 2004.

[4] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 10(6):583–598, Dec. 2001.

[5] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Interpersonal Distance in Immersive Virtual Environments. *Personality and Social Psychology Bulletin*, 29(7):819–833, July 2003.

[6] J. N. Bailenson, K. Swinth, C. Hoyt, S. Persky, A. Dimov, and J. Blascovich. The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 14(4):379–393, Aug. 2005.

[7] C. Basdogan, C.-H. Ho, M. A. Srinivasan, and M. Slater. An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction*, 7(4):443–460, Dec. 2000.

[8] R. A. J. d. Belen, H. Nguyen, D. Filonik, D. D. Favero, T. Bednarz, R. A. J. d. Belen, H. Nguyen, D. Filonik, D. D. Favero, and T. Bednarz. A systematic review of the current state of collaborative mixed reality technologies: 2013–2018. *AIMS Electronics and Electrical Engineering*, 3(2):181–223, 2019.

[9] J. Bell. Networked Music Performance in PatchXR and FluCoMa. In *Proc. International Computer Music Conference*, ICMC '23, Shenzhen, China, Oct. 2023. CUHK-Shenzhen.

[10] S. Benford, A. Crabtree, M. Flintham, A. Drozd, R. Anastasi, M. Paxton, N. Tandavanitj, M. Adams, and J. Row-Farr. Can you see me now? *ACM Transactions on Computer-Human Interaction*, 13(1):100–133, Mar. 2006.

[11] S. Benford and G. Giannachi. *Performing Mixed Reality*. The MIT Press, 2011.

[12] S. Benford and G. Giannachi. Interaction as performance. *Interactions*, 19(3):38–43, May 2012.

[13] S. Benford, G. Giannachi, B. Koleva, and T. Rodden. From interaction to trajectories: designing coherent journeys through user experiences. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 709–718, New York, NY, USA, 2009.

[14] F. Berthaut. 3D interaction techniques for musical expression. *Journal of New Music Research*, 49(1):60–72, Jan. 2020.

[15] F. Berthaut and L. Dahl. The Effect of Visualisation Level and Situational Visibility in Co-located Digital Musical Ensembles. In *Proc. International Conference on New Interfaces for Musical Expression*, NIME '22, June 2022.

[16] F. Berthaut, D. M. Plasencia, M. Hachet, and S. Subramanian. Reflets: Combining and Revealing Spaces for Musical Performances. In *Proc. International Conference on New Interfaces for Musical Expression*, NIME '15, Baton Rouge, LA, USA, 2015.

[17] F. Biocca. The Cyborg 's Dilemma: Progressive Embodiment in Virtual Environments [1]. *Journal of Computer-Mediated Communication*, 3(2):JCMC324, Sept. 1997.

[18] F. Biocca and C. Harms. Defining and measuring social presence: Contribution to the networked minds theory and measure. *Proc. Fifth Annual International Workshop on Presence*, Jan. 2002.

[19] F. Biocca, C. Harms, and J. K. Burgoon. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments*, 12(5):456–480, Oct. 2003.

[20] F. Biocca, C. Harms, and J. L. Gregg. The Networked Minds Measure of Social Presence : Pilot Test of the Factor Structure and Concurrent Validity. In *Proc. Presence*, 2001.

[21] A. Boem, M. Tomasetti, and L. Turchet. Harmonizing the Musical Metaverse: unveiling needs, tools, and challenges from experts' point of view. In *Proc. International Conference on New Interfaces for Musical Expression*, NIME '24, May 2024.

[22] M. Botvinick and J. Cohen. Rubber hands 'feel' touch that eyes see. *Nature*, 391(6669):756–756, Feb. 1998.

[23] S. Bouchard, G. Robillard, J. St-Jacques, S. Dumoulin, M.-J. Patry, and P. Renaud. Reliability and validity of a single-item measure of presence in VR. In *Proc. 3rd International Workshop on Haptic, Audio and Visual Environments and Their Applications*, HAVE '04, pages 59–61. IEEE, 2004.

[24] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, Jan. 2006.

[25] V. Braun and V. Clarke. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4):589–597, Aug. 2019.

[26] W. Broll. Interacting in distributed collaborative virtual environments. In *Proc. Virtual Reality Annual International Symposium*, VRAIS '95, pages 148–155, Mar. 1995.

[27] B. Buxton. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann, July 2010.

[28] D. Byrne. A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity*, 56(3):1391–1412, June 2022.

[29] O. Capra, F. Berthaut, and L. Grisoni. Have a SEAT on Stage: Restoring Trust with Spectator Experience Augmentation Techniques. In *Proc. Designing Interactive Systems Conference*, DIS '20, pages 695–707, Eindhoven, Netherlands, July 2020. ACM.

[30] J. Casanueva and E. Blake. The Effects of Group Collaboration on Presence in a Collaborative Virtual Environment. In *Virtual Environments 2000*, pages 85–94. Springer Vienna, Vienna, 2000.

[31] A. Cassinelli and M. Ishikawa. Volume slicing display. In *Proc. SIGGRAPH ASIA Art Gallery & Emerging Technologies: Adaptation*, SIGGRAPH ASIA '09, page 88, New York, NY, USA, 2009. ACM.

[32] M. Chalmers and A. Galani. Seamful interweaving: heterogeneity in the theory and design of interactive systems. In *Proc. Designing Interactive Systems Conference*, DIS '04, pages 243–252, Cambridge, MA, USA, Aug. 2004. ACM.

[33] W. Chen. *Collaboration in Multi-user Immersive Virtual Environment*. phdthesis, Université Paris Saclay (COmUE), Dec. 2015.

[34] W. Chen, C. Clavel, N. Férey, and P. Bourdot. Perceptual Conflicts in a Multi-Stereoscopic Immersive Virtual Environment: Case Study on Face-to-Face Interaction through an Avatar. *Presence: Teleoperators and Virtual Environments*, 23(4):410–429, Nov. 2014.

[35] A. Clay, G. Domenger, J. Conan, A. Domenger, and N. Couture. Integrating augmented reality to enhance expression, interaction & collaboration in live performances: A ballet dance case study. In *Proc. International Symposium on Mixed and Augmented Reality - Media, Art, Social Science, Humanities and Design*, IMSAR-MASH '14, pages 21–29, Germany, Sept. 2014. IEEE.

[36] T. Combe, R. Fribourg, L. Detto, and J.-M. Normand. Exploring the Influence of Virtual Avatar Heads in Mixed Reality on Social Presence, Performance and User Experience in Collaborative Tasks. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2206–2216, May 2024.

[37] N. Coomer, S. Bullard, W. Clinton, and B. Williams-Sanders. Evaluating the effects of four VR locomotion methods: joystick, arm-cycling, point-tugging, and teleporting. In *Proc. Symposium on Applied Perception*, SAP '18, pages 1–8, New York, NY, USA, 2018. ACM.

[38] D. J. Cox. Collaborations in art/science: Renaissance teams. *The Journal of biocommunication*, 18(2):15–24, 1991.

[39] E. A. J. Croes, M. L. Antheunis, A. P. Schouten, and E. J. Krahmer. Teasing apart the effect of visibility and physical co-presence to examine the effect of CMC on interpersonal attraction. *Computers in Human Behavior*, 55:468–476, Feb. 2016.

[40] Y. de Kort, W. IJsselsteijn, and K. Poels. Digital games as social presence technology : development of the social presence in gaming questionnaire (SPGQ). *Proc. 10th Annual International Workshop on Presence*, pages 195–203, Oct. 2007.

[41] C. DeFanti. *Co-Located Augmented and Virtual Reality Systems*. Ph.D. Dissertation, New York Univ, New York, NY, USA, 2019.

[42] T. Drey, M. Rietzler, and E. Rukzio. Questionnaires and Qualitative Feedback Methods to Measure User Experience in Mixed Reality, Apr. 2021.

[43] S. Fels and F. Vogt. Tooka: Explorations of Two Person Instruments. In *Proc. International Conference on New Interfaces for Musical Expression*, NIME '02, page 6, Dublin, Ireland, 2002.

[44] W. M. Felton and R. E. Jackson. Presence: A Review. *International Journal of Human–Computer Interaction*, 38(1):1–18, Jan. 2022.

[45] J. Freeman, S. E. Avons, R. Meddis, D. E. Pearson, and W. IJsselsteijn. Using Behavioral Realism to Estimate Presence: A Study of the Utility of Postural Responses to Motion Stimuli. *Presence: Teleoperators and Virtual Environments*, 9(2):149–164, Apr. 2000.

[46] J. Freeman, S. E. Avons, D. E. Pearson, and W. A. IJsselsteijn. Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence*, 8(1):1–13, 1999.

[47] D. Friedman, A. Brogni, C. Guger, A. Antley, A. Steed, and M. Slater. Sharing and Analyzing Data from Presence Experiments. *Presence: Teleoperators and Virtual Environments*, 15(5):599–610, Oct. 2006.

[48] M. Gandy, R. Catrambone, B. MacIntyre, C. Alvarez, E. Eiriksdottir, M. Hilimire, B. Davidson, and A. C. McLaughlin. Experiences with an AR evaluation test bed: Presence, performance, and physiological measurement. In *International Symposium on Mixed and Augmented Reality*, ISMAR '10, pages 127–136, Oct. 2010.

[49] D. Gochfeld, C. Brenner, K. Layng, S. Herscher, C. DeFanti, M. Olko, D. Shinn, S. Riggs, C. Fernández-Vara, and K. Perlin. *Holojam* in Wonderland : Immersive Mixed Reality Theater. *Leonardo*, 51(4):362–367, Aug. 2018.

[50] E. Goffman. *The presentation of self in everyday life*. The presentation of self in everyday life. Doubleday, 1956.

[51] J. G. Grandi, H. G. Debarba, and A. Maciel. Characterizing Asymmetric Collaborative Interactions in Virtual and Augmented Realities. In *Proc. Conference on Virtual Reality and 3D User Interfaces*, VR '19, pages 127–135, Osaka, Japan, Mar. 2019. IEEE.

[52] S. Grassini and K. Laumann. Questionnaire Measures and Physiological Correlates of Presence: A Systematic Review. *Frontiers in Psychology*, 11, 2020.

[53] F. B. Guefrech, F. Berthaut, P. Plénacoste, Y. Peter, and L. Grisoni. Revealable Volume Displays: 3D Exploration of Mixed-Reality Public Exhibitions. In *Proc. Virtual Reality and 3D User Interfaces*, VR '21, pages 31–39, Mar. 2021.

[54] R. Hamilton. Trois Machins de la Grâce Aimante: a Virtual Reality String Quartet. In *Proc. International Computer Music Conference*, ICMC '19, Aug. 2019.

[55] A. Henson. *We're In This Together: Embodied Interaction, Affect, and Design Methods in Asymmetric, Co-Located, Co-Present Mixed Reality*. thesis, Carnegie Mellon University, June 2019.

[56] S. Herscher. *Scaling Multi-user Virtual and Augmented Reality*. Ph.D. Dissertation, Dept. Comp. Sci., New York Univ., New York, NY, USA, 2020.

[57] S. Herscher, C. DeFanti, N. G. Vitovitch, C. Brenner, H. Xia, K. Layng, and K. Perlin. CAVRN: An Exploration and Evaluation of a Collective Audience Virtual Reality Nexus Experience. In *Proc. 32nd Annual Symposium on User Interface Software and Technology*, UIST '19, pages 1137–1150, New Orleans, LA, USA, Oct. 2019. ACM.

[58] J. Jerald. *The VR Book: Human-Centered Design for Virtual Reality*. Morgan & Claypool, Sept. 2015.

[59] B. Kates. Incandescent Edges of the Future: Performance Creation with Virtual, Augmented and Carbon Realities. Master's thesis, Drama, Calgary Univ., Calgary, AB, Canada, Sept. 2020.

[60] K. Kim, R. Schubert, J. Hochreiter, G. Bruder, and G. Welch. Blowing in the wind: Increasing social presence with a virtual human via environmental airflow interaction in mixed reality. *Computers & Graphics*, 83:23–32, Oct. 2019.

[61] V. Krauß, A. Boden, L. Oppermann, and R. Reiners. Current Practices, Challenges, and Design Implications for Collaborative AR/VR Application Development. In *Proc. CHI Conference on Human Factors in Computing Systems*, CHI '21, pages 1–15, Yokohama Japan, May 2021. ACM.

[62] J. Kreimeier, S. Hammer, D. Friedmann, P. Karg, C. Bühner, L. Bankel, and T. Götzelmann. Evaluation of different types of haptic feedback influencing the task-based presence and performance in virtual reality. In *Proc. 12th International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '19, pages 289–298, New York, NY, USA, 2019. ACM.

[63] J. Lacoche, N. Pallamin, T. Boggini, and J. Royan. Collaborators awareness for user cohabitation in co-located collaborative virtual environments. In *Proc. 23rd Symposium on Virtual Reality Software and Technology*, VRST '17, pages 1–9, Gothenburg Sweden, Nov. 2017. ACM.

[64] K. Layng, K. Perlin, S. Herscher, C. Brenner, and T. Meduri. *CAVE* : Making Collective Virtual Narrative: *Best Paper Award*. *Leonardo*, 52(4):349–356, Aug. 2019.

[65] M. Le Chénéchal, J. Lacoche, J. Royan, T. Duval, V. Gouranton, and B. Arnaldi. When the giant meets the ant an asymmetric approach for collaborative and concurrent object manipulation in a multi-scale environment. In *Proc. 3rd VR International Workshop on Collaborative Virtual Environments*, 3DCVE '16, pages 18–22. IEEE, 2016.

[66] M. Lombard and M. T. Jones. Defining Presence. In M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and R. J. Schaevitz, editors, *Immersed in Media: Telepresence Theory, Measurement & Technology*, pages 13–34. Springer International Publishing, Cham, 2015.

[67] E. Malbos, R. M. Rapee, and M. Kavakli. Behavioral Presence Test in Threatening Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 21(3):268–280, Aug. 2012.

[68] D. Margery, B. Arnaldi, and N. Plouzeau. A General Framework for Cooperative Manipulation in Virtual Environments. In M. Gervautz, D. Schmalstieg, and A. Hildebrand, editors, *Virtual Environments '99*, Eurographics, pages 169–178, Vienna, 1999.

[69] M. R. Marner, S. Haren, M. Gardiner, and B. H. Thomas. Exploring interactivity and augmented reality in theater: A case study of Half Real. In *Proc. International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*, ISMAR-AMH '12, pages 81–86, Atlanta, GA, USA, Nov. 2012. IEEE.

[70] L. Men and N. Bryan-Kinns. LeMo: Exploring Virtual Space for Collaborative Creativity. In *Proc. Conference on Creativity and Cognition*, C&C '19, pages 71–82, San Diego CA USA, June 2019. ACM.

[71] D. J. Meyer, M. Wiertlewski, M. A. Peshkin, and J. E. Colgate. Dynamics of ultrasonic and electrostatic friction modulation for rendering texture on haptic surfaces. In *Proc. Haptics Symposium*, HAPTICS '14, pages 63–67, Feb. 2014.

[72] P. Milgram and F. Kishino. A taxonomy of mixed reality visual display. *IEICE Trans. Inf. & Sys., D*, 77(12):1321–1329, 1994.

[73] M. R. Miller and J. N. Bailenson. Social Presence Outside the Augmented Reality Field of View. *Frontiers in Virtual Reality*, 2:656473, Sept. 2021.

[74] K. Nakagaki, D. Fitzgerald, Z. J. Ma, L. Vink, D. Levine, and H. Ishii. inFORCE: Bi-directional 'Force ' Shape Display for Haptic Interaction. In *Proc. Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '19, pages 615–623, Tempe Arizona USA, Mar. 2019. ACM.

[75] T. T. H. Nguyen, T. Duval, and C. Fleury. Guiding techniques for collaborative exploration in multi-scale shared virtual environments. In *Proc. International Conference on Computer Graphics Theory and Applications*, GRAPP '13, pages 327–336, 2013.

[76] S. Nilsson, B. Johansson, and A. Jonsson. Using AR to support cross-organisational collaboration in dynamic tasks. In *Proc. 8th International Symposium on Mixed and Augmented Reality*, ISMAR '09, pages 3–12, Orlando, FL, USA, Oct. 2009. IEEE.

[77] K. L. Nowak and F. Biocca. The Effect of the Agency and Anthropomorphism on Users ' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 12(5):481–494, Oct. 2003.

[78] C. S. Oh, J. N. Bailenson, and G. F. Welch. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI*, 5, 2018.

[79] C. Onime, J. Uhomoibhi, and H. Wang. Mixed Reality Cubicles and Cave Automatic Virtual Environment. In *Proc. 15th International Conference on Ubiquitous Computing and Communications and International Symposium on Cyberspace and Security*, IUCC-CSS '16, pages 1–8, Granada, Spain, Dec. 2016. IEEE.

[80] S. Parameswaran and R. Kishore. A Social Presence Model of Task Performance: A Meta-Analytic Structural Equation Model. In *Proc. Americas Conference on Information Systems*, AMCIS '17, 2017.

[81] S. Park. Collaborative Mobile Instruments in a Shared AR Space: a Case of ARLooper. In *Proc. International Conference on New Interfaces for Musical Expression*, NIME '20, pages 190–195, Birmingham, UK, June 2020. Birmingham City University.

[82] L. Phillips, V. Interrante, M. Kaeding, B. Ries, and L. Anderson. Correlations Between Physiological Response, Gait, Personality, and Presence in Immersive Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 21(2):119–141, Apr. 2012.

[83] I. Podkosova and H. Kaufmann. Co-presence and proxemics in shared walkable virtual environments with mixed colocation. In *Proc. 24th Symposium on Virtual Reality Software and Technology*, VRST '18, pages 1–11, Tokyo Japan, Nov. 2018. ACM.

[84] T. Porcino, S. A. Ghaeinian, J. Franz, J. Malloch, and D. Reilly. Design of an Extended Reality Collaboration Architecture for Mixed Immersive and Multi-surface Interaction. In *Proc. International Conference on Entertainment Computing*, ICEC '22, pages 112–122, 2022.

[85] S. Reeves, S. Benford, C. O'Malley, and M. Fraser. Designing the spectator experience. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 741–750, New York, NY, USA, 2005. ACM.

[86] H. Regenbrecht, C. Botella, R. M. Baños, and T. Schubert. Mixed Reality Experience Questionnaire (MREQ). Technical report, University of Otago, 2017.

[87] H. Regenbrecht and T. Schubert. Measuring Presence in Augmented Reality Environments: Design and a First Test of a Questionnaire. In *Proc. Fifth Annual International Workshop Presence*, pages 138–144, Porto, Portugal, 2002.

[88] G. Richard, T. Pietrzak, F. Argelaguet, A. Lécuyer, and G. Casiez. Within or between? comparing experimental designs for virtual embodiment studies. In *Proc. Conference on Virtual Reality and 3D User Interfaces*, VR '22, pages 186–195. IEEE, Mar. 2022.

[89] G. Richard, T. Pietrzak, F. Argelaguet, A. Lécuyer, and G. Casiez. MultiVibes: What if your VR Controller had 10 Times more Vibrotactile Actuators? In *Proc. International Symposium on Mixed and Augmented Reality*, ISMAR '23, pages 703–712, Sydney, Australia, Oct. 2023. IEEE.

[90] J. P. Rolland and H. Fuchs. Optical versus Video See-Through Head-Mounted Displays. In *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press, 2001.

[91] A. Rostami and D. McMillan. The Normal Natural Troubles of Virtual Reality in Mixed-Reality Performances. In *Proc. CHI Conference on Human Factors in Computing Systems*, CHI '22, pages 1–22, New Orleans LA USA, Apr. 2022. ACM.

[92] A. Rostami, C. Rossitto, and A. Waern. Frictional Realities: Enabling Immersion in Mixed-Reality Performances. In *Proc. International Conference on Interactive Experiences for TV and Online Video*, TVX '18, pages 15–27, SEOUL Republic of Korea, June 2018. ACM.

[93] C. Salter, T. Thomasson, and P. Uro. Animate: A Theatrical Exploration of Climate Transformation through the Medium of Extended Reality (XR). In *Proc. 28th International Symposium on Electronic Art*, ISEA '24, pages 642–651, June 2024.

[94] C. Sas, G. M. P. O'Hare, and R. Reilly. Presence and task performance: an approach in the light of cognitive style. *Cognition, Technology & Work*, 6(1):53–56, Feb. 2004.

[95] S. B. Schorr and A. M. Okamura. Three-Dimensional Skin Deformation as Force Substitution: Wearable Device Design and Performance During Haptic Exploration of Virtual Environments. *IEEE Transactions on Haptics*, 10(3):418–430, July 2017.

[96] R. Schroeder. *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. Springer Science & Business Media, Dec. 2001.

[97] T. Schubert, F. Friedmann, and H. Regenbrecht. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, 10(3):266–281, June 2001.

[98] V. Schwind, P. Knierim, N. Haas, and N. Henze. Using presence questionnaires in virtual reality. In *Proc. CHI Conference on Human Factors in Computing Systems*, CHI, pages 1–12. ACM, 2019.

[99] S. D. Scott, M. S. T. Carpendale, and K. Inkpen. Territoriality in collaborative tabletop workspaces. In *Proc. Conference on Computer Supported Cooperative Work*, CSCW '04, pages 294–303, New York, NY, USA, Nov. 2004. ACM.

[100] S. Serafin, C. Erkut, J. Kojs, N. C. Nilsson, and R. Nordahl. Virtual Reality Musical Instruments: State of the Art, Design Principles, and Future Directions. *Computer Music Journal*, 40(3):22–40, Sept. 2016.

[101] T. B. Sheridan and others. Musings on telepresence and virtual presence. *Presence Teleoperators Virtual Environ.*, 1(1):120–125, 1992.

[102] B. Shneiderman. Creativity support tools. *Communications of the ACM*, 45(10):116–120, Oct. 2002.

[103] R. Skarbez, F. P. Brooks, Jr., and M. C. Whitton. A Survey of Presence and Related Concepts. *ACM Computing Surveys*, 50(6):1–39, Nov. 2018.

[104] R. Skarbez, M. Smith, and M. C. Whitton. Revisiting Milgram and Kishino 's Reality-Virtuality Continuum. *Frontiers in Virtual Reality*, 2:647997, Mar. 2021.

[105] M. Slater. How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments. *Presence*, 13(4):484–493, Aug. 2004.

[106] M. Slater, D. Banakou, A. Beacco, J. Gallego, F. Macia-Varela, and R. Oliva. A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality. *Frontiers in Virtual Reality*, 3, 2022.

[107] M. Slater, M. Usoh, and Y. Chrysanthou. The influence of dynamic shadows on presence in immersive virtual environments. In *Proc. Eurographics workshops on Virtual environments*, VE '95, pages 8–21, 1995.

[108] M. Slater and S. Wilbur. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, Dec. 1997.

[109] J. Sol Roo and M. Hachet. One Reality: Augmenting How the Physical World is Experienced by combining Multiple Mixed Reality Modalities. In *Proc. 30th Symposium on User Interface Software and Technology*, UIST '17, Quebec City, Canada, Oct. 2017. ACM.

[110] V. Souza, A. Maciel, L. Nedel, and R. Kopper. Measuring Presence in Virtual Environments: A Survey. *ACM Computing Surveys*, 54(8):163:1–163:37, Oct. 2021.

[111] B. Spanlang, T. Fröhlich, V. F. Descalzo, A. Antley, and M. Slater. The Making of a Presence Experiment: Responses to Virtual Fire. In *Proc. Annual International Workshop on Presence*, Presence, 2007.

[112] M. Speicher, B. D. Hall, and M. Nebeling. What is Mixed Reality? In *Proc. CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–15, Glasgow Scotland Uk, May 2019. ACM.

[113] F. Steinicke, G. Bruder, K. Hinrichs, A. Steed, and A. L. Gerlach. Does a Gradual Transition to the Virtual World increase Presence? In *Proc. Virtual Reality Conference*, VR '09, pages 203–210, Lafayette, LA, Mar. 2009. IEEE.

[114] S. Symons. The Perceptron: A Multi-player Entangled Instrument based on Interpretive Mapping and Intra-action. In *Proc. 19th International Audio Mostly Conference: Explorations in Sonic Cultures*, AM '24, pages 385–391, New York, NY, USA, Sept. 2024. ACM.

[115] S. Thie and J. Wijk. A general theory on presence: Experimental evaluation of social virtual presence in a decision making task. In *Proc. Presence*, volume 1, 1998.

[116] P. Uro, F. Berthaut, T. Pietrzak, and M. Wanderley. Decoupling Physical and Virtual Spaces in Co-Located Collaborative Mixed Reality Instruments with gRAinyCloud. In *Proc. International Conference on New Interfaces for Musical Expression*, NIME '25, Canberra, Australia, 2025.

[117] P. Uro, F. Berthaut, T. Pietrzak, M. Wanderley, and L. Grisoni. Strategies for the Reconciliation of Artistic Intent and Technical Constraints in Mixed Reality Performances. In *Proc. Designing Interactive Systems Conference*, DIS '25, Madeira, Portugal, 2025. ACM.

[118] P. Uro, F. Berthaut, T. Pietrzak, and M. M. Wanderley. Behavioral Measures of Copresence in Co-located Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics*, 31(5):2848–2858, May 2025.

[119] M. Usoh, E. Catena, S. Arman, and M. Slater. Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments*, 9(5):497–503, Oct. 2000.

[120] J. Van Baren. Measuring Presence : A Guide to Current Measurement Approaches. *Deliverable of the OmniPres project IST-2001-39237*, 2004.

[121] P. Vorderer, W. Wirth, F. R. Gouveia, F. Biocca, T. Saari, L. Jäncke, S. Böcking, H. Schramm, A. Gysbers, T. Hartmann, C. Klimmt, N. Ravaja, A. Sacau, T. Baumgartner, and P. Jäncke. MEC Spatial Presence Questionnaire (MEC-SPQ): Short Documentation and Instructions for Application. *Report to the European Community, Project Presence: MEC (IST-2001-37661)*, 2004.

[122] I. Wagner, W. Broll, G. Jacucci, K. Kuutii, R. McCall, A. Morrison, D. Schmalstieg, and J.-J. Terrin. On the Role of Presence in Mixed Reality. *Presence: Teleoperators and Virtual Environments*, 18(4):249–276, Aug. 2009.

[123] Y. Wang and C. Martin. Cubing Sound: Designing a NIME for Head-mounted Augmented Reality. In *Proc. International Conference on New Interfaces for Musical Expression*, NIME '22, June 2022.

[124] J. Webster, L. K. Trevino, and L. Ryan. The dimensionality and correlates of flow in human-computer interactions. *Computers in Human Behavior*, 9(4):411–426, Dec. 1993.

[125] C. Wee, K. M. Yap, and W. N. Lim. Haptic Interfaces for Virtual Reality: Challenges and Research Directions. *IEEE Access*, 9:112145–112162, 2021.

[126] J. Weijdom. Performative prototyping in collaborative mixed reality environments: an embodied design method for ideation and development in virtual reality. In *Proc. 16th International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '22, pages 1–13, Daejeon Republic of Korea, Feb. 2022. ACM.

[127] B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, June 1998.

[128] F. Zaman, C. Anslow, A. Chalmers, and T. Rhee. MRMAC: Mixed Reality Multi-user Asymmetric Collaboration. In *Proc. International Symposium on Mixed and Augmented Reality*, ISMAR '23, pages 591–600, Sydney, Australia, Oct. 2023. IEEE.

[129] V. Zappi, F. Berthaut, and D. Mazzanti. From the Lab to the Stage: Practical Considerations on Designing Performances with Immersive Virtual Musical Instruments. In *Sonic Interactions In Virtual Environments*, SIVE '22. Springer VS, 2022.