Universiteit Gent
Faculteit Letteren en Wijsbegeerte
Vakgroep Kunst-, Muziek- en Theaterwetenschappen

Extended musical spaces

Participatory and creative music making in digitally augmented musical environments

---

ir. Bavo Van Kerrebroeck

Universiteit Gent

Faculteit Letteren en Wijsbegeerte

Vakgroep Kunst-, Muziek- en Theaterwetenschappen

Promotor: Prof. Dr. Pieter-Jan Maes

Doctoraatsbegeleidingscommissie: Dr. Frederic Bevilacqua
Prof. Dr. Luc Nijs

Examencommissie: Dr. Frederic Bevilacqua
Dr. Ine Vanoeveren
Prof. Dr. Helena Daffern
Prof. Dr. Luc Nijs

Prof. Dr. Maude Bass-Krueger (rapporteur)
Prof. Dr. Maximiliaan Martens (voorzitter)

Universiteit Gent
Faculteit Letteren en Wijsbegeerte
Vakgroep Kunst-, Muziek- en Theaterwetenschappen
IPEM – Instituut voor Psychoacoustica en Electronische Muziek
De Krook, Miriam Makebaplein 1, B-9000 Gent, België

Proefschrift voorgelegd tot het behalen van
de graad van Doctor in de Kunstwetenschappen

Academiejaar 2022-2023

Acknowledgements

For a dissertation devoted to the “extended musical space”, I cannot think of a better place and group of people than the one I’ve been lucky to find at IPEM. These past four years have been exciting, enriching, and opened many doors into many fascinating worlds.

I have to start by giving thanks to my promotor. Pieter-Jan, it’s thanks to you I’ve got the chance to start this journey in the first place and thanks to you I’ve managed to end up here. The fact that we have virtual spheres, breathing sonifications, slime visualizations, spatial music compositions, dynamic agents, and other imaginations laying remnant in early prototypes, is for me a sign of our shared curiosity, and makes me look forward to any future encounters in the unexplored. Thanks for your guidance, inspiration, always welcoming attitude, critical insights, as well as spontaneous eruptions in the form of Krowk regattas, cycling trips, and immersive sessions.

Next to my promotor, I would like to thank Dr. Frederic Bevilacqua, Prof. Dr. Helena Daffern, Prof. Dr. Luc Nijs, and Dr. Ine Vanoeveren for their valuable feedback in evaluating this dissertation. It was a privilege to receive your rich and complementary feedback and I hope we can continue our discussions about (XR) music making in the future. I also would like to thank Ola and Ioana for proofreading this manuscript. Any remaining mistakes are my own.

A big thanks to all the IPEM colleagues, collaborators, friends, and visiting researchers. Thank you Giusy, Kelsey, Ola, Adriaan, Stephane, Kristel, Glenn, Niels, Mattia, Canan, Bart, and Marc, for trusting on my XR expertise to assist you in your works and granting me insights into your own endeavours. A special thanks to Kelsey and Ola for the many meetings and relentless enthusiasm in organising SysMus22 together. Thank you, Marc, Edith, Ioulia, Adriana, Fleur, Valerio, Leon, Lousin, Jorg, Aagje, Joren, and Ivan for making these past four years an unforgettable journey. A special thanks to Joren for the back-to-back exchanges of geeky curiosities, to Bart for always having a solution or a glass of wine ready, and to Ivan for complementing my virtual work with tangible creations as well as for the introduction to perhaps one of my biggest loves at IPEM, the Synthi EMS 100. Thank you Aagje for making the whole of IPEM function and to Ann for brightening the rest of De Krook. Thank you, Andrea, Celien, Lennert, and Ella for your inspirational internships and all those hours spent together in the ASIL. A thanks to all participants for lending their time, minds, and bodies for my empirical work, with a special thanks to the musicians in the Piano Phase and Music Moves studies. Thank you Charo, CREW, Tineke, and Studio Radiaal for thinking and working together during your residencies at ASIL. A big thanks to all the members of the Ambisinix (you know who you are), for all the extended, experimental, or just downright chaotic musical jams, let’s stay in touch (I hear

networking is still big these days). A special thanks to my jam-partner Mattia, I hope our musical conversations will continue to unfold.

Besides colleagues and coworkers, the excitement in pursuing this thesis could have only been fuelled when relying on, as I call them, my most loyal support. To my parents, thank you for always being ready to listen and asking the right questions. Your shared attention and limitless care continue to move me. To my brother and sisters, thank you for always having my back, and who knows, I might still invent something new one day. To Jan, Roos, and Mon, thank you for being a second home in Ghent. I will heartfully miss our shared moments, games, and (culinary) degustations. To newfound friends and family in Gent, Toon, Lien, Andreas, Argjira, and Jesse, thank you for the friendship, the walks and talks, the music, the climbs, and welcome moments of distraction. To my dearest Jelva, thank you for your loving presence. Finding each other in times of social distancing, our journey has been a joy, a refuge, and its traces remain, in leaving and in finding each other, time and time again.

Table of Contents

English summary

This thesis explores the potential of extended reality for scientific and artistic research in music making. Extended reality, encompassing augmented, virtual, and mixed reality, is enabled by a suite of technologies aimed at immersing its users in computer-generated, interactive, and multisensorial environments while inducing an embodied sense of presence. This offers the researcher the capability to simulate real-life contexts in the lab with a high degree of experimental control, or create new contexts using “impossible” stimuli such as body illusions, perspective manipulations, and multimodal musical instruments. Through the development of a theoretical and methodological framework, and with a selection of scientific and artistic studies that demonstrate the potential of XR for research into music making, this thesis provides a basis for future investigations into the dynamic and coregulatory processes underlying music making, paving the way towards new and enriching directions for musical expression and experience.

Chapter 1 starts by presenting the philosophical and theoretical foundations of this work, originating from the fields of embodied music cognition, coordination dynamics, and human-computer interaction. It then introduces the concept of extended reality together with its main characteristics of immersion, presence, and virtual embodiment. The chapter uses the “extended musical space” to present an overview of the thesis, followed by the research questions and a note on methodology.

Chapter 2, 3, and **4** are built upon each other, respectively presenting a theoretical paradigm, a methodological framework, and its operationalization in an empirical study to assess performative, embodied coregulatory, and experiential dynamics of music making in XR. Chapter 3 presents a case-study of a piano duet in VR in which a human performer engaged with a human- and algorithmically-controlled virtual human, showing successful music making for the former and suggesting improvements for the latter. The study presented in Chapter 4 applied the methodological framework on a polyrhythmic, musical interaction between two human-controlled virtual humans in XR and showed how musical dynamics remained intact as compared to a real-life setting.

Chapter 5 then focuses on XR’s potential for biofeedback, by analysing behavioural and experiential effects of an adaptive, spatial, auditory stimulus during a spontaneous breathing exercise. The presented study showed how a natural soundscape was most beneficial to induce relaxed states and slow down breathing, while a noise stimulus mimicking a breathing sound motivated healthier breathing inhale-exhale ratios.

Chapter 6 and **7** respectively present four scientific and six artistic studies of music making in XR. Abstracts of each study are presented, together with a discussion on XR designs and implications for future work. The studies of chapter 6 show the potential of XR for

investigations into interpersonal coordination, learning, and feelings of togetherness, while chapter 7 shows its potential for musical performance, composition, and narration. The thesis ends with **chapter 8** in which contributions, limitations, and a conclusion are presented together with a note on future extended reality humanities research.

Nederlandstalige samenvatting

Dit proefschrift verkent het potentieel van extended reality voor wetenschappelijk en artistiek onderzoek in het maken van muziek. Extended reality, dat augmented, virtual en mixed reality omvat, wordt mogelijk gemaakt door een reeks technologieën die erop gericht zijn gebruikers in een door de computer gegenereerde, interactieve en multisensoriële omgeving te plaatsen, terwijl ze een belichaamd gevoel van aanwezigheid behouden. Dit biedt de onderzoeker de mogelijkheid om realistische contexten in het lab te simuleren met een hoge mate van experimentele controle, of om nieuwe contexten te creëren met behulp van "onmogelijke" stimuli zoals lichaams illusies, perspectief manipulaties en multimodale muziekinstrumenten. Door de ontwikkeling van een theoretisch en methodologisch kader, samen met de selectie van wetenschappelijke en artistieke studies die het potentieel van XR voor onderzoek naar het maken van muziek aantonen, biedt dit proefschrift een basis voor toekomstig onderzoek naar de dynamische en coregulerende processen in het maken van muziek en voor nieuwe en verrijkende manieren voor muzikale expressie en ervaring.

Hoofdstuk 1 begint met het presenteren van de filosofische en theoretische grondslagen van dit werk, afkomstig uit de gebieden van embodied music cognition, coordination dynamics en human-computer interaction. Het introduceert vervolgens het concept van extended reality samen met de belangrijkste kenmerken van immersie, aanwezigheid en virtuele belichaming. Het hoofdstuk geeft een overzicht van het werk dat in dit proefschrift wordt gepresenteerd aan de hand van de "extended musical space" en geeft een kader voor de onderzoeksvragen samen met de methodologie.

Hoofdstuk 2, **3** en **4** zijn op elkaar gebouwd en presenteren respectievelijk een theoretisch paradigma, een methodologisch kader, en de operationalisering ervan in een empirische studie om de performatieve, belichaamde coregulerende en ervarings' dynamieken van het maken van muziek in XR te beoordelen. Hoofdstuk 3 presenteert een case-study van een pianoduet in VR waarin een menselijke artiest een interactie aangaat met een door mensen en algoritmen bestuurde virtuele mens, waarbij succesvol muziek wordt gemaakt voor de eerste, en verbeteringen worden voorgesteld voor de laatste. De studie in hoofdstuk 4 paste het methodologische kader toe op een polyritmische, muzikale interactie tussen twee door mensen bestuurde virtuele mensen in XR en liet zien hoe de muzikale dynamiek intact bleef in vergelijking met een realistische setting.

Hoofdstuk 5 richt zich vervolgens op het potentieel van XR voor biofeedback, door gedrags- en experiëntiële effecten van een adaptieve, spatiële, auditieve stimulus tijdens een spontane ademhalingsoefening te analyseren. De gepresenteerde studie liet zien hoe een natuurlijke geluidsomgeving het meest gunstig was om ontspannen toestanden op te wekken en de ademhaling te vertragen, terwijl een geluidsstimulus die een ademhalingsgeluid nabootst, een gezondere inadem-uitademverhouding motiveerde.

Hoofdstuk 6 en **7** presenteren respectievelijk vier wetenschappelijke en zes artistieke studies over het maken van muziek in XR. Samenvattingen van elke studie worden gepresenteerd, samen met een discussie over de XR-ontwerpen en implicaties voor toekomstig werk. De studies van hoofdstuk 6 tonen het potentieel van XR voor onderzoek naar interpersoonlijke coördinatie, leren, en gevoelens van saamhorigheid, terwijl hoofdstuk 7 het potentieel laat zien voor muzikale uitvoering, compositie en vertelling.

Het proefschrift eindigt met **hoofdstuk 8** waarin bijdragen, beperkingen en een conclusie van dit werk worden gepresenteerd, samen met een opmerking over toekomstig onderzoek.

Research output

List of peer-reviewed journal articles

Articles included in thesis

Van Kerrebroeck, B., Rosso, M., & Maes, P. J. (2020). Linking embodied coordination dynamics and subjective experiences in musical interactions: a renewed methodological paradigm. DOCUMENTA, 38(1), 38-60.

Van Kerrebroeck, B., Caruso, G., & Maes, P. J. (2021). A methodological framework for assessing social presence in music interactions in virtual reality. Frontiers in Psychology, 12, 1880.

Van Kerrebroeck, B., & Maes, P. J. (2021). A Breathing Sonification System to Reduce Stress During the COVID-19 Pandemic. Frontiers in Psychology, 12, 623110.

Van Kerrebroeck, B., Crombé, K., Wilain, S., Leman, M., & Maes, P. J. (2022). The virtual drum circle: polyrhythmic music interactions in extended reality. *Manuscript under review*

Abstracts included in thesis

Onderdijk, K. E., Swarbrick, D., **Van Kerrebroeck, B.**, Mantei, M., Vuoskoski, J. K., Maes, P. J., & Leman, M. (2021). Livestream experiments: the role of COVID-19, agency, presence, and social context in facilitating social connectedness. Frontiers in psychology, 12, 647929.

Rosso, M., Gener, C., **Van Kerrebroeck, B.**, Maes, P. J., & Leman, M. (2022). Sensorimotor synchronization dynamics from allocentric and egocentric perspective. *Manuscript under review*

Campo, A., Michałko, A., **Van Kerrebroeck, B.**, Stajic, B., Pokrić, M., Leman, M. (2022). The assessment of presence and performance in an AR environment for motor imitation learning: a case-study on violinists. *Manuscript under review*

Vidal M., **Van Kerrebroeck, B.**, Aguilera, A. M., Maes, P. J., Fritz, T. H., & Leman, M. (2022). Expressive virtual agent during music performance induces critical levels in brain systems of arousal. *Data-collection and analysis on-going*

List of conference contributions

Caruso, G., **Van Kerrebroeck, B.**, & Maes, P. J. (2020). Piano Phase for two pianists in VR. Expanded Animation 2020: Synaesthetic Syntax, Ars Electronica, Linz, Austria, *Online presentation*

Van Kerrebroeck, B., Hermans, C., & Maes, P. J. (2021). Being Hungry – a 3D spatialized audio version presented in a multimodal Mozilla Hubs environment. ICMC 2021, *XR music installation*

Michałko, A., Campo, A., **Van Kerrebroeck, B.**, Stajic, B., Pokrić, M., & Leman, M. (2022). Exploring the potential of augmented reality in instrumental music learning. Musical Togetherness Symposium, Vienna, Austria, *Onsite presentation*

Van Kerrebroeck, B., Crombé, K., Wilain, S., Leman, M., & Maes, P. J. (2022). Joint-action dynamics of polyrhythmic music interactions in augmented-reality. Musical Togetherness Symposium, Vienna, Austria, *Onsite presentation*

Onderdijk, K. E., Michałko, A., & **Van Kerrebroeck, B.** (2022). SysMus22 proceedings. SysMus22, Ghent, Belgium.

List of artistic-scientific studies

Van Kerrebroeck, B., Caruso G., & Maes, P. J. (2020). Piano Phase with virtual pianist. De Krook, Ghent, Belgium. *Public performance of Piano Phase (Steve Reich, 1967) for pianist with virtual partner*

Van Kerrebroeck, B., & Maes, P. J. (2021). Lines And Swarms. ASIL, Ghent, Belgium. A *spatial audio composition tool in virtual reality*

Van Kerrebroeck, B., Hermans, C., & Maes, P. J., (2021). Being Hungry. ASIL lab of IPEM, Ghent, Belgium and online in Mozilla Hubs. *Spatial audio composition presented in IPEM's Art and Science Interaction Lab and on the Mozilla Hubs platform*

Van Kerrebroeck, B., De Mulder, S., & Lareu, B. (2021). Searching Borders. Best submission HacXathon 2021 during 24 Hours of Deep Listening at Tinnepot theatre, Ghent, Belgium. *Immersive virtual experience*

Carmen, L., **Van Kerrebroeck, B.**, & Maes, P. J. (2022). Be Hear Now. ASIL lab of IPEM, Ghent, Belgium. *Immersive, interactive audiovisual installation at IPEM's Art and Science Interaction Lab*

Van Kerrebroeck, B., Van Kets, N., Moens, B., Van Wallendael, G., & Maes, P. J. (2022). MusicMoves. Onze (on)bekabelde cultuur symposium, Ghent, Belgium, *Hybrid (online-offline) XR music performance*

1.

Background

1.1 Introduction

Music is powerful. It moves us, literally (Bernardi et al., 2009; Burger et al., 2013; Schiavio et al., 2014) and figuratively (Juslin & Sloboda, 2013). While the focus of this thesis is not music itself, if such a thing were even possible, the work presented here is centred in the "musical space" in which people experience, compose, perform, and interact with music and each other. Music is considered as the cultural activity and temporal art of organised sound, integrating expressive and meaningful intentions, working mainly through experiential mechanisms rooted in the body (Leman, 2007). This music making is seen as an inherent social and embodied practice, "not just as a social activity, but one that generates specific socialities" (Cook, 2013), in which humans are able to coordinate and coregulate their (inter)actions and intentions (Leman, 2007; Lesaffre et al., 2017). The musical spaces are thus viewed holistically, as spaces in which performers, audience and the environment co-create the music as an integrated whole, and in which the sum outgrows its individual parts. Music making seen in this way aligns with the notion of "musicking" introduced by Christopher Small, in which music is not seen as a representational object, but as a process whose dynamic and active constitution calls for a verb (Small, 1999). Musicking thus refers to the historical and contemporary performance practices in which emphasis is placed on the participatory and social nature of music making, as opposed to the mere presentation of musical material to a passive audience.

While music making is a fundamental part of human life, musical artefacts have accompanied modern humans for millennia (Cross & Morley, 2010), it has been seriously impacted by the advent of electronic and digital technologies in the 20th century. While the technologically mediated nature of musical communication can pose challenges for musical sense-making and the emergence of shared experiences, they also offer an innovative and creative potential (Holland et al., 2013; Landy, 2019; Leman, 2007). Integrating technological means in the practice of music making has enabled new compositional approaches dealing with original and synthesized sounds, led to the widespread recording and diffusion of musical material, and allowed musical playback and interaction in original configurations. Also, with ever evolving technological capabilities and signal resolutions, one could say the medium moves more to the background, with increased transparency promising the illusion of "non-mediation", capable of retaining the original and meaningful sense-making activities it is meant to facilitate. However, it is important to keep in mind how technological development always occurs in a social, economic, political, and cultural fabric, and how with each technical innovation, existing attitudes, practices, and relations evolve in dynamic co-dependency alongside it. Considering the technological lifeworld today (Ihde, 2009), music making in a networked and digitally extended reality offers enormous opportunities, as well as highlights fundamental questions about the sense-making and co-regulatory dynamics that constitute them.

Extended reality (XR), encompassing augmented, virtual, and mixed reality (AR, VR and MR), has been one of the most influential developments for music making and research in the last decade (Çamcı & Hamilton, 2020; Turchet et al., 2021). XR encompasses a plethora of technologies to create new environments, or simulate and enrich existing ones, via computer-generated interactive multisensory displays (Scarfe & Glennerster, 2019; J. Taylor, 1997). Although there is no consensus of a clear definition, XR is generally seen as the umbrella term for the various technologies blending our digital and physical worlds. By immersing its users into digitally augmented environments, it promises an illusory sense of non-mediation (Lombard & Ditton, 1997; Riva et al., 2007). By creating rich and realistic contexts, XR should be able to increase the ecological validity of experimental research while retaining high levels of control (Kothgassner & Felnhofer, 2020). On the other hand, by placing music making in a digital environment, XR also promises to create new and innovative contexts enriching for musical expression and experience (Turchet et al., 2021). Given its potential to improve our understanding and enrich our experience of music making, it is the aim of this work to offer a comprehensive investigation and evaluation of doing music research with XR (chapters 2-4), evaluate its biofeedback potential (chapter 5), and present a selection of scientific (chapter 6) and artistic (chapter 6) demonstrations. The work presented here will demonstrate the value of XR as a methodological tool for interdisciplinary, action-oriented, participatory, and embodied music research. XR technologies will be used to build digitally augmented environments that allow the emergence of the coregulatory dynamics of music making, opening the “extended musical space”, and the creation of rich contexts required for scientific and artistic music research.

1.2 Philosophical foundations

The research into music making presented in this thesis is performed through an interdisciplinary approach taking place in technologically-laden musical environments. It is characterized by the oscillation between an active and utilitarian stance in the development of technological tools, and a more contemplative stance leading to the creation of, and critical reflection on, music making. This pragmatic and hermeneutic approach emphasizes knowledge as originating from the self-reflective inquiries and embodied, action-oriented engagements in a historical, social, and “technological lifeworld” (Heelan & Schulkin, 1998; Ihde, 2009). Before presenting the theoretical foundations, XR, and musical spaces in this work, this section takes a step back and introduces its techno-cultural and interdisciplinary context.

1.2.1 The technological lifeworld

Since XR technology represents such a central part in this thesis, as locus of investigation and as methodological tool for research into music making, consideration must first be given to its foundational characteristics. Technology is considered not so much as an expression of human nature, whether we are tool users or tool makers, but rather as defining an increasingly dominant kind of sociocultural practice (Scharff & Dusek, 2014). In the active and reflective process of inquiry (Hickman, 2014), we use technology to obtain improved conditions and alter or adapt to our environment more effectively, leading to the progressive blurring of the distinction between reality and representation, or a “heterogenesis” of the interface between human being and its sensory environment (Gallese, 2020). Therefore, it is when we work in experienced “situations” involving our practical-technical engagements, rather than in our quest for abstract theoretical knowledge, that we can improve our understanding of the relationships with our surroundings and other people (Mitcham, 2014). Science and technology are thought of as practices in which we are materially and intellectually submerged (Pickering, 2014), allowing for thinking to take place within the “clearing” where our relationship with things and people take place, with the aim of creating the possibility for a “free relation with technology” (Heidegger, 2014; Scharff & Dusek, 2014). These views have led some to speak of the “technological lifeworld” (Ihde, 2009), a technological extension of the socially, linguistically represented, and action-oriented “lifeworld”, where the ontological status of things derive directly from their meaning and experiencing within human life (Heelan & Schulkin, 1998; Husserl, 1970).

Research in such a world, whether scientific or artistic, considers technology an inherent and non-negligible part of its endeavours. It goes beyond the notion of “essential” descriptions of self-world correlations, and leads to the postphenomenological stance of seeing technologies as acting upon our practices and as having their own “technological intentionalities” (Ihde, 2009). These intentional relations are seen as not fixed but “multistable”, fully meaningful in multiple ways (Verbeek, 2014). As in the experienced “situations” mentioned before, this stance requires an acknowledgement of the fact that all self-world correlations are embodied relations, and part of a research process that works through a practical and technologically mediated existence (Ihde, 2009). This process sees action and performance, not knowing, as the basic condition of the world, and allows for “the dance of agency”, the endless accommodation and adjustment between humans and machines (Pickering, 2014; Scharff & Dusek, 2014).

Media is seen as the currency and fabric for such human-machine interactions. Some have stressed the mediated nature of our relation to the world, less defined from direct experience and bodily interaction as from representations in the distance (Debord, 1995;

Gallese, 2020). Our experience can be said to take place in a “mediascape”, or in the more recent notion of the “metaverse”, the environment promoting and facilitating mediation between individuals and reality through technological artefacts (Casetti, 2018). This collectively built environment has strong associations with the “virtual”, that notion of abstract reality that “concerns the potency in what is, by virtue of which it really comes to be”, connoting “a force of existence” and of “potential” (Massumi, 2014). This “virtual lifeworld” has an important social dimension, as a meeting place to exchange expressive intentions, as a form of mediated intersubjectivity, and as a shared practice to build collective memories and create history (Gallese, 2020; T. K. Metzinger, 2018). The virtual could be seen as the expression of a human drive to extend ourselves into thought, of which XR would be only its latest instantiation (Frankel & Krebs, 2021). Digital technology has not instantiated the virtual, but did contribute to its history, and provided new ways to conceptualize, use, and experience it (Grimshaw, 2014). XR, having a big impact as “potential space” and “laboratory of subjectivity” (Frankel & Krebs, 2021), represents a paradigmatic shift in this evolution, and might place us significantly closer to the so-called “hyperreal” world (Baudrillard, 2000). Given the losses and gains of every technological advance (Doel & Clarke, 2013; Stiegler, 2019), it remains important to accompany these advances with comprehensive study and reflection, such as the one aimed for in this work.

1.2.2 Augmented humanities

Traditionally, the humanities are a set of academic disciplines involving a hermeneutic and critical tradition concerned with human culture and society. They employed proper methods and practices on mainly static cultural artefacts for the critical reflection on, and interpretation of meaningful relations. However, computational power introduced by digital technologies, together with empirical methods from the natural sciences, have extended the scale of the material and tools available for humanities research. This has led to a digital turn, in which digital solutions were made available to answer humanities research needs (Six, 2018). This so-called digital humanities field, also known as humanities computing, lying at the intersection of computing and the humanities, moved the technological component from mere support to a full intellectual endeavour, and was “methodological by nature and interdisciplinary in scope” (Kirschenbaum, 2016). These tools motivated a search for patterns throughout the humanistic tradition leading to the so-called “humanities 2.0”, defined as “the identification and representation of patterns by digital means” (Bod, 2013). This has already led to a successor, “humanities 3.0”, in which the original hermeneutic and critical tradition is applied to the digital tools and patterns obtained in the humanities 2.0 (Bod, 2013).

What can be seen as the next step from the digital humanities, or of the humanities 3.0, is the field of the “augmented humanities” (Six, 2018). It requires the humanities researcher to work in the context of human-artefact interaction, rather than with static archives and unidirectional computational analysis. It starts from the notion that cultural artefacts thrive on interaction, as when someone replays a (simulated) historical musical instrument. By placing the cultural artefact in an interactive context facilitated by digital means, augmented humanities research acquires the capacity to manipulate contextual variables, and to test hypotheses while the practice under consideration is being re-enacted. Formulating appropriate hypotheses would then require an understanding of how humans develop meaningful, sense-making interaction with their (musical) environment and the underlying interaction dynamics (Leman & Six, 2018; Mitchell, 2018). This approach could inform the "living culture of interaction", a concept enabled by the advent of interactive multimedia and proposed to deal with the floating and volatile character of cultural (interaction) heritage (Leman & Six, 2018). This living culture of interaction is meant to induce and empower the culturally relevant social interactions, leading to stable and expressive or "homeostatic" outcomes (Leman & Six, 2018). Like the augmented humanities, this living culture is enabled by an interdisciplinary approach, covering aspects of documentation, presentation through interaction, and aims to involve the larger (scientific) community.

This need for an interdisciplinary approach has also been stressed as the requirement for a more proactive stance of the humanities (Lesaffre & Leman, 2020). Such a stance would give the humanities an active role in (co-)steering the innovative developments of the techno-culture, on top of being a reflective, critical, and descriptive discipline. Lesaffre and colleagues (Lesaffre & Leman, 2020) argue how such proactive humanities are built from a tight integration of the arts and the sciences in interactive, multimedia contexts, in which human expressive communication and sense-making are investigated. They argue how the arts have the capacity to create rich, meaning-laden contexts that provide a strong basis for improved scientific understanding through measurement, observation, and manipulation which in turn would allow critical reflections on the artistic process. This integration of the arts and sciences has been discussed elsewhere (Siler, 2011; Tress et al., 2007), with some advocating for a closely integrated "entangled" approach (Lesaffre & Leman, 2020; Wernli et al., 2016) that includes the design and engineering disciplines (Oxman, 2016). Such "entanglement" would move beyond the mere application of methods and results across disciplines, would be driven by questions instead of methodologies, and would result in a close coupling between disciplines where modifications in one discipline immediately affect the other (Lesaffre & Leman, 2020; Oxman, 2016). The work presented in this thesis can be seen as the outcome of such an entangled practice that is centred in the digitally augmented musical environments enabled by XR. Its focus is on the scientific and artistic (musical) potential of such environments while taking design and engineering considerations into account. XR technologies are used and critically reflected upon for their capacity to simulate existing, as well as create new, rich, and meaning-laden contexts required for social, embodied, and expressive music making. The next section will first present a broad overview of the theoretical foundations on which these contexts can be built.

1.3 Theoretical foundations

The studies presented in later chapters will use XR technology to create and critically reflect on digitally augmented musical environments. XR technology is promising for music research as it offers novel and flexible ways of simulating and enriching music making while promising an illusory sense of non-mediation. The theoretical frameworks presented in the following sections are foundational for this approach, as they offer the required scientific insights, methodological tools, and models for experimental scientific-artistic research. Embodied music cognition provides a comprehensive research paradigm to describe and explain musical sense-making. Coordination dynamics offers consistent terminology and mathematical modelling to capture and simulate the spatiotemporal and relational principles of the musical interaction processes. Finally, the human-computer interaction field offers design and engineering guidelines to mediate and enable musical creativity and expressivity. The following sections only provide a brief overview of the concepts most relevant for this work, with comprehensive accounts listed in the references (for embodied music cognition (Leman, 2007; Lesaffre et al., 2017), coordination dynamics (Kelso, 1995; Tognoli et al., 2020), and human-computer interaction (Holland et al., 2013)).

1.3.1 Embodied music cognition

Embodied cognition is based on the idea that cognition is fundamentally grounded in its physical context and does not require a “final” (brain) region in which experience is abstracted and integrated together (Niedenthal et al., 2005). It articulates how our cognitive experiences are shaped and constrained by our bodies, and how cognition arises from our bodily interactions with the world (Thelen et al., 2001). Our sensorimotor processes are thus closely integrated, leading to tight action-perception couplings in the exploration of, and adaptation to our environment. Embodied cognition is closely linked to notions of extended and enacted cognition, referring to the idea that extrabodily processes or tools play a non-negligent role in the constitution of cognition, and the emphasis on the need for active engagement in and with the environment (Newen et al., 2018; Svare, 2006). As such, humans use technology to “reach out” into their environments, assisting in the drive to make sense of their embodied and embedded selves. Embodied interaction with our environment can thus be seen as a means of active inference, and as assisting the interoceptive processes that help to maintain homeostatic balance (Seth, 2013; Seth & Tsakiris, 2018). These notions of feedback and stability have pushed embodied cognition theory to a more dynamic approach (Maes et al., 2018), considering embodiment as only one component in an interconnected network of sensory, motor, affective, and cognitive processes.

Performing cognition research in such an active and dynamic setting has required a more action-oriented stance, resulting in the so-called pragmatic (Engel et al., 2016) and interactive turn (De Jaegher et al., 2010). These turns have argued for more ecologically valid paradigms and a move towards “there where the action is” (Engel et al., 2016). This move is important as, in their view, cognition does not serve to produce models of the world, but serves action, as it is grounded in sensorimotor skills.

A good example of such an action-oriented stance is the area of joint-action research, highly relevant for research into music making. Joint actions are ubiquitous and an essential part of human life, such as when people are playing sports or music, dancing or moving together. They happen when people coordinate their actions to achieve a shared goal, requiring the monitoring, planning, prediction, and communication of actions while keeping one’s own and others’ contributions apart (Sebanz et al., 2006; Sebanz & Knoblich, 2021). This mutual coordination does not only apply to actors’ actions, but to (joint) expressive goals, emotions, and mental states as well (Keller et al., 2014; Sebanz & Knoblich, 2021). Music, with its ability to incorporate and communicate expressive intentions and evoke emotional responses (Leman, 2007), can be seen as a refined form of such interpersonal coordination, and as a rich area for joint-action, and by extension, (social) cognition research.

This potential for music research is captured in the embodied music cognition framework, which emphasizes the role of the human body as a mediator for meaning formation, able to establish an intentional level of musical interaction through corporeal articulations and imitations (Leman, 2007). Validation of this framework has come from empirical evidence of the influence of bodily states and movement on music perception (Maes et al., 2014), shown to automatically engage multi-sensory and motor simulation processes (Schiavio et al., 2014).

The framework of embodied music cognition has been able to accommodate a large body of research. For instance, it has spurred empirical studies focusing on the role of corporeal and corporeal-technological mediators (Nijs & Leman, 2014), and the role of expressive gestures in meaning formation and communication in music making (Desmet et al., 2012). Other studies have looked at body sway (Chang et al., 2019; Demos et al., 2018), physiology (Gordon et al., 2020; Konvalinka et al., 2023), or gaze patterns (Kawase, 2014) to gain insights into the embodied coregulatory processes that help to establish and maintain interpersonal synchrony and leader-follower relations (Keller, 2014). Given the strong subjective experiences that music can evoke, and the significant emotional effects of moving together in time (Juslin & Sloboda, 2013), research in the embodied music cognition framework has also helped to gain a better understanding in the prosocial effects of synchronization and entrainment (Launay et al., 2014; Stupacher, 2019; Tarr et al., 2018).

The introduction of XR technology and its capability to present interactive stimuli sufficiently rich in detail and with appropriate levels of behavioural realism, has allowed researchers to continue these investigations with increased levels of experimental control while retaining high ecological validity (Launay et al., 2014; Tarr et al., 2018). While XR technology has already introduced new research paradigms in other lines of research, such as through highly effective body-illusions, and facilitated controlled and replicable scenarios using virtual humans (De Oliveira et al., 2016; Matamala-Gomez et al., 2021; Petkova & Ehrsson, 2008; Seinfeld et al., 2018), its potential for the embodied music cognition framework is only beginning to be exploited. By combining the social, embodied, and expressive nature of music with the flexible, scalable, and controlled nature of XR, novel experimental scenarios can be brought to a large population and continue the scientific successes stemming from the embodied music cognition framework.

1.3.2 Coordination dynamics

Coupling, feedback, and dynamics have already been part of the theory introduced by the embodied (music) cognition framework in the preceding section. They were central in the view of an embodied actor embedded in, and interacting with its environment, as well as in the coregulatory dynamics and homeostatic processes between acting, thinking, feeling, or music making subjects. While these dynamical notions have mainly been theoretical or descriptive, coordination dynamics is the framework that helps to study, empirically test, and analyse these dynamics using mathematical models and a consistent terminology. For instance, the framework of coordination dynamics has led to insights into the organizing principles underlying sensorimotor (Demos et al., 2019; Heggli, Cabral, et al., 2019) and group coordination (Bégel et al., 2022; Dotov et al., 2022; Shahal et al., 2020), provided methods for the analysis of spontaneous music making (A. E. Walton et al., 2018), as well as helped to describe the exploratory and playful nature of improvisation (Borgo, 2005; A. E. Walton et al., 2015, 2018) and creativity (van der Schyff et al., 2018). As these examples show, the framework of coordination dynamics is not bound to a specific level of observation and can be applied to a plethora of music making activities involving rich and spontaneous coordination unfolding over time.

Coordination dynamics is built on the foundations and principles of dynamical systems theory and complexity science. These are essentially concerned with the dynamics that come from the non-linear coupling between agents and their environment. Both theories study the phenomena that emerge from a collection of interacting, self-regulating agents, which themselves are affected by memory or feedback and appear in an environment that constrains and shapes their behaviour (De Wolf & Holvoet, 2004). Interacting agents can be seen as self-organising by acquiring and maintaining a spatial, temporal, or functional structure without external control. This could then lead to the spontaneous formation of ordered states in the form of properties, behaviour, or patterns that tend to maintain some identity over time. These emergent phenomena possess constitutive characteristics that are relatively insensitive to perturbations and are not explainable by their isolated parts. Dynamical systems theory and complexity science focus on the underlying organizing principles of such emergent phenomena and have helped to model, simulate, and predict, a wide-ranging variety of natural and artificial systems. They describe and mathematically model system dynamics using physical or informational variables that can vary from dopamine levels and light intensity to speed and relative phase, amongst many others.

While it is not the intention to give a comprehensive overview of dynamical systems theory and complexity science here (for a comprehensive dynamical systems account of music making, the reader is referred to (Demos et al., 2014)), it is worth introducing a few central concepts that have a prominent place in coordination dynamics. System dynamics are generally considered to be non-stationary, aperiodic, and evolving over time. They are often scale-free, with properties such as entropy and density persisting locally and globally. Dynamics are characterized by their attractors, specific states or bounded areas to which the system can be pulled, and in which the system exhibits some stability. These attractors can be identified by searching for critical fluctuations in the dynamics of the system when it is nearing them. By changing environmental constraints or the system's parameters, the number and stability of attractors can be changed, leading to so-called bifurcations. By having change, stability, adaptability, and robustness at the core of the dynamical systems theory and complexity science, they are ideally equipped to deal with the sensitive and fragile, dynamic and reactive processes underlying interpersonal coordination.

Integrating these principles, coordination dynamics attempts to find general systems of equations for the lawful, interpersonal coordination patterns at multiple levels of description (Kelso, 2009; Tognoli et al., 2020; Warren, 2006). A key part in the discovery of such laws is the distinction between an order variable that captures the dynamics of the coordination, and state or control variables of individual system parts and environmental constraints. Like in dynamical systems theory, these laws are characterized by non-linear relations, attractors, bifurcations and are sensitive to initial conditions. For coordination dynamics, these concepts have received broader implications and can result in qualitatively different phenomena. For instance, bifurcation mechanisms have shown to improve the stability and recall in the learning of coordination patterns (Kostrubiec et al., 2012), and sensorimotor exploration of dynamics in a state space, jump-started by imitation or instruction, can help to determine control laws for a task (Lamb et al., 2019; Warren, 2006). Initial conditions in coordination dynamics can thus refer more broadly to the importance of context, task, and individual differences, codified as intrinsic dynamics in the system. Finally, an important notion in coordination dynamics is that of multi-stable or meta-stable regimes (Kelso, 2009). They refer to a marginally stable system state from which different attractors are readily accessible. These marginally stable states allow for the emergence of "functionally meaningful information" and the coexistence of competing tendencies for autonomy and coordination (Kelso, 2009). Speculatively, these regimes appear to share similarities with the marginally stable homeostatic processes present in the embodied music cognition framework, and could represent the fertile, flexible, and ambiguous ground needed for spontaneous and creative music making.

Given the capabilities of coordination dynamics in the modelling of human behaviour, it represents an ideal framework for XR technology to create richer and more realistic multisensorial environments. It could support the development of virtual humans with higher levels of behavioural realism which, when experimented on, could in turn help to improve underlying behavioural models. For instance, a paradigm such as the human dynamic clamp (Kostrubiec et al., 2015), in which participants try to synchronize their finger taps with an algorithmically controlled virtual 2D partner, could inspire the development of more realistic settings in XR using fully embodied virtual agents. Other immersive musical environments have already shown their value as natural and responsive biofeedback applications (Blum et al., 2019; Kitson et al., 2019; S. Moran et al., 2016; Prpa et al., 2018) and could profit from the dynamical models formulated by coordination dynamics. Nonetheless, the design and engineering of such environments requires a lot of expertise, which will be the topic of the next section.

1.3.3 Human-computer interaction

The field of human-computer interaction (HCI) research has traditionally been focused on interface and interaction design of digital applications aimed at specific goals such as increased productivity, improved user experience, or the stimulation of creativity (Beaudouin-Lafon, 2004; Frich et al., 2018). It has developed an extensive set of interaction models, methodologies, and design principles and has been informed by psychological theories of human behaviour, engineering theories of improved HCI practice, anthropological theories of situated (inter)action, and theories residing in the resulting HCI artefact (Zimmerman et al., 2010). The HCI field had its own evolution towards an embodied view on interaction, considering the evolution of our interactions with technology as a gradual incorporation of human skills and abilities (Dourish, 2001). For instance, it has seen the introduction of graphical interfaces as shifting the task of managing information to one of managing space, leveraging abilities such as peripheral attention, pattern recognition, and spatial reasoning. The history of HCI is often presented using three HCI waves (Bødker, 2015; Harrison et al., 2007). They refer to the evolution from a focus on engineering and performance optimization, to one on situated action and user-centric design, to the social, tangible, and ubiquitous computing practices of today (Tanaka, 2019). Research into ubiquitous computing is characterized by a focus on everyday settings that includes experiential, aesthetic, and cultural aspects (Harrison et al., 2007).

When looking at our interactions with technology, one can view the manifestation of technology as a tool, as a partner or as a medium (Reidsma et al., 2014; Verbeek, 2014). Building on this distinction and involving notions of extended and situated cognition, the tool can be framed as a natural extension of the body, or as part of the environment and as requiring a dialogue (Leman, 2016). Central in these views is the concept of mediation and its characteristic of transparency, referring to the ability of an interface to decrease the distance between a user's intent and action. Consequently, some researchers have argued for more transparent mediation technologies requiring the involvement of sensorimotor, cognitive, social predispositions, and user related backgrounds (Maes et al., 2018). Others have stressed the need for interfaces with increased "opaqueness" or "disturbance" allowing for an improved "control intimacy" (Fiebrink et al., 2010; Jäger & Kim, 2021; F. R. Moore, 1988). "Opaque" interfaces with "control intimacy" would offer the rich and consistent feedback required for music making to a user, while "disturbing" interfaces would stimulate creativity and exploration based on a challenging, ambiguous, and interpretable interaction (Impett et al., 2015).

Given this emphasis on feedback and interaction, instead of the traditional computing perspective of HCI with its notions of mediation, tools, and users, it might be beneficial to take a performance perspective on music making technology (Jacucci, 2015; J. Spence et al., 2013). Such a view would transform users to actors and performers, and value expression and experience over usability and replicability. This could be especially relevant for XR environments (Benford & Giannachi, 2011), in which the blending of physical and virtual worlds could be seen as the staging of a fictional space, which then guides and constrains an actor's interpretation and actions (Jacucci, 2015). This performance perspective shares similarities with the design principles proposed for "artful VR", specifically the distinction between doing or acting, being or interpreting, and the importance of play in VR (Atherton & Wang, 2020). These principles are also relevant for the often-overlooked role of audio in VR, and stress the need for immersive, interactive, and dynamically generated audio (Çamcı & Hamilton, 2020; Turchet et al., 2021).

Music making, with its idiosyncratic practices and ambiguous goals, has been a welcome and fertile field for the theoretical frameworks and concepts central in HCI (J. Malloch et al., 2019; Tanaka, 2019). One active research area involving music technology and HCI has been that of "sonification", the transformation of data such as movement or physiology into sound. This field investigates sound and interaction design often for the purpose of biofeedback and has led to applications in learning, art, rehabilitation, and sport (Bevilacqua et al., 2016).

Another active research area combining HCI and music making is that of "New Interfaces for Musical Expression". It has led to the incorporation of HCI methodologies (Holland et al., 2013), and the formulation of design principles for digital musical instruments (Fiebrink et al., 2010) and sonic interactions (Serafin et al., 2011). These interdisciplinary approaches have been combined with XR technology, proposing design principles for virtual instruments (Serafin et al., 2016), 3D interaction techniques (Berthaut, 2020), affordances for musical interaction (Çamcı & Hamilton, 2020), and musical engagement (Deacon et al., 2017). Others have emphasized the importance of collaboration in the virtual environment for musical creativity (Barrass & Barrass, 2006; Schlagowski et al., 2022).

Finally, with new networking technologies and continued development on the "Internet of Things" (Turchet et al., 2018), an increasingly relevant research area is that of "networked music performance". Networked music performance has been broadly defined as the "mediated interaction modality characterized by extremely strict requirements on network latency" (Rottondi et al., 2016). It investigates aspects of bandwidth and latency (Driessen et al., 2011; Rottondi et al., 2016), as well as the experience and musicality of performers (Bartlette et al., 2006; Tsioutas & Xylomenos, 2021) in networked music making. Networked music performance has been combined with XR technology (Cairns et al., 2020; Hamilton, 2019; Loveridge, 2020; Turchet & Rottondi, 2022), raising many opportunities for research into embodied music cognition, and the development of inclusive and accessible applications for networked music making.

1.4 Extended reality

The previous sections have set the philosophical and theoretical foundations on which this thesis is built. They aimed at showing the relevance of these disciplines and frameworks for the construction of "digitally augmented musical environments" for "participatory and creative music making". These "digitally augmented musical environments” are created as XR environments or “extended musical spaces” that offer new affordances, and allow interactions with, and as virtual humans.

Such environments are relevant for research into music making, as they allow a high degree of experimental control, and the creation of rich contexts with dynamic, multimodal stimuli that would be too costly or impractical to achieve in the real world (Blascovich et al., 2002; Fox et al., 2009; Pan & Hamilton, 2018). XR technology, by substituting perception of the real world for virtual simulations, and immersing its users in dynamic environments filled with affordances, allows for a “hacking” of the action-perception cycle, and has been described as a technological metaphor for conscious experience (T. K. Metzinger, 2018).

XR technology, and the environments enabled by it, could thus be used to gain further insights into the dynamical coordination processes of music making. This would lead to improved theories and frameworks of embodied music cognition, coordination dynamics, and human computer interaction, which in turn would lead to better XR environments in a cyclical process of theory and praxis. This section will present the nature of XR itself, specifically its main features of inducing presence and a sense of virtual embodiment in its users.

1.4.1 The virtuality continuum, presence, and immersion

The first steps towards how we understand XR today were taken by Ivan Sutherland in a paper titled “The ultimate display”, and with the resulting stereoscopic, head-mounted prototype a few years later (Sutherland, 1968). While this kind of display was focused on the visual sense, another origin for XR could be found in Morton Heilig’s patent for a stereoscopic-television system describing the use of headphones and binaural sound (Heilig, 1960). In contrast to these two origins, which are deeply rooted in the technological component, another approach would be to focus on the media and digitally augmented or virtual environment itself (Steed, 2014). From this viewpoint, Jaron Lanier is often credited to be the first to use the term “virtual reality”. He used it to refer to computer generated immersive environments which involved 3D objects manipulated using a data glove (Lanier, 1988).

Fast-forwarding to today, extended reality (XR), or cross reality, is considered the umbrella term for augmented, virtual, and mixed reality (AR, VR and MR). Generally speaking, all of these terms do not refer to a particular kind of technology or media type, but concern computer generated and interactive content to create digitally augmented, multisensorial environments for its users. They are usually distinguished using the “virtuality continuum”, spanning fully real environments, i.e. the physical world, on one end, and fully virtual ones, i.e. VR, on the other (Milgram & Kishino, 1994). AR lies in between these extremes and refers to the mixture of real and virtual worlds. Such mixture has digital content placed as an overlay in the real world, or real elements placed inside a virtual environment. The latter was referred to as augmented virtuality in the original Milgram paper but will be identified with AR in this work, as the balance between the amount of real or virtual content is not relevant here. AR is distinguished from VR in that it always retains a trace of the physical space the user is in, with digital content interacting with the user and possibly with the real world as well. MR has been a contested term in the literature (Speicher et al., 2019), but refers here to the combination of AR and VR, in the sense that (users in) both environments can interact with each other, or that both environments are simultaneously offered to a user. Finally, XR as an umbrella term, encompasses all variants of AR, VR, and MR. It is sometimes used interchangeably with MR and rarely clearly distinguished from it. Here, XR is seen as the "ubiquitous mixed reality environment", stressing its independence from the technological medium (whether head-mounted, wearable, or mobile), and highlighting its inherent social and dynamic nature in accommodating human-driven avatars and animate data from (neuro)physiological, social, or natural sources (Paradiso & Landay, 2009).

So far, this definition of XR has avoided the inclusion of any particular sensory modality. While the original characterizations of XR have focused on visual displays, more recent accounts have stressed the need to not only account for non-visual stimuli, but also regard them as fundamental aspects of an immersive XR experience (Çamcı & Hamilton, 2020; Turchet et al., 2021). Although vision is often considered a dominant sensory modality, other senses have been shown to modulate visual perception through cross-modal interactions, which could play important roles for our unified experiences through multimodal integration (D. Martin et al., 2022; Shams & Kim, 2010). Multimodal cues are especially relevant in XR, as they ground the user's experience in the physical domain, and can enhance or modulate the virtual experience (for a comprehensive review of multimodality in VR, see (D. Martin et al., 2022)). Sound can play an important role, as it penetrates our physical space, and can provide information in the background and outside the field of view (C. Spence et al., 2020). It is thus ideally suited to deliver proper 3D realism, which has been illustrated by the contribution of spatial audio to presence in immersive virtual environments (Huang et al., 2019; Kobayashi et al., 2015). This importance of sound in virtual environments has led to the notion of "audio-first VR" (Çamcı & Hamilton, 2020), and has stimulated research into the technical aspects for improved immersive sound (Reyes-Lecuona et al., 2022).

A key characteristic of XR is its ability to evoke a feeling of "presence" in its users by "immersion" in a digitally augmented environment. Presence in XR is an active research theme in itself, and its relation to immersion remains heavily debated (for a literature review, see (Agrawal et al., 2020; H. Lee, 2020)). While some consider immersion as a psychological experience resulting from a technological process, others consider immersion as the objective property and “technical capability of the system to deliver a surrounding and convincing environment” (Sanchez-Vives & Slater, 2005; Slater, Lotto, et al., 2009). Some have proposed several dimensions of immersion, for example based on perceptual or narrative elements, or distinguished between the “immersive experience” and the “immersive system” (H. Lee, 2020). Experiential aspects of immersion are considered here as part of the notion of presence, seen as the “human reaction to immersion” (Slater, Lotto, et al., 2009). While there can be a degree of perceptual immersion, evaluated based on the number of sensory modalities accommodated in the digitally augmented environment, and a degree of system immersion, evaluated based on technical capabilities such as field of view or number of audio channels, degrees of “immersive experience” or “psychological immersion” are here considered as part of the feeling of presence.

Presence then, has been defined as “the phenomenon of behaving and feeling as if we are in the virtual world created by computer displays” (Sanchez-Vives & Slater, 2005). It results from the “perceptual illusion of non-mediation” (Lombard & Ditton, 1997; Riva et al., 2007), requiring a low latency sensorimotor loop, statistical plausibility and behaviour-response correlations, and is rooted in activity (Slater, Lotto, et al., 2009). Presence manifests as a multi-layered response, from unconscious physiological processes towards automatic behaviours and reflexes, up to the highest cognitive levels, including the “feeling of being there” (Riva et al., 2011; Slater, Lotto, et al., 2009). It is distinguished from the degree of engagement and involvement, as one can have high presence with low involvement, and is by some even considered independent from it (Skarbez et al., 2017). Feelings of presence have been shown to interact with emotional responses, and can be influenced by a myriad of contextual and social factors still under active investigation (Servotte et al., 2020).

Presence is generally distinguished in three main components: physical, social, and self-presence (K. M. Lee, 2004). Physical presence, sometimes labelled as spatial, environmental, or tele-presence, refers to the “place illusion”, and is generally defined as the “sense of being there” in the virtual environment (Slater, Lotto, et al., 2009). Social presence refers to the feeling of interacting with others, or the degree to which the user “feels to have access to the intelligence, intentions, and sensory impressions of another” (Biocca et al., 2003). It has an interesting variant in the form of co-presence, referring to the feeling of simply being with others (Biocca et al., 2003). Finally, self-presence refers to the feeling of identification with the (virtual) self, and can originate from an embodied, physiological, emotional, or cognitive level (K. M. Lee, 2004). These distinctions illustrate the experiential nature of presence, and can apply across the virtuality continuum, with presence in AR reflecting the extent to which the line between the real and the virtual becomes blurred (McCall et al., 2011). While questionnaires have been developed to measure the different categories of presence or a person's propensity for it (Makransky et al., 2017; Witmer & Singer, 1998), they have been criticized for several reasons. First, there are the methodological issues in probing for experiential aspects retroactively. Second, there is the problem of using static questions to measure a dynamic process. Finally, there is the challenge (or impossibility) in quantifying presence with global measures due to its multi-layered embodied, behavioural, emotional, and cognitive nature. Addressing these challenges will be the topic of chapter 3.

1.4.2 Virtual embodiment

One of the main tools of XR technology, is its ability to immerse its users and offer a sense of virtual embodiment while inducing feelings of presence. It can induce feelings of social presence by generating embodied others in the form of virtual humans. It can induce feelings of self-presence by allowing its users to control a virtual body from a 1st person perspective. Under specific multisensory conditions, XR can be used to generate a "body ownership illusion" or "body distortion illusion", in which people experience fake body parts or whole fake bodies as their own (Kilteni et al., 2015; Normand et al., 2011). Such illusions have led to the paradigm of "body swapping", in which people feel like having a fake body such as a mannequin (Petkova & Ehrsson, 2008), a virtual human (Slater et al., 2010), or even a real, other person (De Oliveira et al., 2016). A comparable setup has also shown capable of inducing "out of body illusions", in which users saw themselves (Ehrsson, 2007) or their virtual bodies (Lenggenhager et al., 2007) from a 3rd person perspective, which led to illusory drifts in their experienced self-location.

As these illusions and their impact on bodily self-location, body representation, and perspective taking illustrate, XR technology offers the capability to selectively influence different representational layers of the human self-model (T. Metzinger, 2014; T. K. Metzinger, 2018). For example, through the body ownership illusion, research has shown how people can experience ownership over virtual bodies of a different gender (Seinfeld et al., 2018), or how the illusion can alter the perception of peripersonal space (Noel et al., 2015). Such changes do not only change internal body representations, but can affect behavioural and physiological responses as well, and have led to applications for improved mental health (Matamala-Gomez et al., 2021). An interesting feature of the full-body illusion is that it creates a unitary full-body experience, while the illusion is induced by (visuotactile) stimulation of only a particular body part. This has stimulated new behavioural and neurological research on the open questions of the underlying (brain) mechanisms for multisensory perception and integration (Ehrsson, 2019).

Research with virtual humans in XR can be distinguished by research on the representation of individuals and their appearance realism, or on the nature of interpersonal interaction and their behavioural realism (Carter & Pollick, 2014). The representation of individuals in XR can vary from minimal displays to fully realistic 3D stimuli including detailed visual appearance, natural movements, and facial expressions. Interpersonal interactions can vary from passive witnessing to more complex social situations (Neyret et al., 2020). Research has shown how people often react similarly to virtual humans as to real people, yet also stressed the importance of context (de Borst & de Gelder, 2015). The perceptual or behavioural realism of virtual humans is often discussed with the uncanny valley hypothesis, suggesting that increased realism of human replicas such as virtual humans reaches a point at which they cause a sense of repulsion in the person viewing (or interacting) with it (Carter & Pollick, 2014; Mori et al., 2012).

Virtual humans can be human or algorithmically controlled as respectively avatars or agents, or as a combination of both (T. K. Metzinger, 2018). The perception and modelling of virtual humans is an active field of research, given their flexible and controllable nature as experimental stimuli. Agents that integrate the behavioural models from the theory of coordination dynamics together with insights from the embodied (music) cognition framework could be of great help for research in the experiential and behavioural processes underlying successful interpersonal coordination (Tarr et al., 2018). While some work has been done placing these virtual humans in musical contexts, as an avatar part of a drumming circle (Kilteni, Bergstrom, et al., 2013), or as a virtual conductor in a 2D display (Reidsma et al., 2014), much work can still be done to increase the expressive capabilities of avatars and agents to allow the emergence of creative musical partnerships (A. R. Brown et al., 2016). Chapter 3 and 4 aim to set steps in this direction by offering a methodological framework capable of evaluating the expressive dynamics in music making, and applying it to both avatar-agent and avatar-avatar musical interactions.

1.5 Extended musical spaces

Having introduced the theoretical frameworks and XR environments allowing for embodied interactions while inducing presence through immersion, this section focuses on the music making taking place within the “extended musical spaces” enabled by them. Music making is considered as the active and creative process that allows for the emergence and experience of meaningful relationships between its participants (Small, 1999). The “extended musical space” is seen as a mixture of a physical, virtual, personal, and social space, co-constructed and shaped by its participants, musical content, and context (Benford & Giannachi, 2011). This musical space is perceived, lived, and produced in an embodied practice (Lefebvre, 2012; Määttänen, 2007), and welcomes the shift from “music-in-a-place” to “music-as-a-place” (Lennox, 2009). Musical spaces for experimentation, performance, composition, and narration will be presented in subsection 1.5.1. While such musical spaces are obviously not new, it is the integration of XR technology that is, and which leads to the “extended musical space”. The “extended musical space” and “musical space” concepts will be used interchangeably from hereon and refer to the former.

Key contributions of XR technology to these musical spaces are its capabilities for immersion, multimodal integration, and embodied networked interaction from a 1st person perspective. These technological developments have led to a reconsideration of the artistic-scientific practice, now required to deal with the modalities of the physical and virtual, the techno-cultural context, new questions of form, content, and performer-audience roles (Catricalà & Eugeni, 2020; Lichty, 2014). The new media artist or engineer, now required for the construction of these musical spaces, becomes organizer and operator (Zabel, 2014), and deals with the XR technological layer and an ultra-mobile spectator endowed with agency (Gallese, 2020).

The physical and virtual dimensions of the extended musical spaces are enabled by a technological layer, informed by the earlier introduced theoretical foundation in section 1.3 (see Figure 1-1). The technological layer includes the technologies to capture, mediate, and display the fine-grained spatiotemporal coordination processes inherent to music making. Capture technology such as motion capture, eye-tracking, and physiological sensors are used to detect bodily movements and emotional markers. Mediation technology in the form of (virtual) musical instruments and hand-held XR controllers are used to transfer expressive cues and gesture trajectories from source to elsewhere (Pysiewicz & Weinzierl, 2016; Serafin et al., 2016). Display technology allows auditory and visual rendering through head-mounted displays and multi-channel speaker setups (Reyes-Lecuona et al., 2022). These technological components are integrated using software, varying from game-engines to lower-level protocols and middleware. Game-engines play a key role in any XR environment, acting as spatiotemporal administrators of multimodal data streams while offering (virtual) affordances. Lower-level protocols and middleware allow translation between the analogue and digital, physical and virtual worlds. Managing and controlling this XR technological layer requires significant expertise, not in the least to deal with the constraints in latency, bandwidth, and computing power (Ruan & Xie, 2021). The specific technological systems used in the studies presented below are listed in section 1.7.

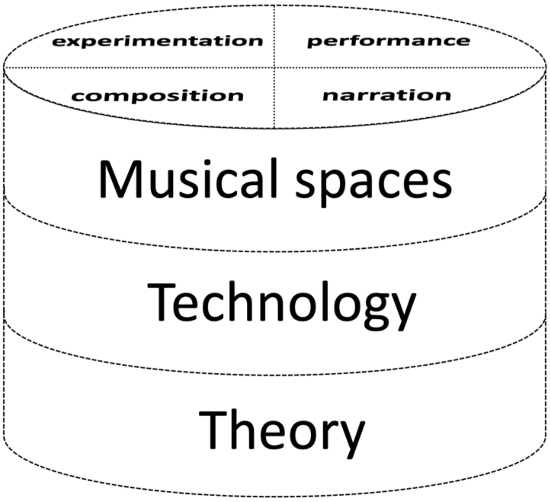


Figure 1-1: Musical spaces in this thesis are enabled by a technology layer and informed by a theoretical foundation

1.5.1 Experimentation, performance, composition, and narration

This thesis categorizes musical spaces as spaces for experimentation, performance, composition, and narration. They are presented below and discriminated using the musical roles and music making goals embedded within them. While most of this thesis will present work in the musical space for experimentation, studies in the other spaces that emerged as natural extensions will be presented in chapter 7. These studies are relevant as they allow for the realisation of experimental insights in concrete, public settings and in their turn can provoke new questions to be experimented on. Musical spaces are “entangled”, with development of the XR environments for music making in one space naturally leading to investigations, development, and insights in another.

Experimentation: The musical space for experimentation originates from a predefined research question and includes an experimenter capturing and observing the musical dynamics portrayed by a carefully selected sample population. Once the experimental context, task, and protocol have been defined, the experimenter only takes a passive role during the actual music making. There is a clear focus on creating the conditions required for answering a research question while reducing and constraining experimental bias. Experimentation in this space allows the underlying XR technology and theoretical frameworks to be questioned and improved. For instance, music making with algorithmically-controlled virtual humans in the musical space would allow improvement of the underlying coordination dynamics models (Tognoli et al., 2020). Simulating musical scenarios would allows investigation of the impact of auditory context (Bergstrom et al., 2017) and audience engagement (Glowinski et al., 2015) on performance quality, or comparison of their effectiveness to real-life situations (Williamon et al., 2014). Networked, virtual musical interactions could be manipulated using controlled delays to analyse sensorimotor coordinative processes of anticipation and adaption (Harry & Keller, 2019) or leader-follower roles (Konvalinka et al., 2010). Musical scenarios requiring cooperative or competitive joint-actions between virtual humans and agents could be designed to learn more about the embodied coregulatory processes, subjective experiences, and prosocial effects (Tarr et al., 2018). Most of the work presented in this thesis investigates the potential of XR technology in enabling these musical spaces for experimentation, and offers a methodological framework to better understand the musical dynamics within it.

Performance: Another musical space is the space for performance. It retains the traditional performer-audience dichotomy to some extent, although XR technology naturally affords a more participatory role to the audience (Mazzanti et al., 2014). Audience and performers do not have to be physically co-located, with networked, virtual environments creating a sense of shared space at the cost of local presence (Hamilton et al., 2011). The musical spaces for performance require renewed consideration of their scenography (Berthaut et al., 2015; J. Spence et al., 2013), can include (bio)feedback (Selfridge & Barthet, 2019), as well as new instruments, affordances, and sound-producing gestures (Berthaut et al., 2011; Serafin et al., 2016). Time no longer needs to be linear, large-scale "live" performances broadcasted over networks are buffered anyhow[1], with spectators potentially choosing their spatiotemporal vantage point and experiential trajectory (Benford & Giannachi, 2011; Thielen et al., 2018). The musical spaces for performance do retain an important focus on the "live" element to create feelings of togetherness (Onderdijk et al., 2021), as well as on the fertile, flexible, and negotiable conditions for musical play, interpretation, and improvisation (Borgo, 2005; Impett et al., 2015; A. E. Walton et al., 2018).

Composition: Musical spaces for composition are lived and practiced primarily by a designer, orchestrator, or composer structuring and evaluating the XR environment, its compositional tools, and the resulting musical composition. XR technology and its ability to create and work with (virtual) space through virtual embodiment and immersive contexts, naturally lends itself to work on the spatial aspects of musical composition, such as in spatial music (Pitman, 2021). Spatial music considers the spatiality of sound as a primary musical parameter, important for emotional expression, creative narratives, and imaginative thoughts (Hagan, 2008; Schmele & Finney, 2011). This interest might not be

---

[1] An extreme example is the live performance of Travis Scott on the Fortnite platform attended by more than 12.5 million people split across 250.000 concert instances (Ball, 2022)

surprising, given the intrinsic link between music and our spatial perception and cognition (Kendall, 2010; Schäfer et al., 2013; Tajadura-Jiménez et al., 2011). Various instruments for spatial music composition have been created (for a review, see (Pysiewicz & Weinzierl, 2016)), often valuing expressive gestures for their creative potential and their link between action and abstract thought (Deacon et al., 2019; Slepian & Ambady, 2012). XR technology has been used to devise spatial music composition tools (Santini, 2019), taking advantage of its capabilities for body movement tracking and 3D data display. This 3D view on the musical work can enrich the compositional process, as it allows full benefit of human abilities such as spatial reasoning and pattern recognition (for a collection of XR composition tools, see (Turchet et al., 2021)). Networked, shared, virtual spaces for composition can lead to collaborative creativity (Men & Bryan-Kinns, 2019), or can blur the lines between performance and composition, when acting and moving in the virtual environment is considered an experience of both (Ciciliani, 2020).

Narration: XR technology has a fundamental impact on narration, furthering the tradition of transmedia or immersive storytelling (Bruni et al., 2022). XR with its specific visual grammar and capabilities for immersion, virtual embodiment, and affordances has allowed new forms of narration to arise, and a new form of storytelling based on non-linear "micro-narratives” (Spampinato & Carticalà, 2021). The user, guest, character, or player shifts from a passive observer to an active discoverer, answering its personal questions of the why, where, what, and who in the story. This has led some to consider the narratives less as “stories” and more as “spaces” (Bruni et al., 2022; Jenkins, 2004), in which the design of the facilitating XR environments requires a balance between interactivity of the medium and control over the narrative’s coherence (Bruni et al., 2022). Sound, with its ability to offer cues outside our field of view, can play an important role in shaping and guiding the experience (Çamcı, 2019). Music composed with synthesized or environmental sounds can evoke mental imagery to real or imaginary spaces (Zelli, 2009) and thus strongly impact the narration experience. This potent combination of XR technology, sound, music, and narration could thus lead to new, rich applications in areas such as cultural experience (C. P. Martin et al., 2020), documentaries (Trommer, 2020), and audio games (Nakagawa et al., 2018).

These musical space categories are neither exhaustive nor mutually exclusive. For instance, musical spaces for learning (Orman et al., 2017; Waddell et al., 2019) or well-being (Bissonnette et al., 2016; Daffern et al., 2019) are not treated here. The spaces themselves share characteristics such as being networked, social spaces, enabled by an interdisciplinary approach, and are informed by the sciences, arts, design, and engineering fields. They all profit from the XR enabled affordances and XR’s capability for immersion and virtual embodiment. Certain aspects such as (virtual) music instruments can serve as experimental controls, performance vehicles, compositional tools, or narrative devices depending on context. The actual realisation of each space often includes elements of other categories, such as when performers and audience co-construct a story or a composition (Wozniewski et al., 2018).

Musical spaces discussed in this thesis are laid out in Figure 1-2. Chapters 2 to 5 present self-contained studies that are fully and originally (as published or submitted) included in this thesis. Chapter 6 contains the abstracts of four experimental studies that use XR technology in experimental music research. The author of this dissertation contributed the technical design and implementation to these studies. One of these studies has led to a published paper, two led to submitted manuscripts, and one is still in the data-collection and analysis phase (see the Research Output listed at the beginning of this thesis). These chapters are all rooted in a space for experimentation and originate from a scientific drive in using XR technology for an improved understanding of music making. The other spaces of performance, composition, and narration will be illustrated by artistic studies of XR technology for music making in sections 7.2, 7.3, and 7.4. The author of this dissertation contributed the design and technical implementation to these studies. The next section will introduce the research questions belonging to each chapter and artistic study.

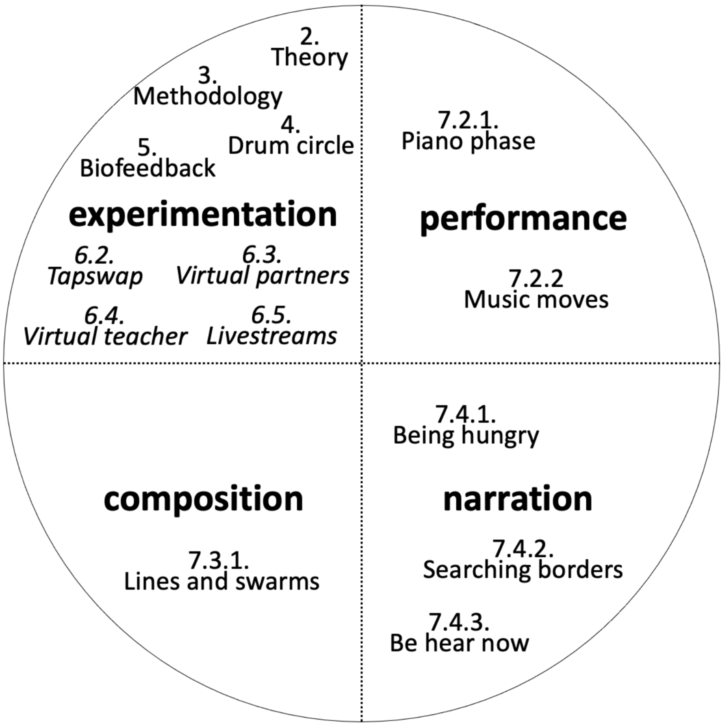


Figure 1-2: Studies presented in this thesis and their chapter, section, and subsections numbers mapped out over the different musical spaces

1.6 Research questions

Having introduced the philosophical and theoretical foundations on which this thesis builds, together with the main characteristics of XR and the musical spaces that are enabled by it, it is now time to frame the research questions that will be addressed in the chapters to come. The table below introduces the overarching main question and its translation into scientific and artistic goals. These are followed by specific research questions that will be addressed in every subsequent chapter.

Main question: How can XR technology contribute to the development of musical environments that enrich embodied and collective music making? This question highlights the central place given to XR technology in this thesis, which will be used to develop musical environments, and enable the musical spaces presented in the previous section. The presented studies will operationalize XR technology in concrete musical settings, offer methodological tools to perform experimental music research, and highlight its challenges and opportunities. Most of the work will take place in the musical space for experimentation, has a scientific focus, and profits from XR's capability of bridging experimental control in the lab (internal validity) with ecological value from simulated real-life settings (external validity) (Kothgassner & Felnhofer, 2020).	
Scientific goal	XR is considered as an experimentally useful tool and surrogate, capable of accommodating the dynamical and embodied interaction processes in music making. Given this potential of XR for experimental research, the scientific goal of this thesis is: → Explore XR's potential for an improved understanding of dynamical interaction processes and subjective experiences in music making
Artistic goal	Development of the XR environments for experimental music research naturally led to more artistic oriented studies taking advantage of the design and insights gained from the musical space for experimentation. These studies took advantage of the (virtual) affordances offered by the XR environments, raising new questions about XR's potential for expressivity and creativity. The artistic goal of this thesis is: → Explore XR's potential for novel expressions and experiences in music making

Table 1-1: Main research question of this thesis

Chapter 2, 3, and 4 can be considered together as they build up from theory, to methodology, to experimental inquiry. To design and build the XR musical environments that are enriching for embodied and collective music making, one first needs to be able to understand the sense-making processes underlying it. Chapter 2 aims to offer a theoretical paradigm capable of doing so, by combining the explanatory power of the embodied music cognition framework (Leman, 2007) with the terminology and quantitative tools of coordination dynamics theory (Demos et al., 2014; Kelso, 2009; A. E. Walton et al., 2015). It starts from the following high-level research question.

Question chapter 2: how to improve our understanding of musical sense-making?

Equipped with the argument that embodied music cognition and coordination dynamics are explanatory and constitutive elements of musical sense-making, one needs to have the proper (technological) means to incorporate and evaluate them. XR technology is ideally suited for this task with its capability for virtual embodiment (Slater et al., 2010) and the integration of algorithmically-controlled virtual humans or agents (Tarr et al., 2018; Tognoli et al., 2020). If the interaction with an embodied agent can establish social presence (the feeling of interacting with another person (K. M. Lee, 2004)), one could consider the embodied music cognition and coordination dynamics requirement for musical sense-making fulfilled. Therefore, chapter 3 develops a methodological framework to answer the following question.

Question chapter 3: how can we evaluate social presence in musical interactions in XR?

The methodological framework developed in chapter 3 evaluates social presence in three interdependent layers: performance quality, embodied coregulation, and subjective experience. Given the potential influence of visual and auditory coupling on musical interactions (Chauvigné et al., 2019; D'Amario et al., 2018; Dotov et al., 2021), the question remains to what extent XR technology impacts music making by mediating musical expressivity and communication. Chapter 4 thus experimentally manipulates audiovisual coupling in an XR environment to answer the following question.

Question chapter 4: to what extent can XR technology facilitate adequate performance quality, embodied coregulation and subjective experiences in musical interactions?

Next to viewing XR technology as an experimental tool, simulating performance environments, and capturing the music making dynamics, it can be viewed as a tool for biofeedback. Biofeedback in XR could be administered in specific interactions such as with the algorithmically-controlled virtual humans in chapter 3, as well as through changes in the audiovisual, immersive environment (Naef et al., 2022). By immersing its users in a virtual, multimodal environment, and exchanging experience of the real world for virtual simulations, it has shown capable of inducing states of relaxation (Blum et al., 2019; Naef et al., 2022), and might be able to induce lucid awareness (Kitson et al., 2019). Combining insights from the coordination dynamics theory with the immersive, auditory component of XR could help to develop biofeedback applications with a natural, realistic experience for its users. Chapter 5 attempts to take first steps in that direction by answering the following question.

Question chapter 5: can we use a dynamic auditory environment to manipulate human behaviour and experience?

Performing music research with XR requires an interdisciplinary approach dealing with technological constraints alongside fundamental research questions. These research questions originate from a deep expertise in knowledge gaps, which can be answered in experimental designs that are co-determined by technological capabilities. Technological innovations that open pathways for observation and analysis in their turn can provoke reflections and highlight other knowledge gaps. Given that XR has been considered as a paradigm shift in experimental research (Kothgassner & Felnhofer, 2020; T. K. Metzinger, 2018), chapter 6 aims to frame consequences for action-oriented music research by answering the following question.

Question chapter 6: what are the opportunities for XR technology in experimental music research?

Pursuing the artistic goal formulated at the beginning of this section, chapter 7 is devoted to artistic studies aimed at exploring and leveraging the opportunities of XR technology for musical expression and experience. Given the scientific insights originating from virtual embodiment, affordances, networked music making, immersion, and presence in XR, the question remains whether if, and how, they can be put to artistic use and shared in public contexts. The starting point of these studies is the following question.

Question chapter 7: what are the opportunities of XR technology for artistic musical expression and experience?

Chapters 3 and 4 respectively took the minimalist piece Piano Phase by composer Steve Reich and a polyrhythmic drum pattern as musical scenarios to investigate embodied music making dynamics in XR. This experimental music research was done in highly controlled lab conditions and thus raises questions about their ecological value. In addition, XR technology allowed the introduction of virtual immersion, human- and algorithmically-controlled virtual humans, and a virtual music instrument in the musical interaction. Since these XR enabled aspects were introduced in a musical space for experimentation, it remains to be seen if they remain capable of facilitating music making when transferred to a public performance context. Section 7.2 thus aims to answer the following question.

Question 7.1: Can we transfer a scientific musical scenario in XR to a public performance context?

As XR offers its users a virtual environment rich with affordances, in which to spatially navigate using a first-person perspective, it offers a direct analogy for our dealing, composing or constructing, with space. Moving spatial composition from the 2D interface to the 3D immersive world (Santini, 2019), and integrating scientific knowledge from the embodied music cognition and human-computer interaction fields, could offer tremendous benefits for creative activities in the musical spaces for composition (for an overview, see (Turchet et al., 2021)). Section 7.3 aims to develop new affordances using

XR technology for the development of such a space and attempts to answer the following question.

Question 7.2: What are the opportunities of XR for spatial audio composition?

Finally, XR has a large impact on narration through the creation of a new multimodal grammar with narrative tools such as micro-narratives (Spampinato & Carticalà, 2021), perspective taking, and the switching between 1st to 3rd person perspective (Bruni et al., 2022). Audio in XR has been systematically underrepresented (Reyes-Lecuona et al., 2022), and while there have been voices aimed at highlighting this gap (Çamcı & Hamilton, 2020), narrations in XR could benefit from a closer look on the role of music and sound. Section 7.4 aims to do just this by the design and development of musical spaces for narration to answer the following question.

Question 7.3: What are the opportunities of XR for musical narrations?

With every chapter now introduced based on its research question, the next section will present the overarching methodology before delving into the studies themselves.

1.7 Methodology

The work presented in this thesis can be distinguished based on the dual scientific-artistic question presented in the previous section, which is on the one hand to better understand, and on the other to enrich music making in XR. Chapters 2 to 6 will present more scientific oriented studies, while chapter 7 will focus on more artistic oriented studies. The research for this thesis originated in theoretical work (chapter 2), leading to a methodological framework and operationalisation in a case-study (chapter 3). This framework was then applied to an experimental study of musical interaction in XR (chapter 4). A distance-based, experimental study taken amidst social restrictions imposed by the Belgian government during the COVID-19 is presented in chapter 5 and investigates auditory immersion and adaptive biofeedback. This is followed by a selection of fundamental and applied experimental research into music making in chapter 6. As outlined in section 1.5, these studies are situated in the musical space for experimentation, with studies situated in the musical space for performance, composition, and narration presented in chapter 7.

The studies are interdisciplinary, with research questions originating in different research fields (see Table 1-2), and action-oriented, searching for knowledge and expressive potentialities in music-making through a "living culture of interaction" (Lesaffre & Leman, 2020). XR technology is used to deal with the balance between experimental control and ecological realism, building on its capacity to simulate immersive musical contexts and obtain responses comparable to those in the real world, all whilst maintaining experimental rigor and standardization (Blascovich et al., 2002; Kothgassner & Felnhofer, 2020).

XR technology is used as a means for immersion in novel musical contexts. When immersion is listed as XR stimuli in Table 1-2, it refers to the capacity of XR to vary levels of immersion, such as when watching a 360 versus a 2D video, watching a performance in a head-mounted display (HMD) versus on a 2D screen, or when switching between AR and VR in a MR application. Next to its capacity for immersion, XR technology was used to create and offer dynamic, interactive, multimodal, or otherwise impossible stimuli, such as body-illusions, in the experimental studies (see Table 1-2). Dynamic, virtual musicians (avatars and agents), instruments, and sonification through spatial, binaural audio were used as (bio)feedback in musical scenarios and rhythm-based interaction paradigms to stimulate, challenge, or enrich creative musical expression and interaction.

1.7.1 Materials and methods

A combination of quantitative and qualitative methods was used to capture the music making dynamics as spatiotemporal coordination patterns and indicators of subjective experiences in the XR technological layer. This layer was briefly introduced in section 1.5 by distinguishing between technologies to capture, mediate, or display music making dynamics. The following paragraphs will give a summary of the specific technologies used in the studies presented in the following chapters. For an overview of the main capture and display technologies used in specific studies, the reader is referred to Table 1-3. See Appendix A: for a list of technological components and schematics of all XR implementations with associated latencies.

Capture: This technology component aimed to capture the spatiotemporal timeseries of the performance output, embodied coregulation, and subjective experience. Performance output was captured as note and breathing onsets from MIDI keyboards, pressure sensors connected to Teensy 3.2 microcontrollers, or computer keyboard presses. Embodied coregulatory timeseries resulted from body movement captured through the Qualisys motion capture system or from gestural XR controllers from the Oculus Quest HMD. These timeseries were then further processed to obtain postural sway or head movement timeseries. Video data from musical interactions were captured and annotated to look for patterns in communicative cues. Low-latency Logitech Brio cameras were used in the Tapswap experimental study. Eye-gaze and pupil data were captured through eye-tracking sensors integrated in the HTC Vive Pro Eye HMD. Finally, questionnaires were taken from participants to probe for experiential aspects through indicators of flow (Engeser & Rheinberg, 2008), (shared) agency, presence (Makransky et al., 2017), and inclusion of other in self (Aron et al., 1992). Two studies in chapter 6 also captured brain dynamics through electroencephalogram (EEG)-recordings but since they were not implemented nor supervised by the author of this thesis, they will not be discussed any further.

Mediate: Mediation of music making dynamics refers to the transport of expressive cues and gestures from source to recipient, the rendering of captured and predesigned stimuli in (immersive) displays, and the coupling of performers through (virtual) music instruments. Transport of data used the cabled and wireless Ethernet infrastructure from the Art and Science Interaction Lab[2] (ASIL) with extensive use of the Dante, Open Sound Control (OSC), and MIDI protocols to send and receive audio and specific control messages. Synchronization of data streams used 120Hz timecodes distributed using a master clock also integrated in ASIL's infrastructure. Latency and bandwidth requirements were experientially tested and measured prior to experimental studies in pilots. Hybrid instruments were used in several studies such as a virtual piano aligned to a physical piano keyboard, physical drum pads that interacted with a virtual drum circle and rotating spheres, and the use of the Oculus Quest controller to draw, enrich, and manipulate visual trajectories in an XR environment.

Display: Display technologies rendered auditory or visual immersive environments or a combination of both. Visual immersive displays consisted of HMDs such as the HTC Vive Pro Eye, the Oculus Quest 1 and 2, and the Microsoft Hololens 2. A smartphone cardboard was used as HMD in the Livestream study together with streamed live 2D and 360 video of performances through respectively Zoom and YouTube. A large 2D projection screen of 7x4m in the ASIL lab was used in several artistic studies. Spatial audio was rendered as 3rd and 6th order Ambisonics on the 80-speaker system in ASIL, or as binaural audio using software plugins (see software below). Finally, a Mozilla Hubs environment[3] was used as a virtual, social space for public demonstrations of the Being Hungry study reported in section 7.4.1.

Software: The game-engine Unity was used to offer virtual affordances, capture interactions, and render visual content and virtual humans in the HMDs. The digital audio-workstation Ableton was used to capture and render audio, as well as to record MIDI data from physical instruments. The visual programming environment Max MSP was used to develop user interfaces in Ableton to control experimental protocols and render Ambisonic audio using the Envelop4Live[4] and ICST plugins (Schacher & Kocher, 2006). Communication between Ableton and Unity was done using the OSC protocol. The Searching Borders study reported in section 7.4.2 also used the Touchdesigner visual

---

[2] https://asil.ugent.be/

[3] https://hubs.mozilla.com

[4] https://github.com/EnvelopSound/EnvelopForLive

programming environment to render real-time spectrograms, and to send them to Unity using the Spout protocol.

Analysis: Data pre-processing was done using Python. Data analysis was done using Python, Matlab, and R. Python was mainly used to calculate phase and tempo timeseries from onset data. Matlab was used to calculate wavelet coherence analyses on body movement data. Finally, R was used to perform statistics with extensive use of the lme-package (Bates et al., 2015) for linear mixed-effect modelling.

Studies	Chapter	Type	Field	XR stimuli
Theory	2	E	embodied music cognition	NA
Methodology	3	E	embodied music cognition	avatar, agent
Drum circle	4	E	embodied music cognition	avatar
Biofeedback	5	E	well-being	binaural audio
Tapswap	6.2	E	neuropsychology	body-illusion
Virtual partners	6.3	E	neuropsychology	avatar, agent
Virtual teacher	6.4	E	education	agent
Livestreams	6.5	E	cognitive psychology	immersion
Piano phase	7.2.1	P	embodied music cognition	agent, immersion
Music moves	7.2.2	P	embodied music cognition	avatar, virtual instrument, immersion
Lines and swarms	7.3.1	C	human-computer interaction	virtual instrument
Being hungry	7.4.1	N	digital arts	spatial audio, immersion
Searching borders	7.4.2	N	digital arts	body-illusion, haptic-visual sound, immersion
Be hear now	7.4.3	N	digital arts	spatial audio, generative visuals

Table 1-2: A list of the studies present in this thesis, their main research field, and use of XR stimuli (E = Experimentation, P = Performance, C = Composition, N = Narration)

Studies	**Capture**				**Display**		
	Onsets	Body movement	Eye-tracking	Questionnaires	HMD	Screen	Spatial audio
Theory	NA	NA	NA	NA	NA	NA	NA
Methodology	X	X	X	X	X		
Drum circle	X	X		X	X		
Biofeedback	X			X			X
Tapswap	X			X	X		
Virtual partners		X	X	X	X		
Virtual teacher		X		X	X	X	
Livestreams				X	X	X	
Piano phase	X	X			X	X	
Music moves	X	X			X	X	
Lines and swarms		X			X		X
Being hungry						X	X
Searching borders					X	X	
Be hear now						X	X

Table 1-3: A list of the studies present in this thesis distinguished by their use of specific capture and display XR technology components

2.

A theoretical paradigm

Published as: Van Kerrebroeck, B., Rosso, M., & Maes, P. J. (2020). Linking embodied coordination dynamics and subjective experiences in musical interactions: a renewed methodological paradigm. DOCUMENTA, 38(1), 38-60.

2.1 Introduction

’You are the music, while the music lasts’ (T.S. Eliot - Four Quartets, 1943)

Combining quantitative methods with qualitative methods or so-called mixed-method approaches can enable a deeper understanding of phenomena by gaining insights from multiple perspectives. However, we argue that their value is not found by merely searching for correlations in quantitative data and qualitative reports of subjective experiences. Instead, the central thesis of this paper is that this value is found by using appropriate procedures and routines that reveal interaction dynamics, meaningful constraints, relations, and influences using the common vocabulary and theoretical framework of coordination dynamics.

The coupling of the quantitative and the qualitative has historically been a hotly debated topic (Sale et al., 2002). From a pragmatic viewpoint, the discussion here is held while acknowledging the fact that research always occurs in a social, historical, political and cultural context (Bresler, 1995; Creswell, 2009) and stresses the fact that “reality” is an ongoing, dynamical and meaningful transaction between environment, mind and sense perception (Barone, 2000; Horne et al., 1998). This discourse is situated in the domain of digital humanities and aims to extend the recently coined “humanities 3.0” concept (Bod, 2013). As such, it acknowledges the value of using digital tools to discover patterns (humanities 2.0) alongside hermeneutic and critical approaches (humanities 1.0) but argues that one should go beyond mere patterns and focus on the underlying principles that create them.

Both qualitative and quantitative methods are valuable and necessary approaches to gain knowledge. In any research domain, qualitative and quantitative approaches have shown their value as respectively inductive and deductive instruments (Creswell, 2009). Qualitative research allows for a deeper understanding and appreciation of phenomena while quantitative research provides a more precise analysis and prediction with the goal of generalisation (Razafsha et al., 2012). In addition, quantitative methods feature a larger distance between researcher and research, exchanging meaning for a higher level of abstraction. However, both are systematic attempts to examine concepts (Bodie & Fitch-Hauser, 2010; Razafsha et al., 2012). Simply put, quantitative methods are particularly powerful in discovering patterns in occurring phenomena (regularities, inter-dependencies, trends), while qualitative methods are particularly well suited for interpreting the found patterns. The combination of quantitative and qualitative data provides additional insights and is now widely applied in research using mixed-method paradigms (Creswell, 2009). However, as both methodologies study different phenomena, combinations should remain complementary and should not be used for cross-validation (Sale et al., 2002). We assume this mixed-method approach here complemented by the vocabulary and features of dynamic systems theory. From the latter we use non-linear quantitative methods for pattern discovery and argue for the use of qualitative methods for their interpretation.

In the present article, we want to contribute to the debate on human interaction with music. In accordance with the embodied music cognition (EMC) paradigm that will be introduced in section 2, we approach musical meaning as an active process: a lived experience, created in-the-moment of people's interaction with music and situated in a specific socio-cultural context and personal "histories" of experiences (Cook, 2013; Leman, 2007). The human body, its motor and (neuro)physiological functioning are thereby attributed a central role. The core assumption is that the observation of the body, and its functioning, may provide access to the subjective realm of musical experience, feelings, and sense-making. In that regard, observable patterns of bodily activity and the subjectively felt quality of that interaction are essentially coupled, making the integrated use of quantitative and qualitative methods necessary. Supported by new technologies for measuring bodily activity (movement, physiology and brain activation) and computational analysis methods, empirical research has profoundly ameliorated our knowledge on the embodied basis of human music interaction (Lesaffre et al., 2017). Yet, some important challenges lie ahead.

A first challenge pertains to the enormous variability in the observed patterns of embodied music interaction, both in time and space, complicated further by the manifold of contextual and personal factors. Given this variability it is hard to interpret and generalise musical behaviours and experiences. Typically, the solution is found in a reductionist approach stripping away the lived experience of a real-life musical interaction to a greater or lesser extent. Yet, it's hardly ascertainable that variation, deviation, dynamical change and surprise are at the core of musical pleasure and sense-making. For this reason, this article approaches human beings and their musical environment as a complex adaptive system as seen from the dynamical systems paradigm. The core insight we adopt from the dynamical systems paradigm is that, instead of looking at the appearance of patterns in their manifold of variability, we need to focus on the understanding of the organisational principles that lead to the manifold of observable patterns of music interaction. It is here that quantitative methods can help in the detection and analysis of patterns while qualitative methods can help with their interpretation and reveal the relation to their underlying organisational principles. According to the paradigm, these principles are generic, regulating pattern formations in physical and biological systems nature-wide. This includes systems that involve human embodied interaction and the subjectively felt qualities of these interactions. What is especially interesting is that dynamics, variability and instability are at the core of these organisational principles, in allowing systems to behave flexibly, adapt to change and evolve towards qualitatively new forms of organisation and behaviour. Finally, the dynamical systems paradigm provides a valuable vocabulary, giving the opportunity to connect the languages of people involved in music interactions (interpreters), and music and cognitive (neuro)science (pattern finders). The methodological model incorporating these notions from dynamic systems and EMC forms the content of section 2.3. A selection of quantitative techniques used for pattern detection and analysis are discussed and introduced in subsection 2.3.1.

Secondly, we need to deepen our knowledge on the nature of the subjective experience of human interaction with and via music. Early attempts of investigation were rooted in the domain of phenomenology (Dura, 2006; Pike, 1967) that will be introduced alongside the EMC paradigm in section 2. The original objective of phenomenology is the systematic attempt to uncover and describe the internal meaning structures of a lived experience (Eberle, 2014). Qualitative methods seem best suited to undertake this task, given that the study of an experience is primarily approached from a first-person perspective (Gallagher, 2012; Randles, 2012; F. J. Smith, 1995). Nevertheless, the complementarity of seemingly decoupled opposites such as the quantitative and qualitative has historically been included in the phenomenological view. Phenomenology arose out of the need to bridge Cartesian dualism between objects “out there” and subjectivity “in here” (Kearney, 1995). A deeper elaboration on this view may foster the development of innovative methods for the integrated study of quantitative and qualitative aspects of human music interaction (Schäfer et al., 2013). To this end, more recent psychological and neuroscientific accounts on the quantitative study of subjective experience are discussed in section 2.3.2.

Finally, we end this paper with a section on four research experiences which incorporate the conceptual model from section 2.3 in their practice. While some of them are still on-going, they aim to illustrate how the quantitative and qualitative techniques from subsections 2.3.1 and 2.3.2 can be used in empirical research of dynamical music interactions.

2.2 Theoretical background

The goal of this paper is to present a model, implementing a renewed methodological paradigm to study dynamical musical interaction processes. An essential feature of this paradigm is the aim to link quantitative coordination *patterns* characterizing a musical interaction, to the subjectively felt *quality* of the interaction. The realisation of this paradigm relies on the combination of different theoretical frameworks. At the core lies the EMC theory that provides a global theory on the intricate relationship between bodily movement and subjective sense-making. As an extension, we propose to integrate the theory on coordination dynamics, to deal with the spatiotemporal variability and complexity inherent to musical interactions, by focusing on the generic structuring principles underlying musical interactions. Finally, we refer to the framework of phenomenology, as a means to integrate a first-person perspective to the experienced quality of a musical interaction, linked to the concepts of intentionality and agency.

Embodied music cognition: EMC is rooted in more general theories on embodied cognition and interaction (Anderson, 2003; Svare, 2006) and embodied forms of phenomenology (Merleau-Ponty, 2013). These theories have led to several complements and extensions such as in the enactive, extended, embedded, ecological, emotional, engaged, expressive and emergent approaches (Hutto & McGivern, 2015). Core concepts of all these approaches are the close action-perception coupling and the interaction with the environment.

The core idea of EMC is that an intentional level of musical interaction is established through corporeal articulations and imitations of sensed physical information provided by the musical environment (Leman, 2007). It emphasises the role of the human body as mediator for meaning formation and places it in an interconnected network of sensory, motor, affective, and cognitive systems involved in music perception. Subsequent accounts have extended the role of environmental and social contexts by emphasizing the importance of collaborative interaction and joint action (N. Moran, 2014). These contexts would enable a sense of participatory sense-making, creativity, meaning formation (De Jaegher & Di Paolo, 2007) and intense subjective experiences (Maes et al., 2018). Others have highlighted the overly dualistic nature of a “body as mediator” and in the distinction between encoding and decoding, nature and culture (Geeves & Sutton, 2015).

The EMC framework is valuable in the way it connects subjective experience and sense-making to situated bodily activity. However, empirical research has generally been struggling to reliably capture the complexities and variability, both in time and space, inherent to embodied musical interactions. A solution that gains increasing impact is to extend the EMC framework with a more dynamical account of music interactions. Within our proposed methodological paradigm, we integrate the interdisciplinary framework of coordination dynamics, originating in the work of JA Scott Kelso, to better capture the complexities and spatiotemporal variability in embodied music interactions.

Coordination dynamics: Coordination dynamics is a theoretical, methodological and analytical framework that aims to understand how patterns of coordinated behaviour emerge, persist, and evolve in living things (Kelso, 2009). This line of scientific inquiry focuses especially on time-varying coordination processes in the human brain and behaviour, making it particularly applicable to the study of embodiment in dynamical musical interactions (Borgo, 2005; Demos et al., 2011, 2019; Maes, 2016; Varni et al., 2012; A. E. Walton et al., 2018). Musical interactions, whether in performance or dance, require an intricate, fine-tuned spatiotemporal coordination of a large number of coupled body parts (of one or more individuals) to reach a coherent and pleasant performance. According to the coordination dynamics approach, understanding the temporal dynamics of such musical interactions is to study the *relationships* that exist between the individual bodies and body parts, rather than studying them each on their own. In other words, bodies and body parts should be studied as a collective system and correspondingly, the unit of analysis should be shifted to the system level. The central thesis thereby is that such a system will self-organise throughout time so that quasi-stable relationships and patterns are established between its interrelated parts. The field of coordination dynamics is thereby of interest for music research as it provides valuable analytical tools to quantify the time-varying relationships and patterns within complex, coordinated musical behaviour. In addition, coordination dynamics brings in a scientific vocabulary of concepts that is suited to match the vocabulary of musicians and dancers, describing their musical experience often in terms of a dynamic interplay of moments of relaxation and tension, balance and instability, complexity and simplicity, predictability and surprise.

Phenomenology: A third theoretical underpinning inherent to our proposed methodological paradigm for studying dynamical musical interaction is the framework of phenomenology. Phenomenology provides a steppingstone to the integration of a first-person perspective to the experienced quality of an embodied (musical) interaction (Gallagher, 2012; D. W. Smith, 2008). A shared importance between phenomenology and the approach presented here is given to the role of the body. Merleau-Ponty, for example, focused on the circular relationship between the objective and subjective dimensions of the body that enable a relation between the perceiving and the perceived (Halák, 2016). Concepts such as empathy and inter-subjectivity that are crucial in any musical interaction rely on this relationship (Duranti, 2010; Zahavi, 2001). Another relevant phenomenological concept is that of perceptual indeterminacy (Merleau-Ponty, 2013). It can be linked to the aforementioned positive traits of variability and instability of a dynamical system as it views the indeterminate as a positive phenomenon from which qualities can emerge. As such, it also shows a correspondence to the concept of emergence in dynamic systems (De Wolf & Holvoet, 2004) and meta-stability in coordination dynamics (Kelso, 2009). A concrete musical example is the emergence of “groove” out of small time-differences in music (Roholt, 2013). A phenomenological account by Casey illustrates the value of investigating the subjective experience in musical contexts. The musical experience is said to be direct, involving a “willing suspension of belief” through the practice of so-called “bracketing” as stated in phenomenological reductionism (Casey, 2010). Interactive musical scenarios as experimental settings could thus provide transparent investigations into the subjective experience by minimising noise from spurious mental processes and thoughts. Some neuroscientific accounts comment on the illusionary experience of a phenomenological unity in the musical experience when perceptual components such as pitch, rhythm, tempo, meter, contour, loudness, spatial location and timbre are processed separately (Levitin & Tirovolas, 2009; Niedenthal et al., 2005). These accounts stress the experiential aspect of music but nevertheless remain vague as to how this experience comes about. Below, we will discuss some possible methods to assess the subjectively experienced quality of musical interactions from a first-person perspective.

2.3 A renewed model for linking embodied coordination dynamics and subjective experiences

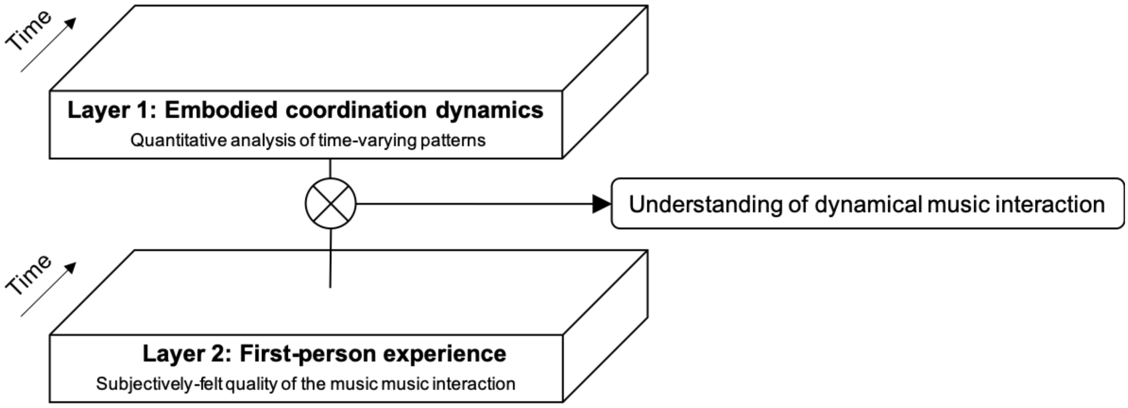


Figure 2-1: A conceptual model of the link between embodied coordination dynamics and the first-person experience

The three theoretical frameworks in Figure 2-1 are combined into a working model, proposed as a methodological paradigm for the empirical study of dynamical human music interactions. As mentioned, the model is rooted in the EMC theory in the sense that human experience and sense-making (layer 2) is inherently linked to bodily action and interaction (layer 1). Yet, the model proposes to extend this basis of the EMC theory in a twofold manner. As a first extension, the model attaches great importance to the time-varying nature of embodied music interactions and their subjective experiences. As a second extension, the model advocates for a systems perspective to the study of human music interaction. As explained above, embodied coordination and sense-making are understood as collective, participatory, and relational processes.

On the level of embodied coordination dynamics (layer 1), the challenge is to better understand how humans jointly construct patterns of order (articulated in bodily movement and sound), and how these patterns sustain, break down, and evolve towards new ordered patterns throughout time. On the level of the first-person experience (layer 2), the challenge is to better understand the affective quality of the joint relationship between interacting musicians, dancers, or listeners. This affective quality pertains to the communicating and negotiating of intentions, to expressivity, the feeling of togetherness, shared agency, and flow among other experiential aspects. Finally, the key challenge related to the model is the investigation if, and how, the time-varying patterns of embodied coordination relate to the affective quality experienced by the musicians, dancers, or listeners involved. This final challenge might include an analysis at multiple levels of observation, for example from body part to individual or group level, and lead to a verifiable formalisation of underlying organisational principles.

As a concept, it is more and more acknowledged within the cognitive and social sciences that the time dimension and (dynamical) systems approach are relevant in research on human interaction. So far however, empirical research in the domain of music found it difficult to reliably capture the time-varying processes in concrete scenarios of human music interaction. An important goal of the present paper is to briefly discuss existing methods that provide opportunities for empirical research to operationalize the ideas inherent to the model presented here.

2.3.1 Layer 1: Quantitative measurement and analysis of embodied coordination dynamics

With increasing technological innovation, researchers now have a wide-ranging choice of tools and sensors for capturing quantitative data. This wealth of possible data poses a considerable challenge to researchers to decide on the most relevant data and eliminate noise given the envisioned research questions. In the context of music, data may pertain to audio recordings, body movements, physiological data, note sequences and many more. In this section, we introduce five quantitative techniques that are well suited to unveil time-varying patterns and processes in sequential, non-stationary, and timeseries data. As such, they are proper candidates to operationalize layer 1 of our model proposed above. The discussed techniques allow us to not only unveil time-varying patterns within a single time series, but equally to unveil patterns across, and relationships between multiple data streams. This makes them particularly relevant to implement the systems approach, advocated for in our model. In addition, working with dynamic systems revalues outliers and individualised research as defined by (Holmes et al. 2013; Bresler 2006) through its capability of working with multiple resolutions. It allows the underlying simple principles to be uncovered while keeping in mind intrinsic dynamics and initial conditions. It tackles the challenge of too many degrees of freedom by focusing on lower-dimensional order parameters. All these challenges are characteristic of non-linear interactions. It is a goal of the methods below to reveal the recurrent patterns, underlying structure and more generally, understand the interactions between components.

Phase space reconstruction: Phase space reconstruction makes it possible to identify patterns and relations between non-recorded degrees of freedom. It refers to the process of obtaining the phase space of a dynamical system from its timeseries. A phase space represents the set of all possible states of a dynamical system such that each state of the system corresponds to a unique point in the state space. Using an influential theorem, one can reconstruct this space using a potentially lower-dimensional timeseries (Takens, 1981).

An example can be seen in Figure 2-2. On the left, it shows a one degree of freedom gyroscope recording of a simple movement with a smartphone. On the right, it shows its associated phase space reconstruction and 3-dimensional patterns corresponding to pitch, yaw, and roll.

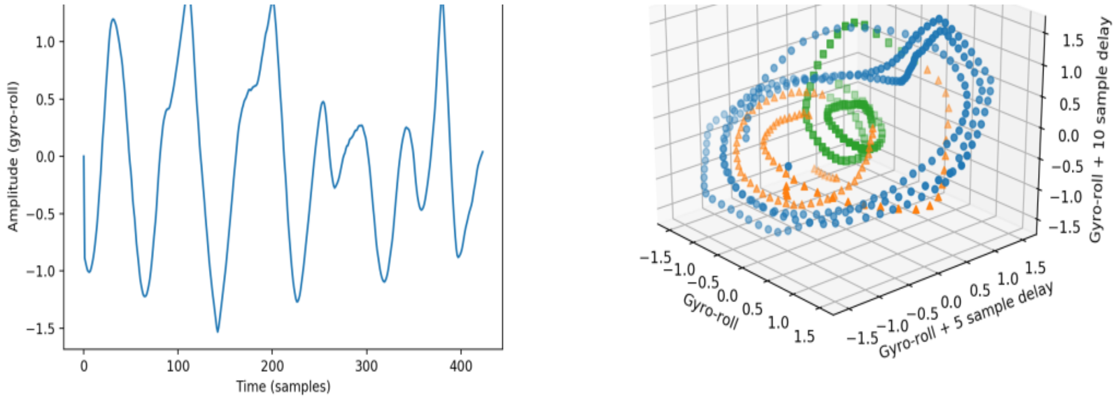


Figure 2-2: Gyroscope data recording of 1 degree of freedom (left) and its phase space reconstruction (right, with three cyclic patterns in different colours)

Recurrent Quantification Analysis: This technique reveals structure in complex timeseries. It takes the reconstructed phase trajectory of a dynamical system and counts the number of recurrences to any particular state. Its basis is a square distance matrix with recurrence elements evaluated using a cut-off limit called a recurrence plot. The advantage of this technique is that it does not require assumptions about data stationarity, data set size or distribution.

Figure 2-3 shows example recurrence plots with phase space trajectories on the x- and y-axes from conditions in which two subjects are instructed to tap their hand along with metronomes. The metronomes start in-phase but gradually de-phase due to different tempi. The left-most plot shows the trajectory of the metronomes' relative phase and shows a predictable system that linearly increases its collective variable. The plots in the middle and to the right respectively show the participants in isolation and looking at each other's hand moving. It shows how variability in human behaviour adds random fluctuations and structure at specific time- and phase-relationships.

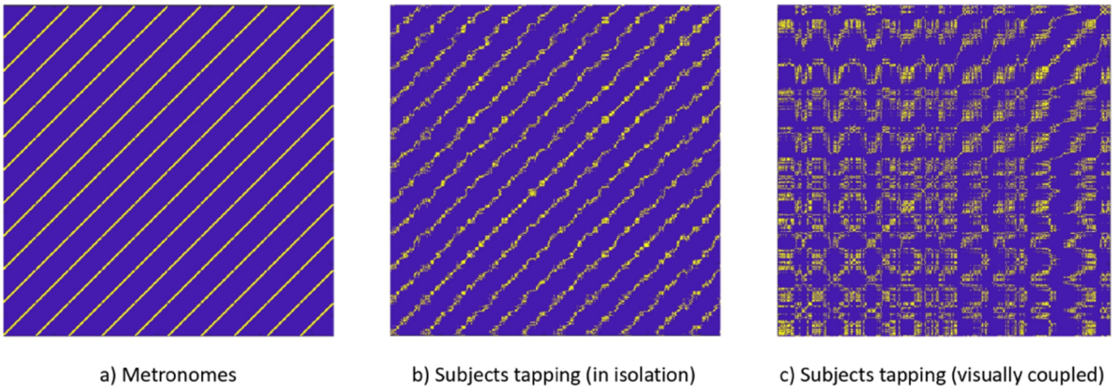


Figure 2-3: Recurrence Plots (RPs) computed from the time series of a system's collective variable (relative phase)

Fractal analysis: Fractal analysis represents a collection of contemporary methods that measure complexity. It is useful in measuring properties of systems that possess a degree of randomness and allows simplification and quantification of complex relationships over multiple spatiotemporal scales. For example, it has been used to show the ability of listeners to predict tempo fluctuations (Rankin & Large, 2009), to measure complexity in musical improvisation (Keller et al., 2011) and to describe gait dynamics synchronised to music (Hunt et al., 2014). An important measure is the fractal dimension which evaluates to what extent properties depend on the resolution at which they are measured. Another technique that is often used is called Detrended Fluctuation Analysis (Peng et al., 1994). It calculates a curve with an exponent that is an indicator of structure appearing at multiple scales (self-similarity), long-memory processes, 1/f noise, and power-law relationships.

Cross wavelet coherence: This technique is a useful to assess synchronisation between subjects. It evaluates coordination through examination of the strength (coherence) and patterning (relative phase) of two timeseries across multiple time scales. It can be applied to non-stationary data and is a form of spectral analysis for non-linear timeseries. It is able to reveal periodicities of local micro-scale structures within global macro-scale patterns (A. E. Walton et al., 2015). An example plot of the transform is shown in Figure 2-4 and was used to assess coordination using movement data of two players playing a shaker instrument. It indicates a higher degree of synchronisation for lower frequencies (0.125Hz or 8 second period) and a regular phase-lag (a quarter cycle or 2 seconds) between the two players indicated by the upward arrows. This observation can then be related back to the musical phrases and their interpretation by the musicians.

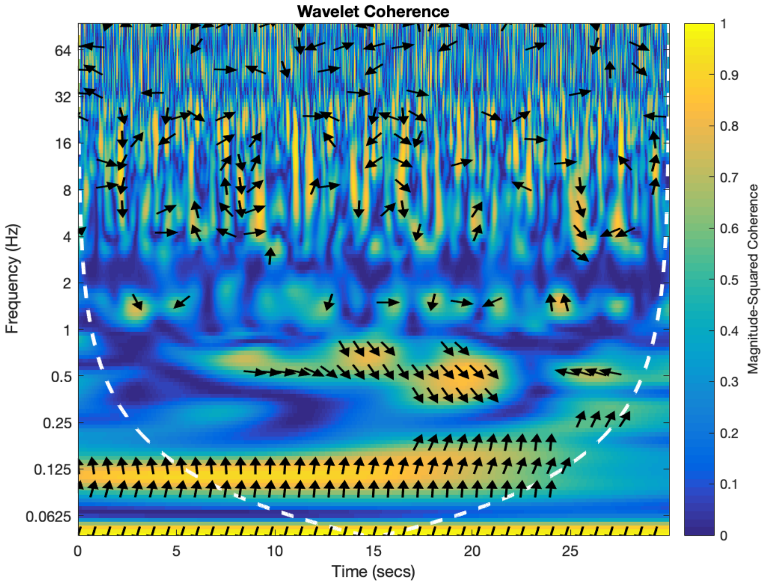


Figure 2-4: Cross wavelet coherence on movement data of two music players

Dynamic models: When relevant (control and order) parameters have been defined using empirical research, and an in-depth understanding of the non-linear coupling between components has been achieved, one can formulate a mathematical formulation of a dynamic model. They are useful for prediction and allow a deeper understanding of phenomena through their formalisation of a general organisational principle. They can be experimentally tested and are built upon simple dynamic models. To name a few, models exist for rhythm perception (Large, 2008), synchronisation (Mörtl et al., 2014) or joint-action (Demos et al., 2019).

The presented techniques obviously do not represent an exhaustive review of techniques for non-linear analysis of dynamic systems. Neural networks might be used to learn underlying dynamics without defining an explicit internal model beforehand. Other (linear) movement analysis methods based on principal components (Toiviainen et al., 2010), topological structure (Naveda & Leman, 2010), probabilistic (Sievers et al., 2013), sequential (Françoise et al., 2012) or functional (Caramiaux et al., 2015) models have shown their value but do not capture fine-grained spatiotemporal structures or non-stationary data well. Linear methods such as Fourier analysis, auto- and cross-correlation can provide helpful directions for subsequent analysis. A good evaluation of the use of both linear and non-linear methods is given in (Ravignani & Norton, 2017) in the context of measuring rhythm complexity.

2.3.2 Layer 2: Assessment of the subjectively experienced interaction quality

In this sub-section we intend to raise an important methodological problem in the study of subjective experience, namely its operationalization in experimental settings. How can we translate the subjective experience into observable variables, in order to measure some dimensions of the subjective experience without interfering with the experience itself?

Partial solutions to tackle such a big problem are already available for adoption. They are presented here as a non-exhaustive methodological overview with practical suggestions for empirical music research. At this layer, the researcher is presented with the methodological challenge of organically integrating the assessment of subjective experience and the measurements of dynamics recorded during musical interactions. As we stressed over the course of the present article, this integration is a condition necessary for a deeper understanding of the phenomenon.

Questionnaires and scales

The first point worth clarifying is the validity of verbal reports from the subjects as quantitative data, as we often rely on their content to measure dimensions of the first-person experience. Interrogating the subject about its own awareness is so far the most direct form of access to the experience: in-depth interviews, focus groups, questionnaires, concept mapping, focused life history narrative, audio/video/document analyses, documentary analysis and case studies can provide access to the domain of subjective experience (Razafsha et al., 2012).

Nevertheless, when it comes to measuring the quality of an ongoing interaction, verbal reports suffer from the intrinsic disadvantage of being mediated by self-referential cognitive processes which eventually lead to verbalization. This implies this sort of data cannot be collected without perturbing or interrupting the flow of the interaction. Hence, there is a need for the report to be referred to a posteriori with respect to the original experience. The solution is sub-optimal, as the subject would refer to a memory of the experience rather than to the experience itself. Furthermore, reports collected a posteriori are usually a summary that are difficult to relate to the time-varying nature of interactions.

Time-varying ratings

In order to facilitate the mediation between the participant and its experience at the time of the measurement, video-audio stimulated recall is a valid approach which has been proposed in music research (Caruso et al., 2016; Desmet et al., 2012). It consists of presenting to the participant a recording of his own performance, so that he or she can associate expressions and intentions to specific moments of the experimental session. An annotation system is then provided to continuously rate subjective parameters of interest over the course of the stimulation to generate a timeseries that can be related to the time-varying measures recorded from the interaction. A practical use-case for annotation will be presented in the next section.

We want to point out that the variable to measure should be carefully selected, since repeating the procedure several times can be tedious for the participants and compromise its own reliability. Since the approach provides measurements which are limited in richness and nuances of content, the researcher should ideally collect as much information as possible by complementary and less systematic means. Open questionnaires and interviews at the end of the session give the participant the opportunity to elaborate on some crucial moment of the experienced interaction, which is potentially a valuable source of information on attributed meanings.

Inference from physiological markers

Besides the above-mentioned methods and based on processes of intentional evaluation by the participant, disposing of a "toolkit" of sensors for measuring biomarkers can provide the researcher with access to some low-level dimensions of the subjective experience. For instance, by analysing the electroencephalogram (EEG) of a person presented with subliminal stimuli can be enough to know whether the stimuli were consciously perceived or not, without asking for any verbal report (Dehaene, 2014). Very far from being any sort of "mind reading", the approach consists of looking for physiological patterns of activations that work as a "signature" for relatively low-level dimensions of consciousness.

The set of tools for the detection of physiological markers spans over the central, peripheral and autonomic levels of the nervous system (Grewe et al., 2007; Steinbeis et al., 2006). Electromyography (P. Ekman, 1992; Tamietto et al., 2009), pupillometry (Laeng et al., 2012), electro-dermal activity, heart-rate and blood pressure among others (Critchley, 2002), have widely proven to provide valid signatures of some components of subjective experience. Over the years, all of these peripheral measures have been correlated to regional brain activity in order to shed light on the hypothesis of so-called "somatic markers". Such quantifiable markers represent states of body arousal which are integrated in the brain to give rise to emergent feeling in the immediate experience of the here-and-now (Damasio, 2003).

We want to stress that physiology can give access to low-level dimensions of conscious experience, such as arousal and basic emotions. Higher-level processes such as meaning attribution and interpretation are out of the reach of these techniques when they are not combined with subjective reports.

2.4 Cases

What follows is a summary of some ongoing studies carried out by the researchers at IPEM, whose line of research is inspired by the integration of coordination dynamics into EMC. Questions they attempt to answer can be summarised as follows. How is coordinated behaviour structured and evolving over time? How do individual actions lead to the emergence of stable forms of coordination between interacting people? How can we interpret the patterns observed in an experiment in light of the participants' subjective experience? And, most importantly, how can we design interactive experimental scenarios in light on the methodological reflections presented in previous sections of this paper?

Joint musical interactions *(A. Dell'Anna)*

In the context of musical interactions, the concept of “*homeostasis”* was proposed as a stable state characterized by an optimal equilibrium of behavioural, physiological and subjective parameters within a system (Mitchell, 2018). According to the proposal, the quality of a collective performance directly depends on the behavioural stability of the individual parts engaged in the interaction. In order to test the validity of such a construct, dell’Anna (Dell’Anna et al., 2020) designed a novel dyadic singing task inspired by the medieval Hocket technique: each participant is provided with a musical score, such that the partners have to alternate with one another in singing individual notes in order to form together the global pattern of the song.

After the task, both participants are presented with audiovisual recordings of their joint performance and asked to continuously move a slider up and down to rate the quality of the interaction. As we previously mentioned, such an approach allows correlation of the course of the performance to a time-varying series of subjective ratings, instead of entirely relying on questionnaires which fail to grasp the evolution of the experience over time. Furthermore, presenting the recorded performance to the participants implies that they do not have to rely solely on their memory for the assessment. In this sense, the method attempts to minimize the mediation between the actual experience and the moment when a participant is asked to recall it.

The influence of Carnatic dance of intentionality in piano performance *(G. Caruso)*

The author of this study is a professional pianist who dedicated her PhD project to the integration of performer’s self-reports into the EMC framework (Caruso et al., 2016). Starting from the EMC notion of mediation, she investigates how performer’s intentions are translated into observable actions and adds a reflective method to assist the artistic practice. She defines a two-way process of bottom-up processing, based on quantitative recordings of a performance, and top-down processing consisting of qualitative annotations from the performer.

The latter component of the process is defined as *performer-based analysis* and is based on the methods of stimulated recall (Bloom, 1953) and thinking-aloud (Van Den Haak et al., 2003). In such a way the performer provides structural, interpretative, and technical annotations to her own recorded performance (1st person perspective) and combines these with extracted features and patterns using quantitative methods (3d person perspective). This method is closely related to the paradigm presented in this paper as the performer-based analysis allows the visualisation of gesture-sound performance patterns with their interpretation through annotations of gesture-sound intentions. An additional step could be a more dynamical account of this approach incorporating non-linear aspects of emergence, self-organisation and sudden (non-linear) qualitative shifts in both the experience and the performance. Such an approach would account for the time-varying feedback loops occurring between the performance and the performer's experience.

Neural bases of coordinated collective behaviour *(M. Rosso)*

Over the past year, one of the authors of the present article started his project adopting a joint finger-tapping task for dyads of participants, during which their brain activity is recorded by means of electroencephalography (EEG). The main goal of the project is to investigate what changes in the brain activity of two people when they pass from behaving as individual units to behaving as a coupled system.

In the paradigm, each participant is instructed to tap the index finger on a sensor, keeping the tempo of a metronome. Depending on the condition, participants can see each other's actions, hear each other's actions or perform the task in isolation. The way the metronomes are programmed is meant to lead the two participants to dynamically explore a whole set of coordinative states over time, recurrently creating a conflict between the timing they are instructed to keep and the timing of the partner's actions. When participants are coupled via a sensory modality, we observed the emergence of spontaneous reciprocal attraction leading to stable coordinative behaviour despite instructions to ignore each other.

Brain dynamics taking place over the course of the interaction are systematically related to the time course of collective behaviour, to its stable states and to its transitions. Questionnaires are introduced at the end of experimental blocks to make sense of the observed patterns in light of the subjective interpretation of the participants. For instance, observed patterns of interpersonal coordination can be experienced as the result of either a cooperative or competitive process. The distinction implies that very different cognitive processes can account for similar observed coordinative patterns, hence the need for orienting and interpreting the analysis of brain dynamics in light of qualitative data.

Simulating musical interactions in virtual reality *(B. van Kerrebroeck)*

The aim of this ongoing study is to investigate the simulation potential of a musical interaction. Its motivation is the search for new, immersive experimental scenarios allowing careful measurement and control of experimental stimuli and to offer insights in technology-mediated (musical) interactions. The study uses virtual reality to compare settings in which a pianist plays with a live or a virtually recorded version of another pianist. The recorded pianist is controlled using principles of coordination dynamics to enable a realistic behaviour. Concretely, it allows the recording to adapt its tempo and playing position in the score based on the playing of the other pianist. To evaluate the simulation, we record behavioural data such as the timing of notes and player movement as well as physiological data such as pupil dilatation and gaze direction. This quantitative data is complemented with a questionnaire gauging experiential aspects of presence, flow, and immersion together with annotations using the performer-based analysis method. Non-linear techniques such as the ones presented in section 2.3.1 are then used to identify occurring patterns in the interaction (that is, in the behavioural and physiological data) and enriched with interpretations using the recorded qualitative data.

2.5 Conclusion

The main contribution of the present work is the proposal of an approach to orient methodological solutions in music research. This approach aims to integrate the dynamic and non-linear aspects from coordination dynamics with the embodied music cognition framework. In addition, it stresses the use of qualitative data from first person perspectives complementing quantitative methods to achieve a better understanding of complex phenomena at a systems level.

Interactions between brains and bodies can nowadays be quantified, described and modelled at a millisecond scale (Heggli, Konvalinka, et al., 2019) and show emergent patterns that they underlie. Bringing the model described in this paper into music research might shed light on the organisational principles underlying patterns in complex cultural phenomena such as musical interactions. A theory-driven use of the tools at the researcher’s disposal is an opportunity to contribute to the hermeneutical turn in digital humanities 3.0 (Bod, 2013). In line with the embodied music cognition theory, we argue that emergent patterns can be better understood by building a knowledge of the time-varying dynamics occurring at the level of the body, conceived as the mediator of interactions with sound and music (Leman, 2007). The experimental design in this line of investigation would ideally develop interactive scenarios that bring together physiological measures, motion analysis and subjective assessment in a way that minimises the inevitable mediation of the experience due to the measurement. Some operational steps in this direction have been initiated, for instance with the performer-inspired analysis (Desmet et al., 2012) and the use of motion capture to “mirror” intentionality in musical performance (Caruso et al., 2016). In parallel, the corpus of existing literature on joint action spanning from the simplest forms of rhythmic interactions (Konvalinka & Roepstorff, 2012) to actual musical practice (Sänger et al., 2013) provides a solid grounding for extending the study of dyads or groups of subjects as interacting systems organised by coordination dynamics. Current empirical music research already employs elements of the model presented here. However, a further cross-disciplinary approach guided by an integration of these different elements might lead to new and deeper understanding of musical interactions.

3.

A methodological framework

Published as: Van Kerrebroeck, B., Caruso, G., & Maes, P. J. (2021). A methodological framework for assessing social presence in music interactions in virtual reality. Frontiers in Psychology, 1880.

3.1 Introduction

Virtual reality (VR) encompasses a plethora of technologies to create new environments, or simulate existing ones, via computer-generated multisensory displays (Scarfe & Glennerster, 2019; Sutherland, 1968; J. Taylor, 1997). Complementary to multisensory displays are technologies for capturing physical body movement to facilitate embodied control and interactions with(in) computer-generated (virtual) environments (Yang et al., 2019). VR technologies hence provide a technological mediation between performed actions and multisensory perceptions, extending the natural sensorimotor capacities of humans into the digital world (Biocca & Delaney, 1995; Ijsselsteijn & Riva, 2003; Kornelsen, 1991). Crucially however, VR typically aims at making its mediation invisible, creating for users the illusory feeling of nonmediation; a feeling coined with the concept of "presence" (Lombard & Ditton, 1997; Riva, 2006). This concept of presence may encompass multiple categories, related to the physical environment, the user's own body, as well as its social environment. The first category – physical presence or telepresence – pertains to the illusory feeling for users of actually being present in another environment than the one they are physically in (Minksy, 1980; Sheridan, 1992; Slater & Sanchez-Vives, 2016). Another category, self-presence, is rooted in the capacity of VR to map the physical body movements of a user onto the moving body of a virtual avatar. The potential of embodying virtual avatars allows creation of the illusory feeling for a user of owning, controlling and being inside another body than its physical one (self-presence) (Braun et al., 2018; De Oliveira et al., 2016; Kilteni, Groten, et al., 2013; Matamala-Gomez et al., 2021). In addition, (bodily) acting within a virtual social context may create a sense of being together (co-presence), or interacting with others (social presence) while actually being physically remote (Garau et al., 2005; Oh et al., 2018; Parker et al., 1978; Parsons, Riva, et al., 2017).

In the current paper, we advocate for establishing VR as a research tool for studying social music interaction and sense-making. We see the relevance of VR precisely within its capacity to create a sense of presence across the different categories described above. Within (empirical) research contexts, this aspect makes that VR can offer accurate control over data measurement, stimuli creation and manipulation, while retaining a level of realism for test subjects comparable to their real-life music experiences. In that sense, VR holds the potential of bringing the external validity ("research in the wild") and internal validity ("research in the lab") of social (music) cognition research closer together (Kothgassner & Felnhofer, 2020; Parsons, 2015). In addition, the use of VR can also widen perspectives and remove obstacles by extending research contexts by designing "impossible stimuli" and new social interactions and scenarios (T. K. Metzinger, 2018; Parsons, Riva, et al., 2017). In the first part of the article, we will discuss in more detail how a VR-based approach has its roots in earlier, human-centred research across a broad range of scientific disciplines and how it holds potential for empirical research on social music cognition and interaction.

In the remainder of the article, we focus on a fundamental prerequisite for establishing the advocated VR-based research method; namely, the idea that VR can actually create a sense of social presence. This is particularly challenging given that music provides a highly particular context of human social interaction. It involves the body as a source of expressive and intentional communication between co-performers, carried out through a fine-tuned and skilful co-regulation of bodily articulations (Leman, 2007; Leman & Maes, 2015). This co-regulation of bodily articulations is a complex process, involving many body parts, and taking place across multiple, hierarchically organized spatial and temporal scales (Eerola et al., 2018; Hilt et al., 2020). Successful co-regulation requires hence that action-relevant information at multiple scales is properly exchanged through the different sensory modalities. In particular the auditory and visual sense are important in signalling communicative cues, related to music-structural aspects and emotional expression (Coorevits et al., 2020; Goebl & Palmer, 2009; Keller & Appel, 2010; Williamon & Davidson, 2002). This complex, embodied nature of music interaction puts considerable demands to communication technologies that aim at mediating social music interactions via digital ways. VR however is, in principle, highly promising as it allows the animation of full-body, three-dimensional virtual humans based on real-time mapping of body movements of actual people captured by motion capture systems (avatars) or based on computer-modelling and simulation of human behaviour (agents) (Cipresso, 2015). These animated, three-dimensional avatars and agents can be observed by others from a freely chosen and dynamic first-person perspective providing a foundation for the complex information exchanges required for successful music interactions. This turns a VR environment into a potential digital meeting space where people located in physically distinct places, together with computer-generated virtual humans, can interact musically with one another. However, it is crucial to further assess the quality of social interactions with virtual avatars and agents (Kothgassner et al., 2017) and to assure the required levels of realism and social presence.

For that purpose, a main objective of this article is to introduce a methodological framework to assess social presence in virtual music interactions. We thereby consider social presence as a multi-layered concept rooted in, and emerging from, the behavioural and experiential dynamics of music interactions. We are able to assess these dynamics by integrating direct data measurement related to performance output, body movement and (neuro)physiological activity with subjective self-report measures. As such, the framework facilitates the design of interactions, avatars, and agents to obtain empirical data and investigate aspects of the subjective experience such as for example empathy, intimacy, and togetherness. Direct assessment is preferred over post-experimental reports to avoid the influence of self-referential cognitive processes or interfering in the interaction. In the second part of the paper, we apply the framework to a case-study of a social music interaction in VR. The case-study presents real and virtual interactions between two expert pianists, a pianist with an avatar and a pianist with an agent and demonstrates similarities and differences revealed throughout the framework's layers. Finally, we conclude the paper with a discussion on the relevance of our framework using insights from the case-study's analysis and present directions for future work.

3.2 VR: A research tool for studying social interactions

Around the 21th century, VR started to develop as a valuable methodological tool in human-centred research, including the social and cognitive (neuro)sciences (Biocca, 1992; Biocca & Delaney, 1995; Blascovich et al., 2002; Fox et al., 2009; Pan & Hamilton, 2018; Parsons, 2015; Parsons, Gaggioli, et al., 2017; Slater & Sanchez-Vives, 2016), philosophy (T. K. Metzinger, 2018), the humanities (Cruz-Neira, 2003), product design (Berg & Vance, 2017), marketing (Alcañiz et al., 2019), medicine (Riva et al., 2019), and healthcare (Teo et al., 2016). Although VR-based research exhibits a richness in variety and discipline domains, the relevance of VR in human-centred research can, in general terms, be captured by two specific traits; namely the ability of VR to simulate existing, "real-life" contexts (simulation trait) and its ability to extend human functions or to create new environments and contexts (extension trait).

The simulation trait of VR relates to the inherent paradox in traditional approaches in empirical research. To obtain valid results and insights, the researcher is motivated to observe phenomena "in the wild" without interventions. However, this approach allows little control over stimuli, often has to cope with a number of confounding variables and provides challenges to perform reliable measurements. On the other end, the researcher performs experiments in a controlled lab setting to obtain generalizable results. This approach however is often overly reductionistic and not ecologically valid. The use of VR technology allows to bridge these extremes by simulating real-life settings in a controlled environment. It can be understood as an alternative empirical research paradigm (Blascovich et al., 2002), offering substantial additional benefits over traditional research practices in laboratory or field conditions. The use of VR allows precise control over multimodal, dynamic and context rich stimuli (Parsons, Gaggioli, et al., 2017) while retaining a level of realism required for realistic responses. Despite the need for technological expertise, research practices using VR are becoming more accessible and standardized and can thus provide representative sampling and better replicability (Blascovich et al., 2002). Given the digital nature of creating VR contexts and the requirement of appropriate sensorimotor sensors, VR technology also offers flexibility in the means of and choices in recording data.

A second trait can be related to Marshall McLuhan's (McLuhan, 1964) understanding of technology as an extension of the human body, mind and biological functions. This view resonated in the early accounts of VR pointing to the ability of VR to create sensorimotor and social experiences not possible or desirable in the actual physical world. Accordingly, VR was defined in terms of a "medium for the extension of body and mind" (Biocca & Delaney, 1995), creating "realities within realities" (Heim, 1994) or "shared/consensual hallucinations" (Gibson, 1984; Lanier, 1988) "bounded […] only by desire and imagination" (Benedikt, 1991). Important to note is that, in most of current human-centred research, this ability of VR is seldomly employed as a form of mere escapism from the physical world. In contrast, and in line with Lanier's ideas, VR is mostly used to "make us intensely aware of what it is to be human in the physical world, which we take for granted now because we're so immersed in it" (Lanier, 1988). Accordingly, the use of "impossible stimuli" and illusions generated in VR have contributed substantially to a better understanding of profound aspects of human embodied cognition and social interaction (T. K. Metzinger, 2018; Parsons, Gaggioli, et al., 2017). For instance, VR technology is capable of selectively modulating our perception of space (Glennerster et al., 2006), time (Friedman et al., 2014), (social) cognition (Tarr et al., 2018) and the body (Petkova & Ehrsson, 2008). It has the potential to influence different representational layers of the human self-model (T. K. Metzinger, 2018) leading to phenomena such as virtual embodiment, (virtual) body swapping (De Oliveira et al., 2016; Petkova & Ehrsson, 2008) and increasingly frequent and complex "social hallucinations" (T. K. Metzinger, 2018).

Given these traits and their potential, the use of VR in music research has increased over the recent decade (Çamcı & Hamilton, 2020). A first category of studies primarily leveraged the simulation trait of VR. They created real-life virtual settings in which to investigate various topics such as music therapy (Bissonnette et al., 2016; Orman, 2004), music education (Orman et al., 2017; Serafin et al., 2017), music performance (Glowinski et al., 2015; Williamon et al., 2014) and the relation between sound and presence (Kern & Ellermeier, 2020; Kobayashi et al., 2015; Västfjäll, 2003). A good example of simulating a real-life setting is given by Glowinski, who investigated the influence of social context on performance (Glowinski et al., 2015). Specifically, Glowinski and colleagues asked participants to perform a musical task in a virtual concert hall while controlling for audience gaze. Other studies focused more on extending real-life contexts. They range from the search towards new virtual instruments (Berthaut et al., 2015; Hamilton, 2019; Honigman et al., 2013; Serafin et al., 2016) to new interactions and the development of interaction design principles (Atherton & Wang, 2020; Deacon et al., 2017). While we made a clear conceptual distinction between the simulation and extension trait, this distinction is opaquer in practice. An effective research paradigm has been to simulate a musical scenario in VR, subsequently extending some human function such as modulating the feeling of body ownership using virtual embodiment, to investigate behavioural changes (Kilteni, Bergstrom, et al., 2013).

A critical requirement however for using VR as a research tool in the study of social music interaction is the ability to establish social presence: the illusory feeling of actually being together and interacting meaningfully with human-embodied avatars or computer-controlled agents in VR. Research on social presence may contribute to social music cognition and interaction in two important ways. First of all, referencing again the quote by Lanier (Lanier, 1988), social music interaction in VR forces researchers to think about, and develop knowledge on, the general nature of human social cognition and sense-making, "which we take for granted now because we're so immersed in it". We therefore advocate that the definition, the measurement, and the testing of social presence in VR is rooted in (neuro)cognitive research on social human cognition and sense-making. In the present article, the core aim is to propose and test such an approach to the definition and measurement of social presence in musical VR (see section 3.3). Secondly, under the condition that social presence can be reliably established, it becomes possible to accurately control and manipulate the many variables that characterize a music interaction, including the context in which the interaction occurs. For instance, it becomes possible to control the perspective that people have on one another, the distance at which they are positioned, the sensory coupling between people, the appearance of people (for example facial expression, age, gender), environmental properties, the actual musical behaviour and bodily performance of VR agents (for example timing, quantity of motion), among other variables. This offers almost limitless possibilities to extend the empirical investigation of the principles of social music interaction and sense-making within (simulated) ecologically valid music environments. In the following section, we describe the methodological framework that we propose to define, measure, and test social presence in VR music interaction contexts.

3.3 A methodological framework to assess social presence in VR

Most research so far has relied on self-report questionnaires to assess the subjective feeling of social presence (Cui, 2013; Oh et al., 2018). The mere use of subjective ratings however poses important limitations, as these provide only indirect and post-hoc measures of presence, lack subtlety and are often unstable and biased (Cui, 2013). In the current article, we propose an alternative, pragmatic approach, considering social presence as emerging from the performative, behavioural and experiential dynamics inherent to the social interaction. This allows the assessment of social presence using a combination of qualitative, performer-informed methods and quantitative measures of the performance, behaviour and (neuro)physiological responses of users by operationalizing them into concrete, direct, and measurable variables. Crucially, in this pragmatic approach, we define the level of social presence as the extent to which social behaviour and responses in simulated VR contexts resemble behaviour and responses in corresponding real-life musical contexts (Johanson, 2015; Minksy, 1980; Scarfe & Glennerster, 2019; Slater, Lotto, et al., 2009).

To allow comparison between virtual and real-life scenarios, we rooted our framework for social presence in the "interaction theory", which currently is the most dominant theory in the social sciences to understand social cognition and sense-making (De Jaegher & Di Paolo, 2007; Fogel, 1993; Froese & Fuchs, 2012; Gallagher, 2001; Gallotti & Frith, 2013; Kiverstein, 2011; Schilbach et al., 2013). Proponents of the interaction theory consider social cognition essentially as an embodied and participatory practice, emerging in real-time co-regulated interaction not reducible to individual processes. In line with this account, we consider successful co-regulation as a foundational criterion to establish social presence in VR. Importantly, in our framework, we consider social co-regulation both from the viewpoint of the quantifiable bodily and multisensory patterns of interpersonal interaction, as from the viewpoint of the intersubjective experience and participatory sense-making (De Jaegher & Di Paolo, 2007). Together with the actual musical outcome, these two interrelated aspects of social co-regulation form the three main layers of our framework to determine the degree of social presence in VR music contexts. Layers of the framework are shown in Figure 3-1.

3.3.1 Layer 1: Performance output

The performance output layer relates to the (un/successful) realization of musical ideas or goals, which may be strictly prescribed in musical scores, loosely agreed upon, or emerge in the performance act itself, depending on the performance type and context. Music performance analysis has been advanced by research and development in the domain of music information retrieval, providing ample techniques and methods for assessing music performance properties (Lerch et al., 2020). These are typically extracted from audio recordings, although other multimodal signals such as body movement are increasingly being used. Further, we advocate for taking into account time-varying features related to timing, synchronization, (joint) multiscale recurrence patterns and complexity measures as these may signal the quality of the performance output. These quantitative measures should ideally be complemented with qualitative, performer-inspired methods to reliably interpret the quantitative outcome measures. They include subjective evaluations in the form of aesthetic judgements of the performance output by the performers themselves.

3.3.2 Layer 2: Embodied co-regulation

A successful musical output relies on a skilful, joint coordination of co-performers' actions and sounds. In line with the interaction theory on social cognition described above, we consider social music interaction as a dynamic and continuous unfolding process of co-regulation, in which performers mutually adjust to one another in a complex interplay of action and multimodal perception. This process of co-regulation integrates various levels and mechanisms of control, ranging from low-level spontaneous coordination based on dynamical principles (Kelso, 1995; Tognoli et al., 2020), to higher-level learning, predictive processing and active inference (Gallagher & Allen, 2018; Koban et al., 2019; Sebanz & Knoblich, 2009). In our proposed methodological framework, we specifically aim at capturing patterns, relationships, and recurrences in the process of co-regulation at the level of the interacting system as a whole. We thereby advocate for the integration of timeseries analyses from the domain of dynamical systems theory, as these are ideally suited to unveil dynamic patterns of interpersonal coordination across multiple body parts and temporal and spatial scales (Eerola et al., 2018; Hilt et al., 2020). Patterns can then be found on multiple levels such as in the attention dynamics from a participant's gaze direction, in expressive gestures with communicative cues from head nodding as well as in the structures of full-body movements resulting from body sway synchronization. The ability to quantify bodily and multisensory patterns of co-regulated interaction between music performers in VR-mediated music contexts is foundational in our approach, as in our view, successful co-regulation is a decisive factor in performers' feelings of social presence.

3.3.3 Layer 3: Subjective experience

This layer deals with the subjectively experienced interaction qualities and sense-making processes of individuals. It contains a combination of quantitative and qualitative methods to link mental states, (expressive) intentions and meaning attributions to observations in other layers. The quantitative methods include the analysis of (neuro)physiological signals as they can give access to low-dimensional aspects of the conscious experience. For instance, electromyography (P. Ekman, 1992) and pupillometry (Laeng et al., 2012) among others, have proven to provide valid markers of cognitive and affective user states in virtual performance contexts such as attention and workload measures, vigilance, affect, and flow (Schmidt et al., 2019). Heart-rate and skin-conductance (Meehan et al., 2002) as well as electroencephalography (Baumgartner et al., 2006) represent good candidates as they are capable of directly assessing the feeling of presence in virtual performance contexts. Complementary, from a more qualitative and performer-oriented point of view, one can ask participants for time-varying ratings of their intentional (joint-)actions and expressions through audio-video stimulated recall

(Caruso et al., 2016). In addition, via self-report questionnaires, one can probe for mental states such as social presence, flow and feelings of togetherness (Lessiter et al., 2001; A. J. Martin & Jackson, 2008; Witmer & Singer, 1998). Finally, because of the multi-layered nature of social presence, open questions and semi-structured interviews focused on individual experience of the virtual other can help to fill analysis gaps and interpret quantitative findings across the different layers.

3.3.4 Operationalization within a performance setting

Layered frameworks have been helpful in earlier research for structuring the investigation into music interactions (Camurri et al., 2001; Leman, 2007). The layered framework here extends these approaches with a focus on the complementary nature of a mixed-method approach and the time-varying aspects, viewing the musical interaction as consisting of multiple interdependent parts. Layers are functionally coupled by non-linear relations allowing the emergence of patterns in time-varying dynamics in each layer. They frame and couple the dynamics of quantitative bodily and multisensory coordination patterns with the (inter)subjectively felt qualities of the music interaction. The framework aims to serve as a template to map this dynamic landscape of time-varying dynamics and aid in the uncovering of insightful states and transitions in a broad range of social interactions. It allows the investigation of social presence in different performance settings and distinguish interactions using a specific operationalization of qualitative and quantitative methods in each layer. One performance setting might have time-delay or phase as variable of interest in the performance layer, while another might focus on frequency. Some settings will require the observation of neurophysiological signals while others might focus on movement data or self-reporting. Application of the framework can then identify interactions with for example close coordination and intense subjective experiences but inferior performance such as when two tennis players are struggling to have long rallies but nevertheless experience heightened attention and synchronized movements. Other interactions can have successful performance outcomes, but fail in creating fertile metastable dynamics (Kelso, 1995; Tognoli et al., 2020) in other layers. Examples can be found in the interactions between human users and current state-of-the art artificial intelligence in video games, humanoid robots or virtual assistants that lack successful (embodied) co-regulation and dynamic intentional relations.

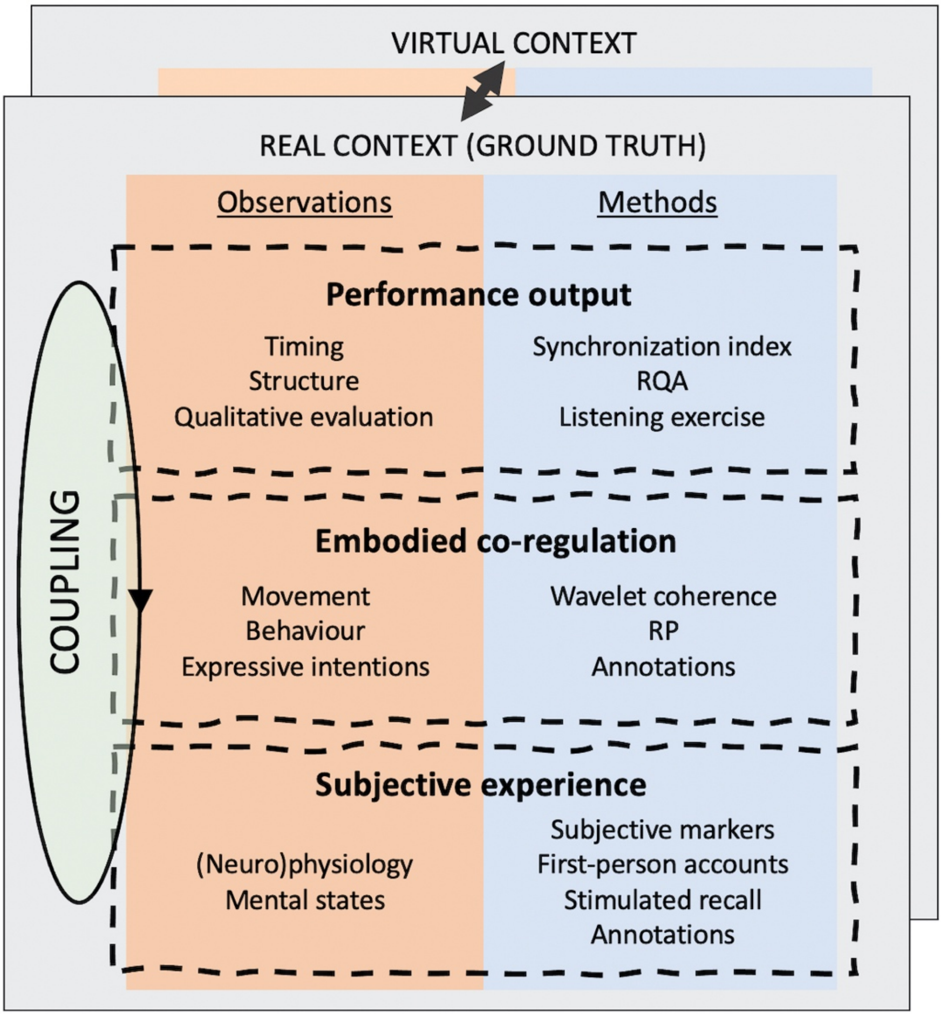


Figure 3-1: Overview of the methodological framework to operationalize social presence in VR music contexts.

3.4 A case-study: Piano Phase

'Piano Phase' (1967) is a composition for two pianos, written by minimalist composer Steve Reich. The piece applies his phase-shifting technique as structuring principle of the composition and is written for two pianists. In short, starting from unison, intermittent gradual tempo changes cause increasing phase shifts between the melodic pattern played by each pianist, which in turn, in their dynamic variety of interlocking, lead to the emergence of a variety of harmonies until pianists are back in unison.

The piece was chosen as a case study, as it provides an excellent musical case to assess social presence in VR music performance across the different layers in our proposed methodological framework. First, the performance output, the instructed phase shifts throughout time, can be objectively assessed and compared across different performances. Second, the performance requires skilled co-regulation between pianists in order to successfully perform. And third, as also Reich acknowledges, the performance of the piece has profound psychological aspects, related to sensuous-intellectual engagement, and strong (inter)subjective experiences of heightened attention, absorption and even ecstasy.

3.4.1 Research question

The case study was meant to empirically evaluate and test our pragmatic approach to the definition and measurement of social presence in real-time VR music performances based on the proposed methodological framework. For that purpose, we designed different performance contexts that enabled us to compare performances of Piano Phase in VR, with a corresponding (baseline) performance of Piano Phase under normal, “real-life” conditions (see 3.4.3 Design). In all conditions, we captured an elaborated set of quantitative data related to the experience, behaviour, and performance of the pianist duo. In addition, we complemented this data with qualitative methods to integrate experiences and intentions from a performer point of view. Based on this quantitative and qualitative data and guided by the layered analysis model inherent to the proposed methodological framework, we could then conduct comparative analyses across the performance contexts to evaluate social presence the test subject experienced in VR.

3.4.2 Participants

The case study involved three expert pianists: one test subject and two research confederates. All pianists had over 10 years professional music experience. The test subject (female, 38 years, in the following termed “Pianist 1”) was not familiar with the piece Piano Phase through earlier performances and did not have earlier VR experiences. A second pianist (male, 32 years, in the following termed “Pianist 2”) functioned as research confederate in Condition 1 and 2 and had concert experiences in performing Piano Phase. Finally, the third pianist (female, 44 years, in the following termed “Pianist 3”) was another research confederate in Condition 3 and co-author of this study. She had no experience in performing the piece but did have experience with VR.

3.4.3 Design

The experiment consisted of three conditions as presented in the schematic overview in Figure 3-2. In each condition, Pianist 1 (the test subject) performed Piano Phase together with a research confederate while wearing a HMD (Pianist 2 in Condition 1 and 2 and Pianist 3 in Condition 3). The fundamental distinction between the three conditions was the level of realism of the confederate partner as perceived by Pianist 1:

- **Condition 1 (ground truth):** Pianist 1 and 2 performed Piano Phase under normal, “real-life” concert conditions. Pianist 1 visually perceived Pianist 2 in a natural, physical manner.
- **Condition 2:** Pianist 1 and Pianist 2 perform together in real-time, but Pianist 1 visually perceived Pianist 2 as a human-embodied virtual avatar.

- **Condition 3:** Pianist 1 performs together with a computer-controlled virtual agent of which the performance behaviour was rooted in an earlier recording of Pianist 3.

For more information on the display methods, see “Performance setting and virtual simulation displays” in 3.4.4 Materials and apparatus below.

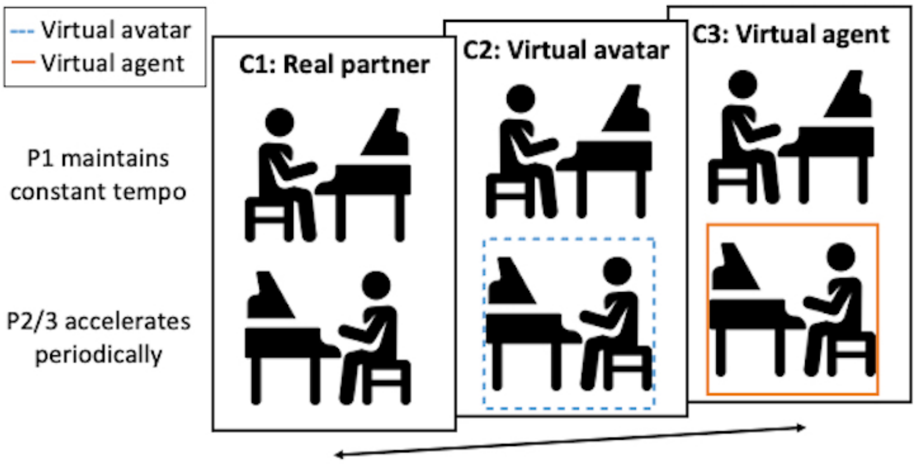


Figure 3-2: Conditions in the case-study
(P1/2/3 = Pianist 1/2/3, C1/2/3 = Condition 1/2/3)

3.4.4 Materials and apparatus

Piano keyboards: The piano keyboards used for the performances were digital interface MIDI controllers. Pianist 1 and 3 played on a Yamaha P60 and pianist 2 played on a Roland RD700SX. MIDI information was processed in Ableton Live 10 to generate piano sounds using a Native Instruments Kontakt 6 plugin. Speakers were placed underneath each piano keyboard to assure a coherent sound source localization throughout the different performances.

Data measurement setup: Multiple technologies were used to capture and measure bodily behaviour and performance aspects of the pianists into quantitative timeseries data. Concerning the performance, we recorded MIDI data from the piano keyboards including MIDI note numbers, note-on/off times and note velocities. We captured full body movements of all pianists (3D position, 120 Hz) using a multi-camera Qualisys optical motion capture system (OQUS 7+ cameras). In addition, video was recorded using a four-camera Qualisys Miqus system. Further, we captured how Pianist 1 distributed her body weight on the chair using four pressure sensors mounted underneath the four legs of the piano chair. Finally, we tracked the eye movements of Pianist 1 using the Tobii eye-tracking technology from the HMD of model HTC Vive Pro Eye. An overview of the technical set-up is shown in Figure 3-4.

Performance setting and virtual simulation displays: The full experiment took place in the Art and Science Interaction Lab of IPEM, a 10m-by-10m-by-7m (height) space surrounded by black curtains that resembles a realistic performance space. The two piano keyboards were placed opposite to each other under an angle of about 60° so pianists could turn towards each other (see Figure 3-3).

- **Condition 1:** In this condition, Piano Phase was performed under normal, real-life conditions. To match the two other performance conditions, we asked Pianist 1 to wear a VR HMD with the pass-through camera activated in order to avoid that the constraints of the HMD would function as a confounding factor while maintaining a normal, physical exchange of auditory and visual information between Pianist 1 and 2.
- **Condition 2:** In this condition, Pianist 1 was disconnected from visual information coming from the physical environment. Instead, all visual information related to the room, piano keyboard, hands, and the co-performing Pianist 2 was simulated in Unity and provided to Pianist 1 via the HMD. The real-time simulation of Pianist 2 into an avatar was realized based on a full body movement capture using the Qualisys system described above of which the data was streamed to Unity (see Figure 3-4). An important consideration in the study was to also simulate the hands of Pianist 1 as earlier research indicated that this may substantially increase the feeling of (self-)presence (Banakou & Slater, 2014). We used the Leap Motion system for that purpose which allowed tracking and display of fine finger movements.
- **Condition 3:** In this condition, Pianist 1 was disconnected from all direct visual information as similar to Condition 2. However, this time, Pianist 1 performed Piano Phase together with a computer-controlled, adaptive virtual agent. The full body movements and MIDI piano performance of the virtual agent were rooted in pre-recorded timeseries data of an actual pianist (Pianist 3), who was asked to perform the same role as the one of the virtual agent (see below, Task). For the virtual agent animation, we used the Kuramoto model to automatically phase-align these pre-recorded timeseries data and the accompanying audiovisual VR display to the real-time performance of Pianist 1. This allowed the accurate control of the phase of the musical part of the virtual agent with respect to Pianist 1 and hence, to perform the piece dynamically as prescribed by composer Steve Reich. Apart from the virtual agent, all other display factors were similar as in Condition 2.

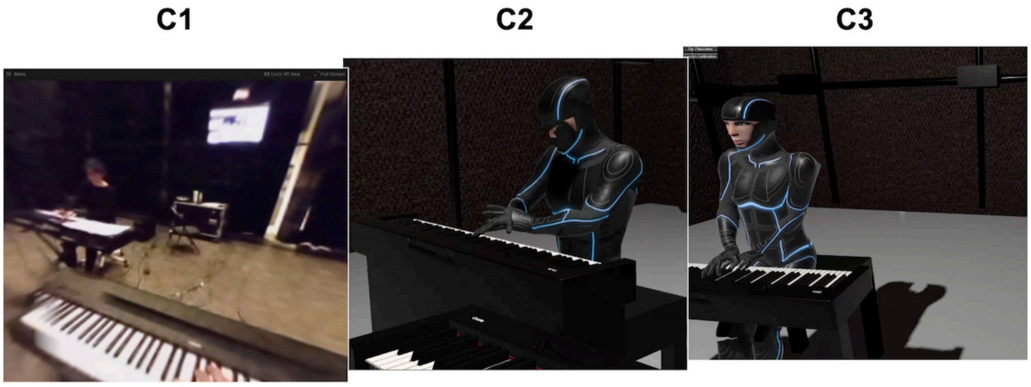


Figure 3-3: View of Pianist 1 (the test subject) in each condition as seen through Pianist 1's head-mounted display

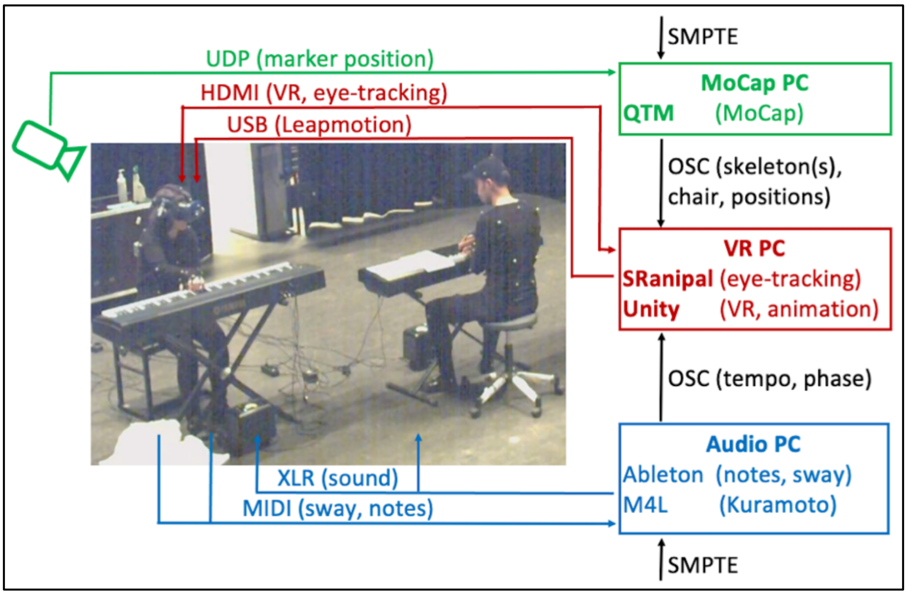


Figure 3-4: Schematic overview of the case-study's technical set-up

3.4.5 Task and procedure

The task for the pianist duos in each condition was to perform Piano Phase as prescribed by the composer Steve Reich. The compositional idea of Piano Phase is to repeatedly play the same melodic pattern together while at certain instances perform gradually increasing phase-shifts across these patterns until pianists play in unison again. The first bars of the piece are shown in Figure 3-5. Reich's compositional instruction is as follows: "The first pianist starts at 1 and the second joins him in unison at 2. The second pianist increases his tempo very slightly and begins to move ahead of the first until, (say 30 to 60 seconds) he is one sixteenth ahead, as shown at 3. The dotted lines indicate this gradual movement of the second pianist and the consequent shift of phase relation between himself and the first pianist. This process is continued with the second pianist gradually becoming an eight (4), a dotted eight (5), a quarter (6), etc., ahead of the first until he finally passes through all twelve relations and comes back into unison at 14 again." (Reich, 2002). In the performance, the pianist that is assigned the top part of the score keeps a constant tempo, while the pianist that is assigned the bottom part performs the gradual phase-shift by gradually increasing his/her tempo. In our study, Pianist 1 was always assigned the top part of the score, while Pianist 2 and 3, were assigned the bottom part in Condition 2 and 3 respectively. We fixed the number of repetitions for each bar at eight for better experimental control and kept the tempo at 72 BPM (one beat for six 16th notes or a dotted quarter note).

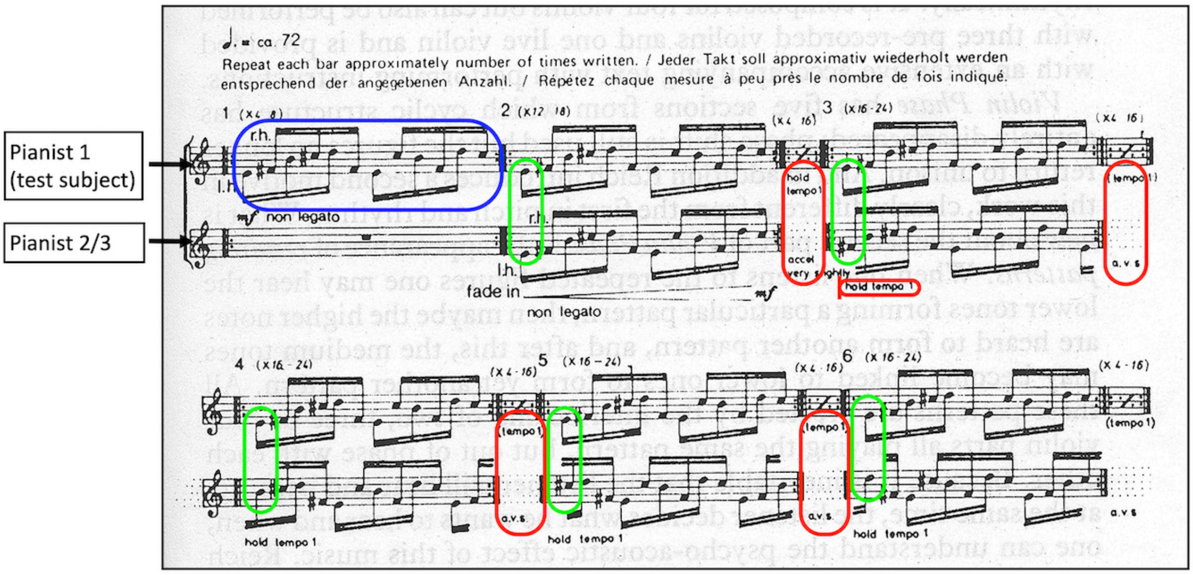


Figure 3-5: Annotated first 11 bars of Piano Phase (1967) by Steve Reich
(blue = melodic pattern, green = phase-difference, red = tempo instruction)
(Reprinted with kind permission from Universal Edition)

A month before the experiment, Pianist 1 was asked to prepare for a performance of the musical composition. Pianist 1 received an audio recording of her part with isochronous notes, uniform velocities, and linear accelerations to help with practicing. Upon arrival, the test subject was told the experiment consisted of a preparation phase followed by three repetitions of the full piece. She was then given a questionnaire and asked to change into a motion capture suit afterwards. A data skeleton was built using some recordings of her walking and playing the piano after which she was asked to calibrate the HMD's eye-tracking.

After the explanations, the test subject practiced the task with pianist 2 for about 15 minutes without wearing the HMD. When both pianists indicated they were ready for the performance, the test subject was given the HMD to get accustomed to the virtual environment after which they performed the three conditions. A questionnaire and break were given after each condition. The test subject was not told that the agent in condition 3 was computer-controlled. The experiment concluded with a semi-structured interview about the experience. A week after the experiment, the test subject was asked to listen and evaluate randomized audio recordings of each condition as well as fill in a final questionnaire.

3.4.6 Analysis

This subsection describes the analysis in the application of the methodological framework on the case-study. It presents the choice of quantitative and qualitative methods used in each layer to obtain insights that will be discussed in section 3.4.7 below.

3.4.6.1 Layer 1: Performance output

In this layer, we investigated if pianists succeeded in executing the compositional instruction. In Piano Phase, both pianists repeat a 12-note pattern of which one pianist accelerates at specific moments for a specific period. This should result in an alternation of stable periods characterized by a consistent relative phase relationship of the note patterns of both pianists and intermittent periods characterized by gradual shifts towards an increased relative phase of the note patterns.

First, we used note onsets of both pianists to determine tempo, inter-onset-intervals (IOIs) and relative phase between pianists. One full phase cycle was defined as 12 notes. Using the relative phase, we calculated the synchronization index (SI) as a measure of stable periods characterized by a consistent relative phase between pianists (Mardia & Jupp, 2008). A SI of 1 represents perfect synchronization and 0 no synchronization.

Next, we looked at musical structure using time-dependent joint Recurrence Quantification Analysis (RQA) of the relative phase. RQA is a non-linear technique for the assessment of dynamic systems and allows us to identify transitions and behaviour of a system by analysing patterns of recurrences in a low dimensional phase space from potentially higher dimensional timeseries (Marwan et al., 2007). Since relative phase between pianists represents the driver of the musical composition, RQA on a phase space of relative phase allows assessment of transitions and dynamics in the musical performance. RQA metrics such as the recurrence rate (RR), determinism (DET) and trapping time (TT) were calculated to measure the percentage of recurrences, the percentage of recurrences that are stable and the average length of stable recurrences. We used an embedding dimension of 4 and a time-delay of 0.3s for a joint RQA. The minimal diagonal length for calculating DET was set to 0.3s. Joint RQA parameters were obtained by looking at extrema of mutual information and false-nearest neighbour metrics (Marwan et al., 2007). The radius was set to 0.55 to produce around 10% of recurrence across conditions.

Finally, we complemented this quantitative data analysis with subjective evaluations. Five days after the experiment, the test subject received six, 15-second, randomized, audio recordings from each condition. She was asked to score each recording on a scale from 0 to 10 on expressiveness (dynamics, accents), timing (rhythm, tempo), interaction quality (collaboration) and appreciation (engaging, positive). In addition, she was asked to leave general remarks for each recording.

3.4.6.2 Layer 2: Embodied co-regulation

In this layer, we assessed movement, behaviour and (expressive) intentions of the pianists and the means through which pianists actually achieved a successful execution of the musical score. When recording the stimuli for the experiment, pianists used head nods to communicate successful transitions and divided the tasks of counting repetitions and measures amongst themselves. Communication and co-regulation between pianists played an essential role for a successful performance and realization of the compositional idea behind Piano Phase.

For that purpose, we recorded head movements as 3D spatiotemporal series to obtain communicative cues and signals of mutual understanding in the piano performance (Castellano et al., 2008). To detect correlated frequencies and their phase angles, we performed a wavelet coherence analysis on the series' main principal component.

We complemented this analysis of dynamics with annotations of specific, expressive gestures in the performance. Concretely, we identified head nods between pianists using the ELAN software (Sloetjes & Wittenburg, 2008) to see whether communicative cues at transitional moments remained consistent across the different performances.

As a measure of coupling and attention towards the other, we recorded the test subject's gaze direction and calculated the angle with the confederate's head position.

We measured postural sway timeseries using the pressure sensors in the piano chair. We summed these series and used a normalized and unthresholded recurrence plot to look at stable periods and identify transitional moments in the performance (Marwan et al., 2007). The recurrence plot had an embedding dimension of 5 and time-delay of 0,35s which were defined using mutual information and false-nearest neighbour metrics. Radii were set to produce around 10% recurrences for each condition ([Radius, RR%] equal $[0.065, 10.66]_{C1}$, $[0.185, 10.63]_{C2}$, $[0.104, 10.55]_{C3}$).

3.4.6.3 Layer 3: Subjective experience

In this layer, we looked at physiological signals and self-reported scores as windows into the test subject's experience after and during the interaction. The immersive tendencies questionnaire (Witmer & Singer, 1998) was taken before the experiment. Self-reported scores were obtained using the flow short scale (A. J. Martin & Jackson, 2008), the presence questionnaire (Witmer et al., 2005) and three custom questions about the overall interaction ("Did you enjoy the interaction", "How close did you feel to your musical partner" and "How natural did you experience the interaction with your partner"). Additional presence questionnaire items as proposed by Slater and Lessiter were included in the presence questionnaire as well (Witmer et al., 2005). A semi-structured interview about the overall experience was conducted upon conclusion of the experiment. We sent a general question to describe the experience in each condition together with randomized audio recordings from each condition.

We recorded pupil dilatation using the built-in eye-tracking functionality in the HMD as an estimate of the intensity of mental activity and of changes in attention or arousal (Laeng et al., 2012).

3.4.7 Results

This section presents the results from analysing the case-study data. We evaluate each layer of the methodological framework in each condition with results shown in Figure 3-6, 7 and 8. Given the fact that only one dyadic couple was observed in all conditions, results are descriptive and meant to provoke reflections leading to the discussion section below.

3.4.7.1 Layer 1: Performance output

The score has a tempo indication of 72BPM and participants performed it slightly faster ($[\text{mean, STD}]_{\text{tempo}}$ equalled $[74.73, 4.44]_{C1}$, $[75.90, 4.62]_{C2}$, $[74.10, 5.67]_{C3}$ BPM). IOI variability was comparable across conditions with a higher variability for Condition 3 ($[\text{Mean, STD}]_{\text{IOIvar}}$ equalled $[7.47, 3.63]_{C1}$, $[7.26, 3.41]_{C2}$, $[12.55, 2.73]_{C3}$ ms). As the agent was programmed to be attracted towards 72BPM, the slightly higher tempo of the performance made its tempo corrections larger when accelerating and synchronizing. Across conditions however, Condition 3’s tempo was closest to the instructed 72BPM.

Bars in the composition represent stable or accelerating tempos with respectively constant or shifting relative phase periods in the performance. These periods and their associated transitions were determined by thresholding the synchronization index as indicated by the grey bars in Figure 3-6, 7 and 8. The third condition had a more variable bar length distribution with extremes of a long period of synchronization at 240 seconds and a turbulent moment of successive (de)synchronization around 280 seconds. Interestingly, a longer period of synchronization is found at the same relative phase in Condition 1 as well. This relative phase and resulting harmony might represent an attractor for the pianists in which it was easy to accelerate in but difficult to accelerate out of.

Overall, increasing relative phase and a fluctuating synchronization index were present across conditions as shown in Figure 3-6, 7 and 8 [B]. The third condition did have a turbulent moment in the beginning and the middle of the performance as well as a sudden transition at the end. These moments resulted from the delay in tempo tracking for the virtual agent, an erroneous note in the stimuli and the test subject jumping to synchronization with the agent towards the end.

Joint RQA shows alternating periods of (de)synchronization (Figure 3-6, 7 and 8 [C]) and comparable average values for RR, DET and TT across conditions. Condition 3 does contain more variability, especially during the earlier identified turbulent moments in the performance. Condition 1 has a slightly higher average DET value indicating more stable synchronization. Fluctuations in DET values from (de)synchronizing are slightly larger in Condition 2 compared to Condition 1.

Test subject's scores on expressiveness, technical content and interaction quality of each audio excerpt are indicated in Figure 3-6, 7 and 8 [A]. Condition 3 received lower scores although differences with the other conditions seem less severe as compared to the self-reported presence scores presented in layer 3 below (mean scores: C1=7.33; C2=7.17; C3=5.5). Three excerpts from Condition 3 did get the remark "I do not understand the intention of the pianists".

In conclusion, performances in each condition were executed relatively well with stable tempos and fluctuating synchronization and RQA measures as instructed by the score. Increased DET fluctuations in Condition 2 were indicative of a very good performance given they are indicative of clearer alternating moments of (de)synchronization. Subjective scores by the test subject were good for Condition 1 and Condition 2 and just above average for Condition 3. Analysis of the performance layer thus shows a good execution of the composition in Condition 1 and 2 and shows more trouble performing successfully in Condition 3. Underlying reasons might be found in the embodied dynamics of coordination and communication that will be discussed in the next layer.

3.4.7.2 Layer 2: Embodied co-regulation

The normalized, principal component of the head movement timeseries of both pianists is shown in Figure 3-6, 7 and 8 [E]. Wavelet coherence on these timeseries is shown in Figure 3-6, 7 and 8 [G]. These plots show maxima at multiples of half the average tempo of the performance (72BPM or 1.2Hz) with a less outspoken pattern for Condition 3. As the score has two beats per measure, it shows how pianists synchronized their head movements coherently with the musical structure. Head movement in the [0.9, 1.5]Hz band showed a flat distribution of relative phase angles between pianists across conditions ([mean, resultant vector length] equalled $[73°, 0.105]_{C1}$, $[-140°, 0.127]_{C2}$, $[82°, 0.070]_{C3}$). Phase shifts in the music might have transferred to head movements and might suggest pianists are mainly keeping time for themselves.

Annotations of head nodding between pianists are shown in Figure 3-6, 7 and 8 [F]. Condition 1 contained cues from Pianist 2 towards the end, Condition 2 contained several synchronized head nods between pianists and Condition 3 shows the absence of communication cues from Pianist 1 towards the virtual agent. Synchronized head nodding in Condition 2 also goes together with a closely coupled gaze and regular RQA measures.

This close coordination was felt by the test subject as she commented after the experiment that she missed the "posture mirroring" of the other pianist in Condition 3 as presented in Condition 2.

Next, one can see differences between conditions for the gaze angle throughout time, between pianists in Figure 3-6, 7 and 8 [D]. Condition 1 has Pianist 1 mainly looking forward as the other pianist sat at an angle of 70°. Figure 3-6 [D] shows a significant decrease of the gaze angle at 300 seconds in Condition 1 just before a longer musical synchronization period. This transition is followed by a head nod of Pianist 2, showing how attention shifted towards Pianist 1. Condition 2 also shows an important transition early in the performance after which Pianist 1 keeps gaze directed towards Pianist 2. This attention shift resulted in a closer coordination as illustrated by synchronized head nods.

Figure 3-6, 7 and 8 [H] shows RPs of the postural sway of Pianist 1 in each condition. At first sight, one can see a phase transition in Condition 1 at 300 seconds. This moment showed a decrease in gaze angle and the start of a musically synchronized period. Looking at postural sway, it shows how the test subject adapted her posture at that time for the remainder of the performance. In addition, one can see the clusters of rectangular recurrence regions that, given a certain delay, correspond to the different bars relatively well. Recurrence values for Condition 3 are regular but smaller on average compared to other conditions. This finding indicates an even spread of recurrences within each point's radius or a uniform noise component for the phase space trajectories.

3.4.7.3 Layer 3: Subjective experience

For the qualitative aspects, we focused on the subjective experience of the test subject after the interaction. The analysis goals were to evaluate how the test subject experienced each performance globally and in specific moments.

Global scores on the interaction from the custom questions showed satisfactory enjoyment for Condition 1 and Condition 2 ([C1,C2,C3]=[7,7,4]), the most natural interaction in Condition 1 ([C1,C2,C3]=[4,1,1]) and interestingly, higher scores for experienced closeness to the musical partner in C2 ([C1,C2,C3]=[4,6,1]). Pianist 1 commented for Condition 2 how *"the VR environment added solely an interesting, fun element, that was almost discarded when the actual playing took place. Because I had a good interaction with my partner, the feeling was very close to the one of a performance that happens in real conditions. The fact that the other pianist was responsive to me was enough to convince me that the situation was real and made me enjoy it thoroughly".* Condition 3 had lowest scores on enjoyment, closeness, and naturalness. Comments of the test subject revealed frustration caused by a non-intentional virtual agent: *"It took me about one to two minutes to realize that my virtual partner was, in fact, not present. … My main focus was on trying to understand the intentions of something that was quite* *obviously not going to follow mine. … as opposed to the second condition, where the fact that I felt the presence of a real person made me connect immediately to an image that was obviously not real, in the third condition I felt almost repulsed by the visual element"*.

The immersive tendencies questionnaire did not detect anomalies in the test subject's profile (involvement=6.38, focus=6.57, games=5.00). Presence and flow scores are shown in Figure 3-9. Flow scores were relatively close across conditions ($mean_{C1}$=6.00, $mean_{C2}$=5.89, $mean_{C3}$=5.67) with slightly less challenge-skill matching and autotelic experience in Condition 3. Presence scores were comparable across conditions as well, with Condition 3 having higher scores on interface quality and involvement and lower scores on immersion. Additional questions in the presence questionnaire (Witmer et al., 2005) had comparable scores across conditions for the "sense of being there" (C1=6, C2=7, C3=7), low scores in "spatial presence" for C2 (C1=6, C2=1, C3=7) and low scores in "similarity to a real place" for Condition 3 (C1=3, C2=5, C3=1).

Pupil dilatations of the test subject are shown in Figure 3-6, 7 and 8 [I]. It remained too challenging to extract meaningful insights or perform comparisons across conditions due to a technical issue in Condition 1 (missing data for 2.5min) and different light conditions across conditions (pass-through camera in Condition 1 while a visualization of the virtual environment in Condition 2 and Condition 3). We do report the normalized data in the figures as an example of a quantitative, physiological method for measuring mental activity in our proposed methodological framework. Participant's comments on their overall experience can be found in the appendix.

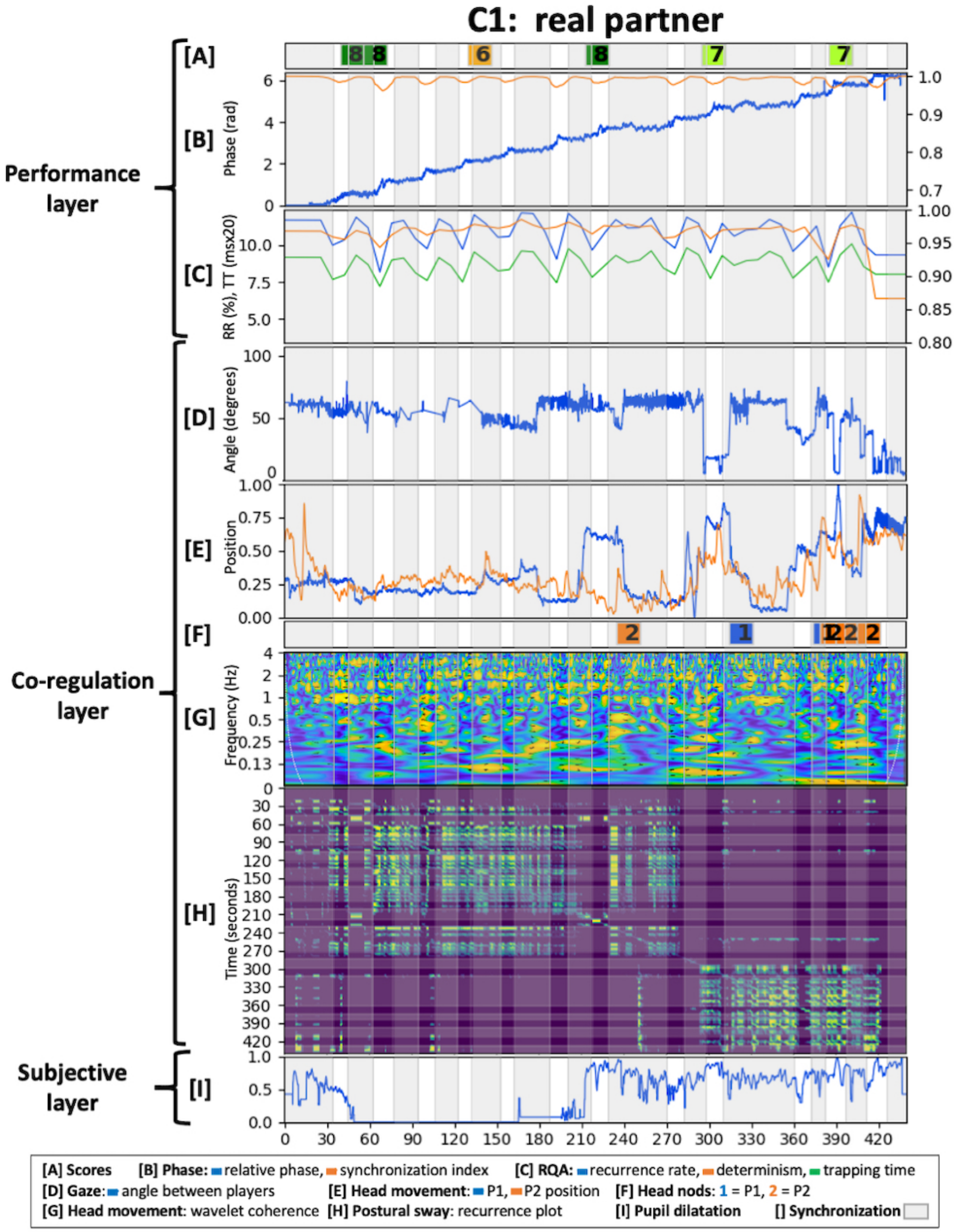


Figure 3-6: Analysis of Condition 1
(P1 = test subject, P2 = real partner)

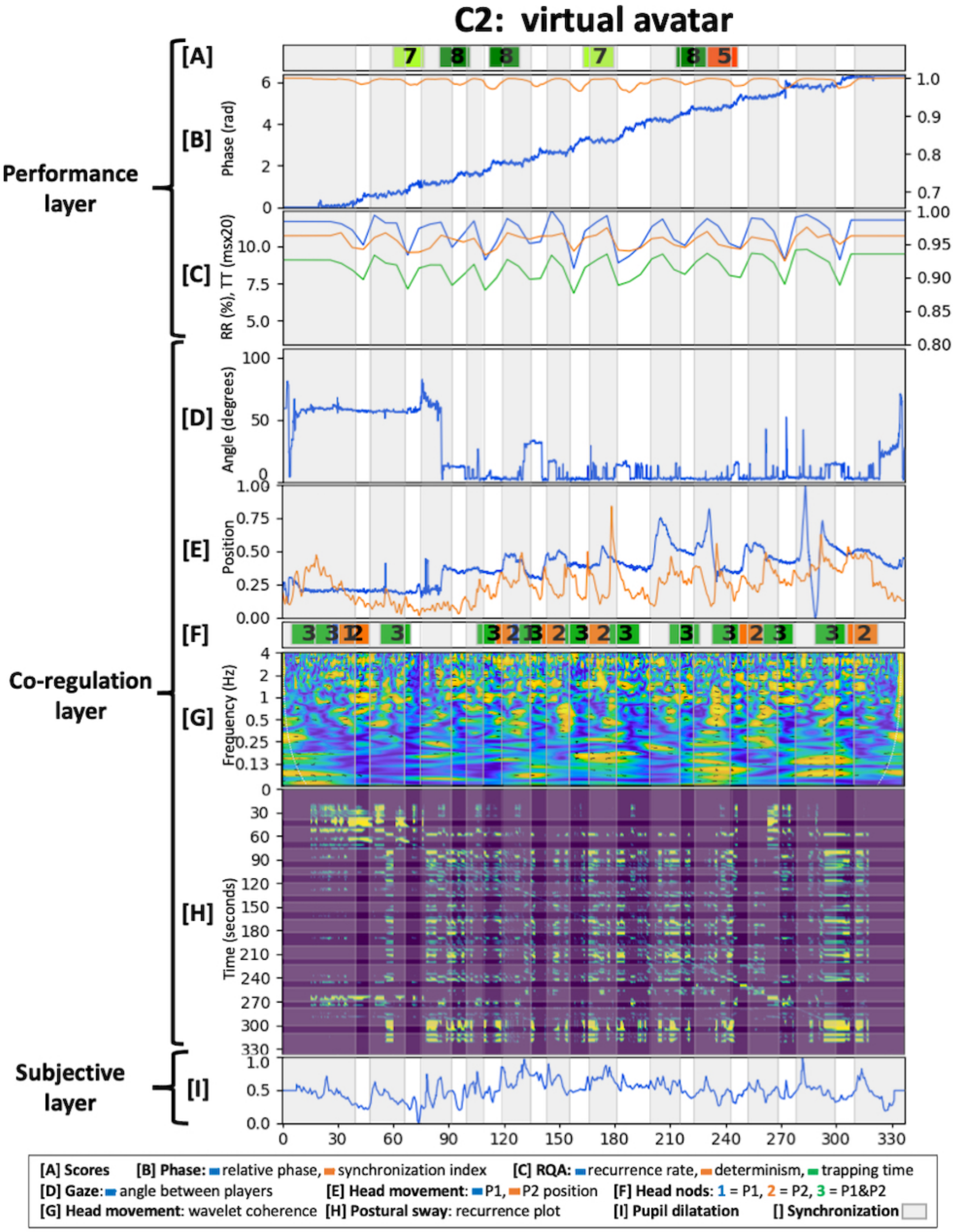


Figure 3-7: Analysis of Condition 2
(P1 = test subject, P2 = virtual avatar)

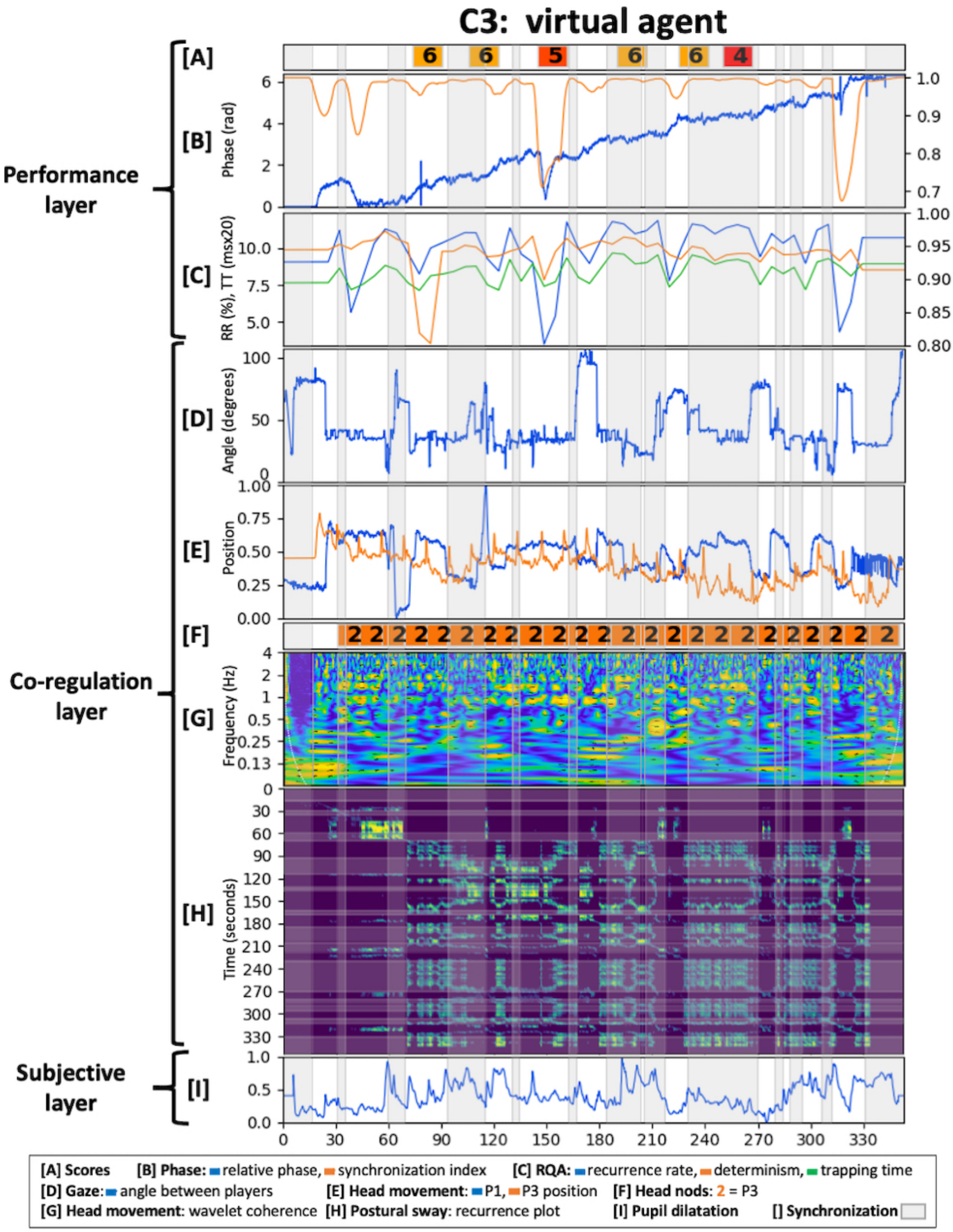


Figure 3-8: Analysis of Condition 3
(P1 = test subject, P3 = virtual agent)

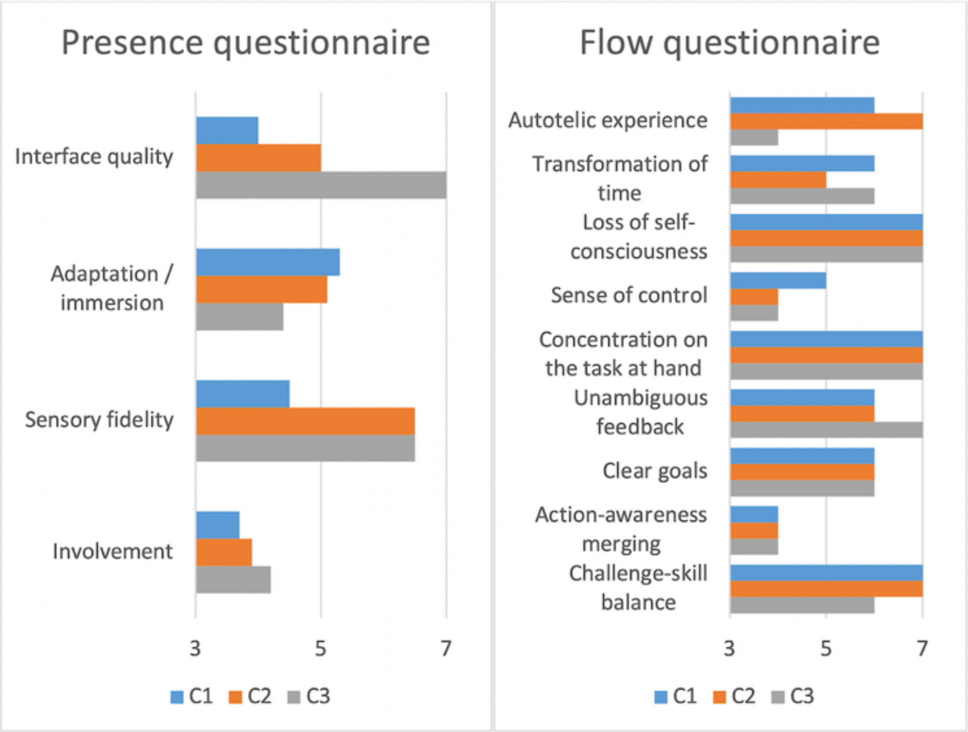


Figure 3-9: Presence and flow scores of Pianist 1 in all conditions

3.4.8 Discussion

Creating a shared context is a delicate process and emerges out of the on-going process of participatory sense-making between closely coupled and coordinated individuals (De Jaegher & Di Paolo, 2007). Once established, 2nd person information in the interaction will have characteristics such as self-directedness, contingency, reciprocity, affective engagement and shared intentional relations (C. Moore & Barresi, 2017). This might have taken place in Condition 2 as it shows an excellent musical performance with close coordination between pianists. Pianists mirror each other's posture, move and de-phase together in time, with joint-actions like synchronized head nods at specific moments in the score. It seems as if both pianists have the interpersonal coordination necessary to co-create the musical structure and perform music as a group, anticipating and adapting to each other effectively (D'Ausilio et al., 2015; A. Walton et al., 2016). Head movement analyses showed synchronized musicians moving along with the metrical time of the score across conditions. Phase angles between these movements drifted during the performance reflecting the changing relative phase resulting from instructed accelerations. Musicians embodied musical time, keeping time for their own stable or accelerating part in the composition and, given the absence of external timekeeping, created time together (A. E. Walton et al., 2018).

A shared context implies the participants are involved in participatory sense-making (De Jaegher & Di Paolo, 2007) with joint-actions and the presence of intentional relations in subjectivity or consciousness, or a form of intersubjectivity (Zahavi, 2001). Taking the second person approach to social understanding requires an understanding of these intentional relations (C. Moore & Barresi, 2017). The relevance of virtual scenarios, avatars and agents for this approach is exemplified by the application of our methodological framework on Condition 2 and Condition 3 to our case-study. Self-reports from the test subject directly referred to "the presence of a real person" about the avatar in C2 and the realization about the agent in Condition 3 "that my virtual partner was, in fact, not present" while she was "trying to understand the intentions of something [the agent]". The comments also stressed the difference between interacting (social presence) as opposed to being with another (co-presence) (Garau et al., 2005; Parsons, Gaggioli, et al., 2017). The other layers in our framework further demonstrate the diffuse intentional relations in Condition 3 by a turbulent coordination and less movement synchronization resulting from a lack of communication between the pianists. As the test subject commented about a perceived "non-intentional agent", it might have been unclear for the test subject when the performance was in a (de)synchronizing measure of the score.

While the agent in Condition 3 had natural movements recorded from real performances, the test subject did not synchronize as well musically and behaviourally as compared to the other conditions. Nevertheless, musical structure was still reflected in the test subject's body sway corresponding to findings from earlier research that showed the transfer of musical structure in test subject's movements (Demos et al., 2018) and possibly indicating individual time keeping. The virtual agent and musical structure of Condition 3 could have been too rigid for the test subject to coordinate effectively, resulting in informationally and behaviourally decoupled musicians. The agent might have lacked adaptive flexibility (A. Walton et al., 2016) and the ability to actively distort or co-determine the musical structure (Doffman, 2009; A. E. Walton et al., 2018).

Flow questionnaire scores were comparable across conditions. The presence questionnaire showed a low sense of spatial presence in Condition 2 possibly resulting from the closer coordination demonstrated by the synchronized head nodding, the gaze angle and self-reports. Higher scores for involvement and interface quality in the presence questionnaire of Condition 3 might have resulted from the frustration of interacting with the non-intentional agent. It might have made the virtuality of the agent become more obvious and caused the test subject to become more (negatively) emotionally involved and become more aware of playing with the piano interface. Musical performance in Condition 3 was not satisfactory for the pianist despite relatively satisfactory relative phase progression and JRQA metrics. On the other hand, scores indicate a successful execution and enjoyment of the performance in Condition 1 and Condition 2.

We argue that the combined successful musical performance, close coupling, and shared intentional relations across different layers between the interacting individuals constitutes the feeling of social presence. Our methodological framework allowed us to frame and couple patterns in these dynamics while adhering to the proposed multi-layered notion of presence (Riva et al., 2004) that is “rooted in activity” (Slater, Lotto, et al., 2009). VR has been the core enabler in this framework through its unparalleled flexibility in controlling stimuli and as an approximation of the ideal, nonmediated interactions we have in real life.

3.5 General discussion and future work

The design and analysis of the case-study based on our proposed methodological framework has allowed us to describe performative, behavioural, and experiential interaction dynamics across real and virtual conditions. The comparison of dynamics from the virtual interactions with the real-life setting has provided the means to evaluate similarities and differences that we argue are needed to confirm the required level of social presence for ecologically valid social cognition research. While music interactions represent a particular setting for the study of social cognition, they are able to create shared contexts with object-centred interactions that involve emotional engagement with joint attention and joint goal-direction actions (C. Moore & Barresi, 2017). As such, their analysis could support the move towards a 2nd person approach to social understanding and help close the gap between 1st person experiential and 3rd person observational approaches (Schilbach et al., 2013).

A first extension of this study would be moving from a case-study to a full experiment involving a large number of participants. One could then move from the descriptive analyses above to statistically substantiated and quantifiable claims about social presence within specific (music) interaction contexts using the methodological framework introduced in this paper. Recording (neuro)physiological signals in the subjective layer using biosensors or a hyperscanning setup (Dumas et al., 2011) would also help to support these claims.

One could vary aspects of the virtual environment as presence modulators to influence the perceived level of realism. An interesting direct modulator would be to blend virtual and real worlds using augmented reality technology. To gain deeper insights into the constitutive aspects of the feeling of social presence, an interesting variation would be experimenting with different performance settings by varying the environment’s acoustics or the inclusion of an audience. For example, one could evaluate the changes in coordination dynamics and experience of the pianists by processing the audio to simulate a dry practice room or reverberant concert hall. With inclusion of an audience, one could vary its engagement (Glowinski et al., 2015).

Besides controlling the (perceived) realism of the environment, one could also modulate the (perceived) realism of the virtual avatar and agents. One could include some form of emotional content by varying facial expressions, include eye-blinking and gaze, have the agent mirror posture of participants, or include behavioural cues at transitional moments of (de)synchronization in the form of head nods. Interaction with the virtual agent might be improved by incorporating elements of surprise and controlled variability. One could script specific actions, blend multiple animations providing richer gesture sets or leverage machine learning techniques to learn new interaction dynamics that balance the exploration and exploitation of possible behaviour states. Specifically, the Kuramoto model used in our case-study could be extended to incorporate richer dynamics and sudden transitions as in the models developed in other research (Mörtl et al., 2014; Shahal et al., 2020; Tognoli et al., 2020).

Another avenue of investigation would be the inclusion of a multi-user VR scenario. Instead of having one test-subject interacting with a real human, avatar, or agent, one could place both pianists in the virtual environment and analyse the dynamics of both participants individually and together. This would require the combination of detailed finger tracking and full body movement as well as low latency, synchronized, audiovisual and tactile content. Multi-user VR scenarios can also be designed as networked performances with investigations into aspects of spatial presence. These are technically challenging but feasible using the technology presented in this paper.

Finally, the methodology presented here is readily available to offer better understanding in existing dynamical theories of action and perception (Warren, 2006), social psychology (Tarr et al., 2018) as well as open new research pathways such as VR based music cognition research and support the investigation of subjective qualities prevalent in musical interactions such as presence, flow, agency and togetherness (G. Herrera et al., 2006; Nijs et al., 2012; Shirzadian et al., 2017). The set-up could be used to transfer existing paradigms in joint-action and amnestic re-embodiment (T. K. Metzinger, 2018; Suzuki et al., 2012) to musical interaction scenarios as well as for applications in education and creative works. The latter was demonstrated by a public performance in our Art-Science and Interaction lab using a modified version of the virtual environment described in this paper[5].

3.6 Conclusion

We introduced a multi-layered methodological framework incorporating qualitative and quantitative methods to assess the feeling of social presence in social music interactions in virtual reality. We then applied this framework on a case-study involving a duet piano

---

[5] A video is available here: https://youtu.be/GlVaMPCotzM (Accessed on the 20th of November 2022)

performance in which an expert pianist played a musical composition with another expert pianist; a human-embodied avatar controlled by an expert pianist; and a computer-controlled agent. The case-study showed excellent performances with close interpersonal coordination in behavioural and experiential layers for interactions with the real pianist and virtual avatar and a good performance without interpersonal coordination for the interaction with the virtual agent. The analyses demonstrated the value of our proposed framework in assessing social presence as well as in highlighting opportunities and challenges in developing better virtual interactions and models of virtual humans.

4.

The virtual drum circle

Submitted and currently in review as: Van Kerrebroeck, B., Crombé, K., Wilain, S., Leman, M., Maes, P. J. (2022). The virtual drum circle: polyrhythmic music interactions in extended reality. *Journal of New Music Research.*

4.1 Introduction

Human joint action can be considered one of nature's most remarkable features, occurring across a myriad of activities in which people coordinate actions towards a shared goal. Ample research has studied joint actions (Sebanz & Knoblich, 2021), deepening insights into their underlying behavioural and neuronal control principles (Koban et al., 2019; Richardson et al., 2015). Action-oriented research has been rewarding for investigations into cognition, stemming forth from the notion that cognition subserves action and is grounded in sensorimotor skills (Engel et al., 2016). For instance, research has shown how coordination through joint action is able to cause prosocial effects such as participatory sense-making (De Jaegher & Di Paolo, 2007) and feelings of shared control or joint agency (Bolt & Loehr, 2017; Loehr, 2022). They are also able to modulate and merge self-other representations (Heggli, Cabral, et al., 2019; Tarr et al., 2014) due to co-activation of action perception networks (Overy & Molnar-Szakacs, 2009). In the last decades, the study of human joint action has been accelerated by the integration of various new technologies, such as motion capture devices, neuroimaging techniques for measuring (inter)brain activity, and more lately, displays and interfaces from the domain of extended reality. More widely, extended reality has been integrated in various scientific disciplines as an innovative methodological tool.

Extended reality settings, encompassing virtual, mixed, and augmented reality, have certain qualities that make them ideally suited for experimental research. For instance, they offer high degrees of experimental control while retaining acceptable and continuously improving levels of realism (Blascovich et al., 2002; Parsons, Gaggioli, et al., 2017; Van Kerrebroeck et al., 2021). In addition, using extended reality technology has created new fundamental questions (T. K. Metzinger, 2018) and offers flexible and new ways of presenting experimental stimuli such as body swapping (De Oliveira et al., 2016), the illusion of backwards time travel (Friedman et al., 2014) and conversing with your future self (Şenel & Slater, 2020). It allows efficient data collection and eases the sampling for, and replication of, experimental work (Blascovich et al., 2002). Extended reality for the arts has enabled new multimodal experiences and non-linear narratives endowing visitors with agency and virtual embodiment (Baker, 2018). Given these traits, it is not surprising how extended reality finds its way into domains such as the social and cognitive (neuro)sciences (Parsons, Gaggioli, et al., 2017), philosophy (T. K. Metzinger, 2018) and the arts (Baker, 2018; Turchet et al., 2021).

Musical expression and interaction in the arts and the sciences have welcomed and experimented with extended reality early-on (Cipresso et al., 2018). Music as a research field is generous, in that it brings together many disciplines. It allows the study of human behaviour, cognition, and perception as well as a focus on engineering and computer science challenges in the musical practices of creating, composing or improvising, interpreting and listening. As such, music in extended reality has attracted a lot of attention (Turchet et al., 2021) and motivated others to outline detailed research agendas (Çamcı & Hamilton, 2020). However, despite its wide and rapid spread, combinations of multi-user setups involving real-time, expressive interactions and low-latency networking are few in number (Hamilton et al., 2011; Pai et al., 2020; Schlagowski et al., 2022) and remain underdeveloped (Turchet et al., 2021).

Music also represents an ideal practice in which joint-action dynamics can be investigated as it involves anticipatory and adaptive processes through fine-grained, bodily coregulation and expressive intentions across multiple timescales (Keller, 2014; Leman, 2007). External rhythms play an important role in these embodied dynamics as shown by the entrainment of dancers' body movements to the musical pulse (Burger et al., 2014; Miura et al., 2013; Miyata et al., 2017; Naveda & Leman, 2010) and the integration of the music's metrical structure within the listeners' body parts (Keller, 2014; Toiviainen et al., 2010). These rhythms can also facilitate interpersonal motor coupling with shared periodic movements, which have been shown capable of predicting coordination (Eerola et al., 2018; Lang et al., 2016).

When musicians interact successfully, they exchange a continuous bidirectional information flow that allows effective coupling into an organic whole with characteristic traits (Demos et al., 2018; A. E. Walton et al., 2018). This coupling has been shown to positively correlate with self-rated "goodness" of performance (Chang et al., 2019) involving embodied dynamics that reflect musical structure and expression (Demos et al., 2018). Investigating joint-action in music performance and interaction could thus help to shed light on the sensorimotor, affective, and cognitive processes facilitating coordination. In addition to music's coregulatory function, it can enhance learning (E. Moore et al., 2017) and induce positive prosocial effects (Stupacher et al., 2017). For instance, synchronizing with music has been shown to lead to a sense of connectedness between people (Demos et al., 2012) with (shared) intentions playing an important role (Goupil et al., 2021). Due to this social and interactive nature of music, music has been described as an embodied language stressing the sense of joint agency induced by the motor actions evoked by sound (Dell'Anna et al., 2021).

When people coordinate, they must rely on informational channels that allow coupling that is neither too rigid nor too loose, allowing exploration and exploitation of emerging dynamics (Kelso, 2009; Warren, 2006). Visual coupling has been shown to increase synchrony between singers (D'Amario et al., 2018; Palmer et al., 2019), people in rocking chairs (Demos et al., 2012), listeners (Dotov et al., 2021; Vuoskoski et al., 2016), pianists (Kawase, 2014), dancers (Chauvigné et al., 2019; De Bruyn et al., 2008) and to reduce variability and individual differences in coordinating duos (Miyata et al., 2017). Visual and auditory coupling can both facilitate this coordinating function with their relative influence potentially depending on context (Chauvigné et al., 2019; De Bruyn et al., 2008; Demos et al., 2012; Dotov et al., 2021; Keller, 2014). Alleviating or adding one sensory channel could lead to compensatory mechanisms (Keller, 2014) and enhanced responses (Chang et al., 2017; Dotov et al., 2021). Due to co-representation of contributions and performance, even the belief that people are acting in another room can influence coordination dynamics (Atmaca et al., 2011; Milward & Sebanz, 2016).

While joint-action research and investigations into dyadic coregulation has made a lot of progress (Sebanz & Knoblich, 2021), a significant limitation of much of this research is that it focuses on unintentional or controlled tapping tasks (Chauvigné et al., 2019; Konvalinka et al., 2010). As music is an intrinsically social activity (S. C. Brown & Knox, 2017), there is a need for investigations that appropriately balance ecological validity and experimental control (D'Ausilio et al., 2015; Dotov et al., 2021). Extended reality technologies promise to fill this methodological gap as well as offer new pathways for experimental research (Kothgassner & Felnhofer, 2020; T. K. Metzinger, 2018). They promise to offer social presence (Van Kerrebroeck et al., 2021), virtual embodiment (Kilteni, Groten, et al., 2013), perspective taking (F. Herrera et al., 2018), unique affordances (Berthaut, 2020) and prosocial attitudes towards virtual characters (Tarr et al., 2018). However, care must be taken in the design of virtual settings such as for the requirement to animate virtual characters with human-like motion data (de Borst & de Gelder, 2015; Pan & Hamilton, 2018). As extended realities technologies are still improving on ecological validity, it remains unclear whether these new technologies can mediate the rich and fine-grained expressions and intentions inherent to musical interactions (Keller, 2014; Leman, 2007).

This study aims to investigate the potential of extended reality technologies for experimental, musicological research. It operationalizes this aim by investigating to what extent visual coupling in the form of a virtual partner influences embodied coregulatory patterns in a rhythmic, dyadic task in extended reality. Since music has been shown to influence coordination and increase prosocial effects in a controlled setting (Demos et al., 2012; Stupacher et al., 2017), this study also compares the impact of music on coordination and experiential dynamics to the impact of a metronome. This paper starts with a description of the experimental design including the experimental task and protocol. The next section describes the extended reality system developed to perform this research together with the quantitative methods used to perform data analysis. This is followed by a section with experimental results presented using a framework describing interpersonal dynamics in a performative, embodied coregulation and experiential layer (Van Kerrebroeck et al., 2021). Results are interpreted in the discussion section using the embodied music cognition (Leman, 2007) and coordination dynamics (Kelso, 2009; Warren, 2006) frameworks together with suggestions for future work.

4.2 Research question

The main research question of this study was to what extent audiovisual coupling influences interpersonal coordination dynamics of two people engaged in a musical activity (See Figure 4-1). More specifically, the goal was to see to what extent visual coupling and the auditory context can facilitate a successful performance, effective embodied coregulation and positive prosocial effects. Visual coupling was varied using the visibility of the partner across not-seeing, seeing-as-avatar, and seeing-as-real levels (partner realism). The auditory context was varied by accompanying the performance either by a metronome or a polyrhythmic backing track (musical background). See Figure 4-2 for an overview of the experimental design.

A first hypothesis was that the polyrhythmic backing track would help with anticipatory and adaptive sensorimotor processes of participants as it was more information rich than the metronome. Participants would thus show a more precise performance of smaller onset asynchronies and smaller variability for the conditions with the polyrhythmic backing track. As the guiding, visual stimulus disappeared during successful coordination periods (see Task below), the second hypothesis was that participants had to compensate for this loss of support by relying more on their bodily coordination. With increased partner realism, there should thus be an increase in shared periodic movements at the common pulse and, potentially, periodic movements at the polyrhythmic tempi across participants. A final hypothesis was that increased partner realism and music as background would be motivational for players and increase prosocial effects leading to higher scores of (shared) agency, self-other merging, and flow.

Figure 4-1: Participants in the seeing-as-real condition

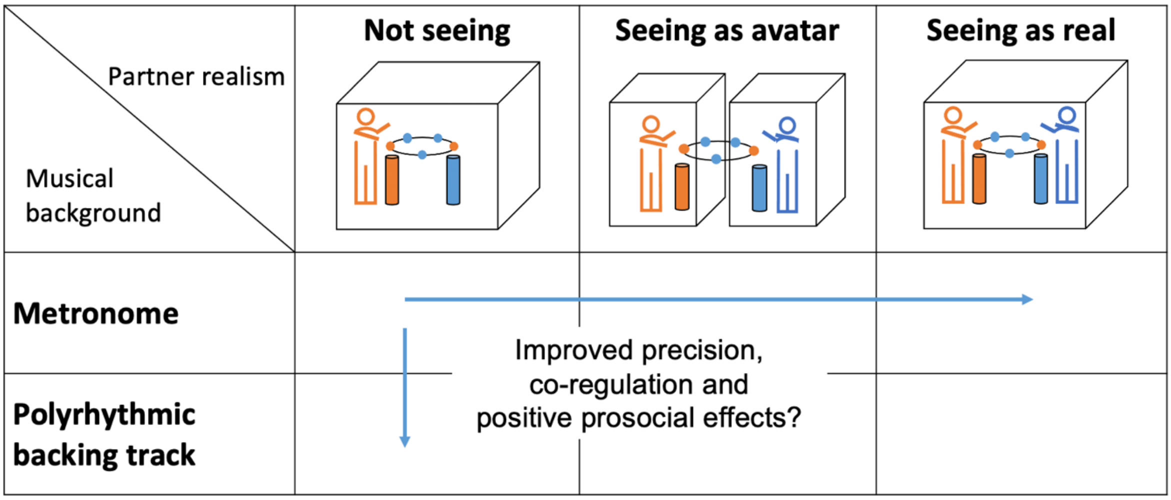


Figure 4-2: Experimental conditions with the "partner realism" and "musical background" factors and hypotheses

4.3 Method

4.3.1 Participants

16 dyads (13 women, 19 men, $[M, SD]_{age}$ = [30.2, 8.4] years, range: 18-48 years) with varying levels of musical expertise (musical training scores from the Musical Sophistication Index (MSI) (Müllensiefen et al., 2014), [M, SD] = [4.1, 1.9], range = 1-6.57) were recruited to participate in the experiment. Participants in each dyad were recruited together and knew each other beforehand to make them feel more at ease interacting with the avatar of their musical partner. There were no restrictions on gender, age, and relationship. The experimental protocol was reviewed and approved by the ethical commission of Ghent University, Belgium. Covid pandemic restrictions in effect were social distancing to 1.5 meter and the wearing of face masks when moving between labs.

4.3.2 Materials and apparatus

The experiment took place in the Art and Science Interaction Lab (ASIL) of Ghent University, Belgium. The technical set-up for this experiment was extensive and an overview is shown in Figure 4-3. Motion capture and avatars were used to investigate the full-body interactions between players. A virtual drum circle, drum pads and questionnaires were used to investigate performative and experiential aspects. All are listed below.

Motion capture: Participants wore full-body motion capture suits with 42 markers. Participants were tracked using independent Qualisys set-ups in two separate rooms (about 7 m distance from one another).

Extended reality – embodied avatars: Qualisys software was used to render real-time, human-controlled, avatar visualizations in a Unity server application that was synchronized to two Unity client applications using the Photon networking framework. These avatars were then streamed wirelessly over WiFi to two HoloLens 2 head-mounted displays for both participants. See Figure 4-4 for a first-person view of the avatar visualisation.

Extended reality - drum circle: A virtual drum circle with rotating spheres was rendered in Unity and displayed in both participant's HoloLenses. Note onset times of participants were compared to the timings of spheres passing in front of participants using a Max for Live application. See Figure 4-4 for a first-person view of the rotating spheres and drum circle.

Drum pads: Participants triggered sounds by tapping on drum pads that sent MIDI notes to Ableton software which rendered them in a percussive sound[6]. Drum pads were custom built and registered note onsets using a pressure sensor connected to a Teensy 3.2 microcontroller. Drum pads were approximately 17 cm in diameter and set at waist height (see Figure 4-1).

Questionnaires: Self-reports of participants were taken using Microsoft Forms questionnaires upon arrival and after each trial. They were asked to indicate whether they had received percussion training in the past, the length and intensity of the dyad's relationship, evaluate their relationship with the self-other integration (SOI) questionnaire (Aron et al., 1992) and fill in the MSI questionnaire before the experiment. Agency, shared agency, SOI and absorption scores from the Flow questionnaire (Engeser & Rheinberg, 2008) were recorded after each trial. Scales are presented in the analysis framework section below.

---

[6] See https://youtu.be/S7LjGwHRqyY for drum sounds of participants with polyrhythmic backing track

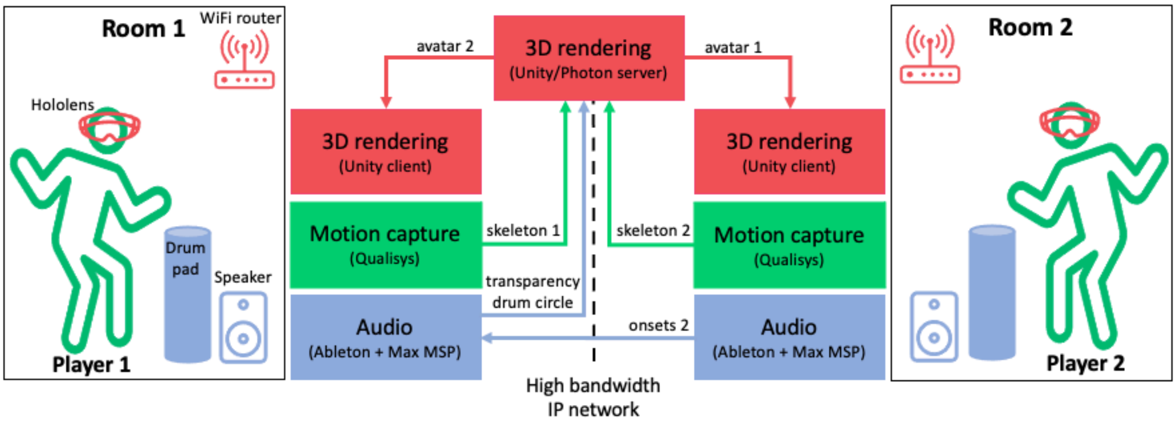


Figure 4-3: The extended reality system for real-time, network, full-body interaction (red = 3D rendering, green = motion tracking, blue = audio)

Figure 4-4: View on the other player as human-controlled avatar and the drum circle as rhythmic stimulus

4.3.3 Task

Dyads were instructed to tap a 2:3 polyrhythmic pattern together at a fixed tempo. Per dyad, one participant was instructed to tap the binary part of the polyrhythm (Inter Onset Interval (IOI) = 1059ms, further referred to as the "binary task"), while the other was instructed to tap the ternary part (IOI = 706ms, further referred to as the "ternary task"). Hence, the smallest unit (tactus) had a duration of 353ms, while the common pulse occurred every 2118ms.

The metronome consisted of a bell sound at pulse level that was also present in the polyrhythmic backing track. The backing track additionally consisted of the standard bell pattern (Agawu, 2006) with other percussive sounds playing on the tactus level[6]. Each condition lasted 242 seconds so participants could perform the polyrhythmic 2:3 pattern a maximum of 114 times.

An augmented reality visual stimulus in the shape of a drum circle was shown in every condition to assist with drumming (see Figure 4-4). The isochronous sequences for each participant were indicated as instructions by rotating virtual spheres on the drum circle and participants were allowed to improvise by skipping taps. When both participants successfully drummed the polyrhythmic pattern, meaning both players tapped within 62.5ms before or after their instructed tap, the stimulus increased in transparency in 5 incremental steps. When they made mistakes, the stimulus gradually returned.

The task was designed to be challenging enough to lead to alternating successful and unsuccessful periods of coordination in tapping the polyrhythmic pattern. The freedom to improvise was included to add aspects of playfulness and expressivity as well as to render the task more challenging for participants with musical experience. The disappearing instruction was designed to induce a form of reward and punishment according to reinforcement learning principles.

4.3.4 Experimental design

Dyads performed polyrhythmic music interactions in six conditions (see Figure 4-2). Experimental factors were partner realism with 3 levels (not-seeing, seeing-as-avatar, seeing-as-real) and musical background with 2 levels (metronome, polyrhythmic backing track). We measured performative, embodied coregulatory and experiential outcome variables that are detailed in the Analysis framework section below. Conditions were randomized with the constraints of not-seeing and seeing-as-avatar as well as both the metronome and polyrhythmic back track conditions always performed together to decrease technical overhead in switching setups. An example order was SarMu-SarMe-SaaMe-SaaMu-NsMu-NsMe (labels are abbreviated from condition names, see Figure 4-2).

4.3.5 Procedure

The experiment lasted two hours. Upon arrival, health and safety rules were explained, and participants received a verbal explanation of the experiment, and gave their informed consent to participate in the experiment. They were then asked to change into motion capture suits and brought to one of two labs. After 45 minutes of motion tracking calibration and virtual skeleton building, both participants performed 2 minutes of their (binary or ternary) part of the task individually with a metronome and as instructed by the virtual drum circle. Next, while still being physically separated in different rooms, they were brought together virtually as human-controlled avatars to familiarize themselves with the real-time virtual animations. After a brief exploration during which participants were encouraged to wave to each other and walk around, they performed another 2 minutes of the task together and with the polyrhythmic backing track as musical background. After these familiarizations, they were physically brought together to exchange some brief impressions and upon which they started the six conditions.

4.3.6 Analysis framework

The analysis framework is based on the earlier work of Van Kerrebroeck, Caruso, & Maes (Van Kerrebroeck et al., 2021). It consists of three interrelated layers containing performative, expressive, and experiential dynamics between interacting players. The framework contains both qualitative self-reports and quantitative measures captured in real-time that are detailed below. A core feature of the framework is the comparison of interaction dynamics from a simulated, virtual context with the corresponding real-life context.

Layer 1: Performance output: The first layer was used to evaluate performance in relation to the goal prescribed by the musical task. The analysis was done using the BListener algorithm, a multivariate tracker of IOIs that uses Bayesian inference to predict timing constancy as a prediction error (Leman, 2021). It can deal with multiples of a basis IOI and is thus useful for dealing with player's improvisations that skip one or multiple tapping onsets in this study. Moreover, global features of timing constancy have shown to correlate with subjective estimates of performance quality and agency (Leman, 2021) and thus allow the linking of insights from this layer to the subjective layer below. Blistener initiates several trackers at the start of each performance that predict incoming IOI-data, which are continuously updated for each player and their instructions. Prediction errors of trackers are then summed over time leading to one global "prediction error" measure per trial for the joint performance of both players and for individual performances of player and (binary or ternary) instruction. BListener parameters, trackers for one trial and model parameters are presented in the appendix below. For more details about the BListener algorithm, the reader is referred to (Leman, 2021).

Layer 2: Embodied coregulation: This layer aimed at evaluating and comparing the embodied dynamics across conditions that players used to coordinate successfully. Movement data were used to compute postural position timeseries and analysed using a Quantity of Motion (QoM) measure (Gonzalez-Sanchez et al., 2018) and wavelet coherence value (Grinsted et al., 2004). The instantaneous postural position was calculated by projecting the centre back marker of each player on the axis connecting the left and right foot markers. The QoM was calculated as the sum of all differences of consecutive samples of the postural position, that is, the first derivative of the position time series (Gonzalez-Sanchez et al., 2018). Wavelet coherence was calculated on concatenated postural sway timeseries of each player. Only parts of timeseries that had a transparent stimulus were kept to reduce the variability between players and focus the analysis on the good interaction bouts. A coherence value per trial was then obtained by summing respective coherence values in 0.2Hz frequency bands centred around the common pulse, the binary and ternary rhythms present in the instruction (common pulse: 2188ms or 0.472Hz, binary: 706ms or 1.416Hz, ternary: 1059ms or 0.944Hz).

Layer 3: Subjective experience: The goal of this layer was to evaluate subjectively experienced interaction qualities and link mental states, (expressive) intentions and meaning attributions to observations in other layers. Self-reports from questionnaires inquired about the length of participants' relationship (5-point Likert scale with 1=Never, 2=Less than once per month, 3=1-3 times per month, 4=1-3 times per week, 5=More than 3 times per week), their SOI score, whether participants were trained in percussion, and the musical engagement and training factors from the MSI. Another questionnaire was taken after each trial and inquired about perceived quality of the performance, feelings of (shared) agency, SOI, and flow. Agency and shared agency scores were recorded using 7-point Likert scales as response to the question "To what extent did you feel agency/shared agency over the musical performance" (1 = Not at all, 4 = Partly, 7 = Very much). The post-session questionnaire can be found in the appendix.

Statistical models: Outcome measures from each layer were evaluated using linear mixed effect (LME) and cumulative link mixed (CLM) models. Variables and their interactions were progressively added to a null model based on whether they significantly improved the model fit. Variables are listed in Table 4-1 and final models in Table 4-2. Models in layer 1 and 2 included the dyad as a random effect while outcome measures from questionnaires in layer 3 included the individuals as a random effect. Models' intercepts corresponded to the player with the ternary task, not seeing the other binary player, in the first trial, playing with the metronome background and having lowest scores for self-reported scores. Outliers in the data were detected using 1.5 times the IQR as metric as well as after model fitting using residual quantile plots with two-sided outlier tests (α = .025). The removed outliers are reported below.

Data processing and analysis: Timeseries trimming and alignment was done using Python. Wavelet coherence values reported below were calculated using Matlab. Other outcome measures such as the prediction error and quantity of motion were calculated using R and Matlab. Statistics were performed in R using the lme4 package (Bates et al., 2015) for the linear mixed effect models and the ordinal package (Christensen, 2015) for the cumulative link models. Tabular reports in the appendix were produced using the sjPlot package (Lüdecke, 2014) with analyses of variance, deviance and contrasts done using the Anova and the emmeans functions. Five trials from three dyads were excluded due to missing data in one trial and a misunderstanding of the task in the others.

Type	Predictor	Levels
Experimental conditions	partner_realism	not-seeing, avatar, real
	musical_background	metronome, music
Experimental characteristics	task	ternary, binary
	trial_count	1 to 6
Self-reports	relation_frequency	5-point Likert scale
	relation_length	number of months
	percussion	0 or 1
	training	1 to 7
	engagement	1 to 7
	SOI	1 to 7

Table 4-1: Experimental model predictors

	Outcome measures	Model
Layer 1	prediction_error (individual performance)	1 + (partner_realism + musical_background)*task + trial_count + training + relation_frequency + (1|dyad)
	prediction_error (joint performance)	1 + partner_realism + musical_background + trial_count + SOI + engagement + training + (1|dyad)
Layer 2	QoM	1 + partner_realism*musical_background*task + trial_count + relation_length + SOI + (1|dyad)
	coherence	1 + partner_realism + musical_background + trial_count + training
Layer 3	agency	1 + partner_realism + musical_background + trial_count + (1|individual)
	shared_agency	1 + partner_realism + trial_count + percussion + relation_length + (1|individual)
	SOI	1 + partner_realism + trial_count + training + (1|individual)
	flow	1 + partner_realism + musical_background + SOI + relation_frequency + relation_length + (1|individual)

Table 4-2: LME and CLM models in each methodological layer

4.4 Results

This section reports the quantitative results from LME and CLM model fitting. Model statistics are presented for each methodological layer together with figures of outcome measure means, standard errors (SE), and model fits across levels of partner_realism and musical_background. A summary of all significant effects across layers is given in Table 4-3. Model parameters, confidence intervals and corresponding statistics are reported in the appendix.

4.4.1 Layer 1: Performance output

Two statistical analyses were done on the prediction errors obtained from the BListener algorithm: one focused on individual players with their instruction and one on the joint performance between players.

Individual performance: Eleven out of 182 data points (coming from 16 dyads performing in six conditions minus five excluded trials) were removed as outliers. Prediction error data were log-transformed to assure normally distributed model residuals. Exponentiated estimates and CIs are presented below. The model explained 63% of the variance with a majority portion explained by the fixed effects (conditional R^2 = 0.628, marginal R^2 = 0.442). There was a main effect of task ($\chi^2(1) = 44.2$, $p < .001$) and musical background ($\chi^2(1) = 17.5$, $p < .001$). Pair-wise comparisons revealed a significant difference between musical backgrounds for the binary (t(146) = 4.172, $p < .001$) but not for the ternary ((t(146) = 1.794, $p = .075$) task. Task (ß = 1.78, CI = [1.43, 2.22], t(151) = 5.22, $p < .001$) and training were significant predictors in the model (ß = 0.941, CI = [0.885, 1.00], t(151) = -198, $p = .050$), while musical background was not ($p = .075$). Trial count ($\chi^2(5) = 21.8$, $p < .001$) and relation frequency ($\chi^2(4) = 30.2$, $p < .001$) showed significant effects with prediction errors linearly decreasing with trial count (ß = 0.799, CI = [0.700, 0.912], t(151) = --3.35, p = .001) and linearly increasing with relation frequency (ß = 1.93, CI = [1.36, 2.73], t(151) = 3.73, $p < .001$). The 5th degree term of trial count (ß = 1.23, CI = [1.08, 1.40], t(151) = 3.07, $p = .003$) and the quadratic term of relation frequency (ß = 1.35, CI = [1.09, 1.68], t(151) = 2.78, $p = .006$) were significant as well.

Joint performance: Three out of 91 data points were removed as outliers. Prediction error data were log-transformed. The model explained 67% of the variance (conditional R^2 = 0.669, marginal R^2 = 0.375). Musical background ($\chi^2(1) = 6.60$, $p = .010$) and engagement ($\chi^2(1) = 5.33$, $p = .021$) were revealed as significant effects with the polyrhythmic backing track (ß = 0.801, CI = [0.674, 0.951], t(74) = -2.57, $p = .012$) and more engagement (ß = 0.769, CI = [0.612, 0.965], t(74) = -2.31, $p = .024$) decreasing the prediction error. Trial count was revealed as a significant main effect ($\chi^2(5) = 16.3$, $p = .006$) with prediction errors decreasing linearly with trial count (ß = 0.762, CI = [0.617, 0.940], t(74) = -2.57, p =

.012) together with a positive 5th degree term (ß = 1.35, CI = [1.10, 1.66], t(74) = 2.88, *p = .005*).

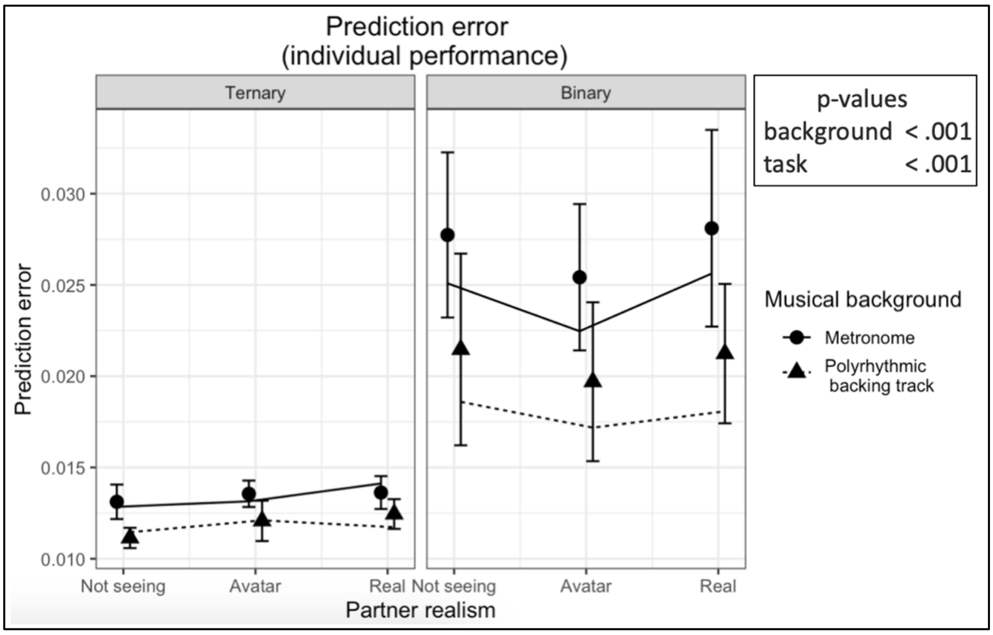


Figure 4-5: Prediction error means, standard errors, and LME model fits for the individual performance across partner realism and musical background

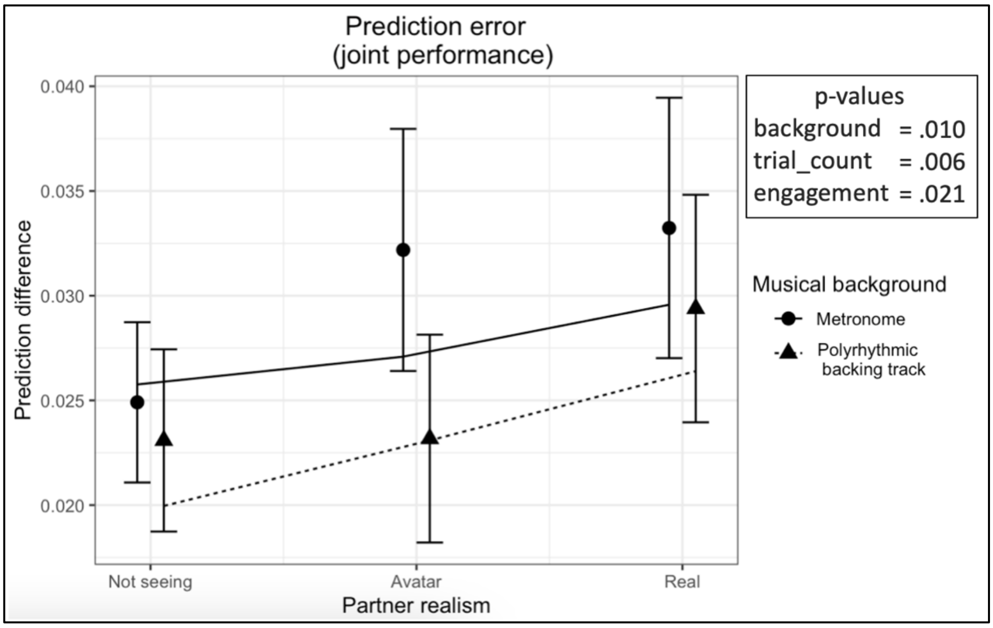


Figure 4-6: Prediction error means, standard errors, and LME model fits for the joint performance across partner realism and musical background

4.4.2 Layer 2: Embodied coregulation

In this layer, one analysis was performed on the QoM of players individually and one on the wavelet coherence value of their joint performance. For the latter, only bandpower for the ternary rhythm was retained as mean bandpower for the common pulse (M = 0.19, SE = 0.003) and binary rhythm (M = 0.21, SE = 0.004) were significantly lower ($p < .001$) than ternary bandpower (M = 0.28, SE = 0.006). In addition, the binary bandpower did not significantly differ from power in bands at four (M = 0.21, SE = 0.003, $p = .907$) and five (M = 0.22, SE = 0.003, $p = .360$) times the tactus frequency.

QoM: The model explained 55% of the variance with an equal portion explained by the fixed and random effects (conditional R^2 = 0.545, marginal R^2 = 0.274). There were significant main effects for partner realism ($\chi^2(2) = 11.2$, $p = .004$), musical background ($\chi^2(1) = 16.0$, $p < .001$), relation length ($\chi^2(1) = 8.03$, $p = .005$) and SOI ($\chi^2(5) = 17.2$, $p = .004$) as well as interaction effects between partner realism and musical background ($\chi^2(2) = 12.3$, $p = .002$), partner realism and task ($\chi^2(2) = 7.56$, $p = .023$), musical background and task ($\chi^2(1) = 5.81$, $p = .016$) and partner realism, musical background and task ($\chi^2(2) = 8.99$, $p = .011$). Post-hoc tests revealed a significant difference between not seeing and seeing as avatar (t(145) = -2.41, SE = 0.039, $p = .045$) and not seeing and seeing as real (t(146) = -3.31, SE = 0.043, $p = .003$) although these results remain difficult to interpret due to interaction of partner realism with task and musical background. The polyrhythmic backing track significantly increased the QoM (ß = 0.172, CI = [0.016, 0.328], t(157) = 2.18, $p = .031$) while the relation length decreased it (ß = -0.003, CI = [-0.005, -0.001], t(157) = -2.83, $p = .005$).

Coherence: Three out of 91 data points were removed as outliers. Wavelet coherence data were log-transformed to assure normally distributed model residuals. As including random effects did not improve model fit ($p = .5687$), they were not included in the model. The fitted linear model explained a significant and substantial portion of variance (R^2 = 0.298, F(9, 78) = 3.67, $p < .001$, adjusted R^2 = 0.217). Training (F(1, 78) = 6.12, $p = .016$) was a significant main effect, and musical background (F(1, 78) = 3.88, $p = .052$) and trial count (F(5, 78) = 2.27, $p = .055$) were almost significant. Training increased the coherence (ß = 1.05, CI = [1.01, 1.09], t(78) = 2.47, $p = .016$) just like seeing as real (ß = 1.12, CI = [1.00, 1.24], t(78) = 2.07, $p = .042$), while a polyrhythmic backing track decreased it (ß = 0.924, CI = [0.853, 1.01], t(78) = -1.97, $p = .052$).

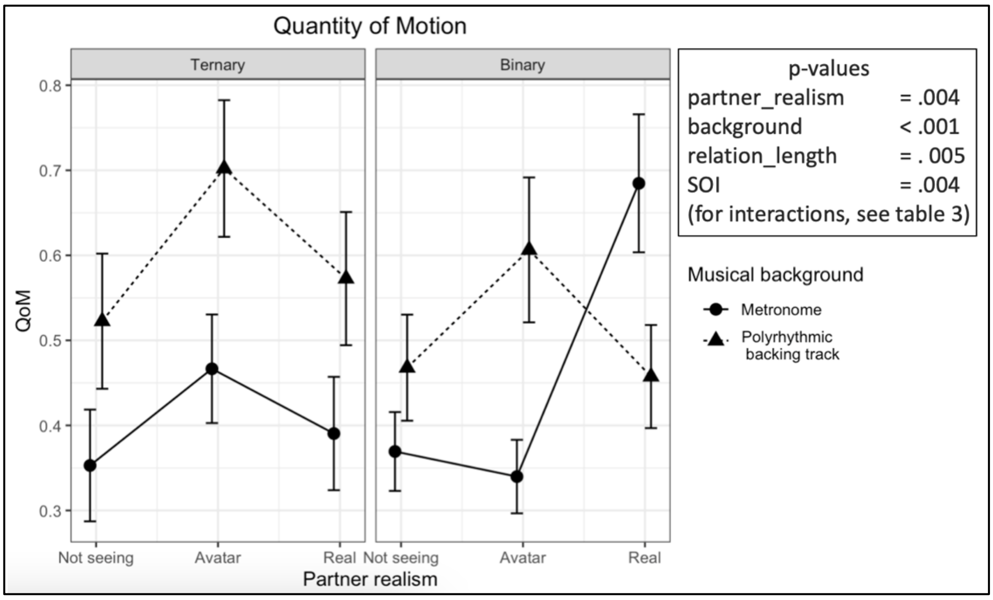


Figure 4-7: Quantity of motion means, standard errors, and LME model fits of players in the ternary and binary task across partner realism and musical backgrounds

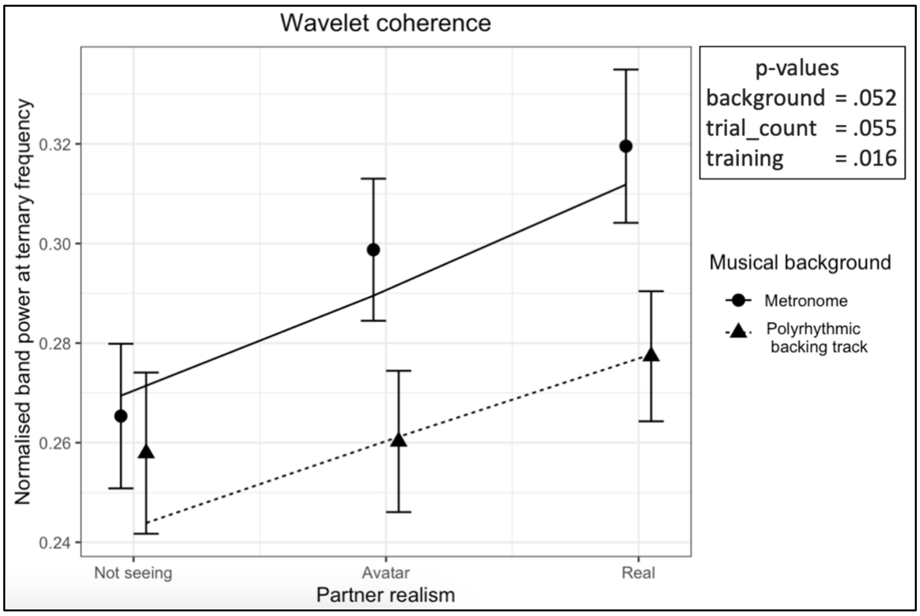


Figure 4-8: Wavelet coherence in the ternary rhythm band means, standard errors and LM model fits across partner realism and musical backgrounds

4.4.3 Layer 3: Subjective experience

This section presents the results from participant self-reports retrieved from questionnaires. Figure 4-9, 10, 11, 12 below show model predictions of the agency, shared agency, SOI, and flow outcome measures.

Agency: The model explained 47% of the variance (conditional R^2 = 0.469, marginal R^2 = 0.165) and revealed a main effect of musical background ($\chi^2(1) = 12.0$, *p < .001*) and trial count ($\chi^2(5) = 34.0$, *p < .001*). Specifically, participants were 2.75 times more likely to report higher scores of agency when playing with the polyrhythmic backing track (odds-ratio (OR) = 2.75, CI = [1.54, 4.92], *p = .001*). Agency increased linearly (OR = 5.86, CI = [2.84, 12.1], *p < .001*) and decreased in the 5th degree (OR = 0.337, CI = [0.168, 0.680], *p = .002*) with trial count.

Shared agency: The model explained 39% of the variance (conditional R^2 = 0.388, marginal R^2 = 0.228). Main effects were revealed for partner realism ($\chi^2(2) = 16.6$, *p < .001*) and trial count ($\chi^2(5) = 26.4$, *p < .001*) with weaker effects revealed for percussion ($\chi^2(1) = 6.03$, *p = .014*) and relation length ($\chi^2(1) = 4.15$, *p = .042*). The seeing as avatar (OR = 3.12, CI = [1.55, 6.27], *p = .001*) and seeing as real (OR = 4.06, CI = [1.90, 8.71], *p < .001*) conditions were respectively 3.12 and 4.06 times more likely to result in higher shared agency scores compared to not seeing. Playing percussion also increased the odds of reporting higher feelings of shared agency (OR = 2.77, CI = [1.26, 6.07], *p = .011*) just like the trial count (OR = 4.83, CI = [2.37, 9.82], *p < .001*), while the relation length slightly decreased it (OR = 0.99, CI = [0.972, 0.999], *p = .036*).

SOI: The model explained 69% of the variance (conditional R^2 = 0.694, marginal R^2 = 0.215). Main effects were revealed for partner realism ($\chi^2(2) = 20.8$, *p < .001*), trial count ($\chi^2(5) = 48.9$, *p < .001*) and a weak effect for training ($\chi^2(1) = 3.61$, *p = .057*). Both seeing as avatar and seeing as real significantly increased the odds of reporting higher SOI compared to not seeing (avatar: OR = 2.32, CI = [2.1.15, 4.66], *p = .019*; real: OR = 6.49, CI = [2.83, 14.9], *p < .001*). Training was a significant, positive predictor in the model (OR = 1.53, CI = [1.02, 2.29, *p = .038*). Trial count had a strong, positive effect for its linear term (OR = 10.43, CI = [4.83, 22.5], *p < .001*) and weaker positive and negative effects respectively for its 4th (OR = 2.71, CI = [1.275, 5.77], *p = .010*) and 5th degree (OR = 0.476, CI = [0.241, 0.941, *p = .033*) terms.

Flow: Five out of 182 data points were removed as outliers. The model explained 54% of variance (conditional R^2 = 0.541, marginal R^2 = 0.297). Musical background ($\chi^2(1) = 8.46$, *p = .004*), prior SOI score ($\chi^2(5) = 14.3$, *p = .014*) and relation length ($\chi^2(1) = 6.02$, *p = .014*) were significant main effects revealed by the model. The flow score increased with a polyrhythmic backing track (ß = 0.035, CI = [0.011, 0.059], *p = .004*) and with increased relation length (ß = 0.001, CI = [0.0002, 0.002], *p = .015*). Quadratic (ß = 0.109, CI = [0.033, 0.184], *p = .005*) and cubic (ß = 0.077, CI = [0.015, 0.138], *p = .016*) terms for the prior SOI score were significant predictors in the model.

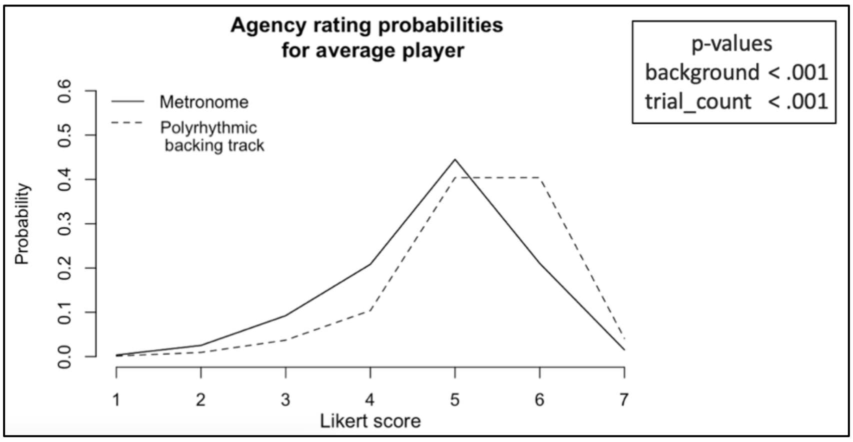


Figure 4-9: Agency rating probabilities for an average player across musical backgrounds

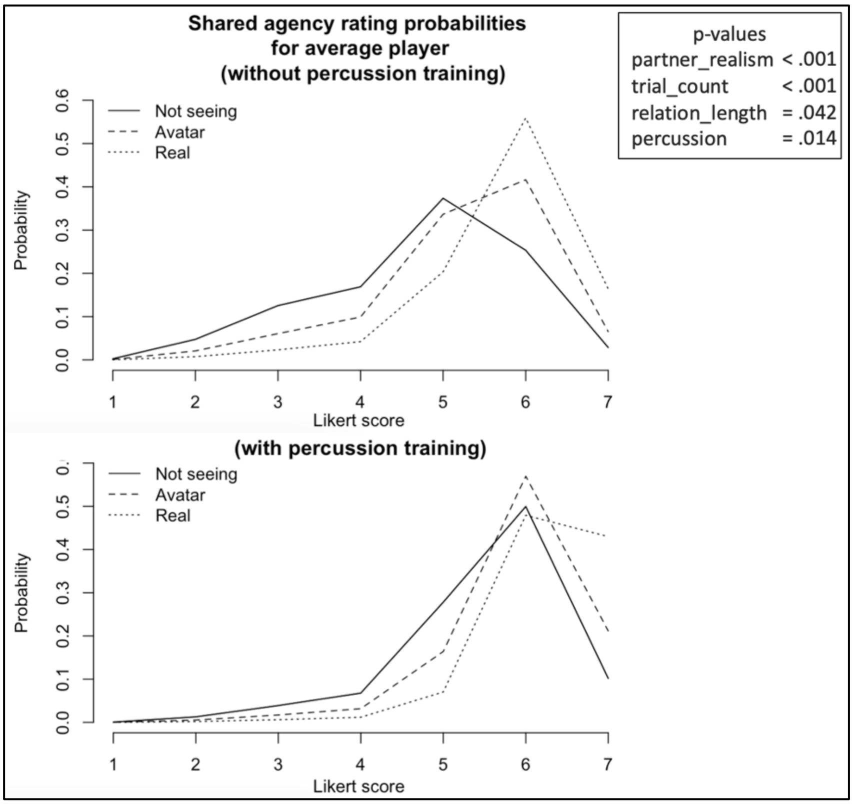


Figure 4-10: Shared agency rating probabilities for an average player with and without percussion training across partner realism

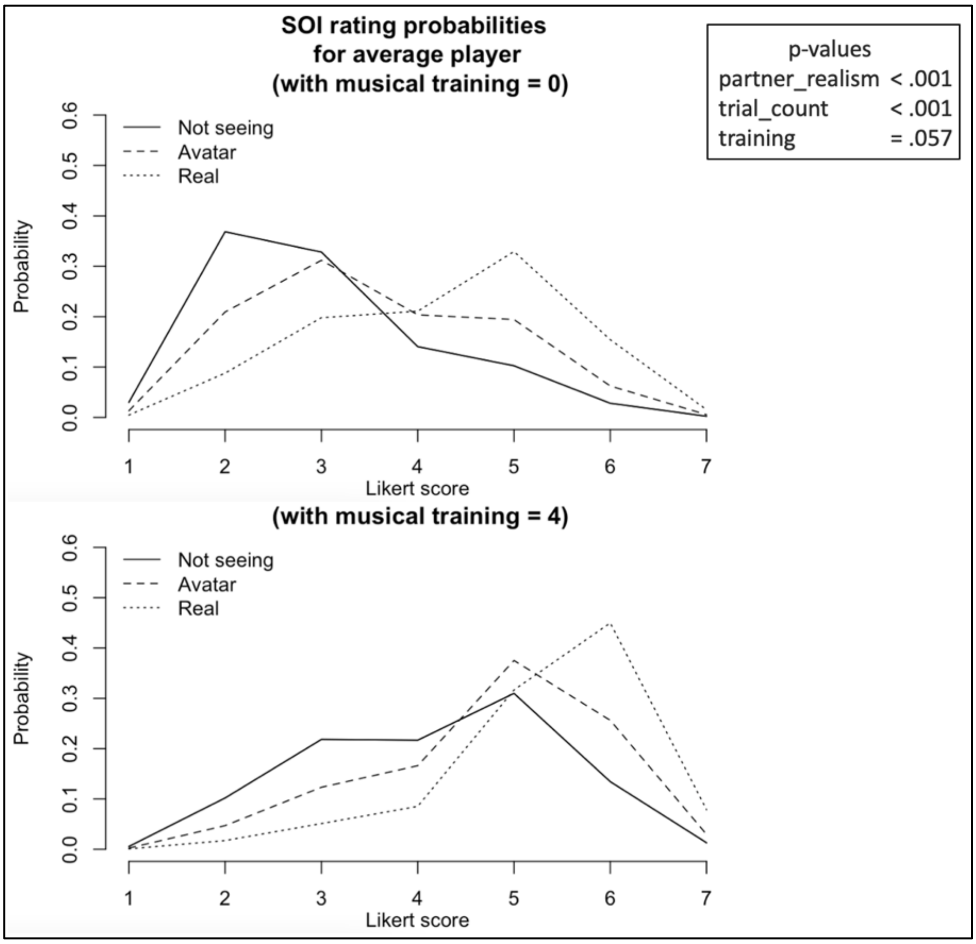


Figure 4-11: Self-other merging rating probabilities for an average player with a musical training equal to 0/7 or 4/7 across partner realism

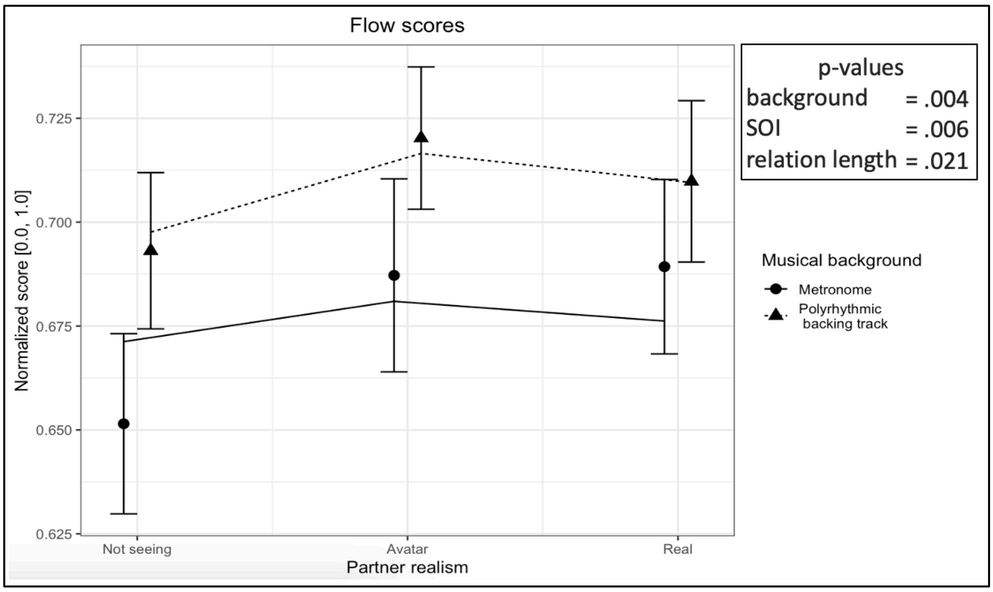


Figure 4-12: Flow scores, standard errors, and LME model fits for all players across partner realism and musical backgrounds

	Outcome measure	**Predictor**	**Significance**
Layer 1	prediction_error (individual performance)	musical_background	< .001
		task	< .001
	prediction_error (joint performance)	musical_background	.010
		trial_count	.006
		engagement	.021
Layer 2	QoM	partner_realism	.004
		musical_background	< .001
		relation_length	.005
		SOI	.004
		partner_realism: musical_background	.003
		partner_realism: task	.023
		musical_background: task	.016
		partner_realism: musical_background: task	.011
	coherence	musical_background	.052
		trial_count	.055
		training	.016
Layer 3	agency	musical_background	< .001
		trial_count	< .001
	shared_agency	partner_realism	< .001
		trial_count	< .001
		relation_length	.042
		percussion	.014
	SOI	partner_realism	< .001
		trial_count	< .001
		training	.057
	flow	musical_background	.004
		SOI	.014
		relation_length	.014

Table 4-3: Significant effects in each methodological layer

4.5 Discussion

This experimental study aimed to investigate to what extent the coregulatory joint-action dynamics between two interacting, musical players are impacted when performing in an extended reality environment. It did this by operationalizing a methodological framework introduced in earlier research (Van Kerrebroeck et al., 2021) to analyse coregulatory dynamics across a performative, embodied, and subjective layer while varying the players' visual coupling and auditory context. The next paragraphs will first discuss the results in each analysis layer, followed by this study's broader implications and directions for future research.

Performance output: Overall, players performing the binary rhythm had a larger prediction error and thus performed worse than the ternary rhythm. Given the fact that the binary rhythm's tempo was further away from a preferred "resonance" tempo (van Noorden & Moelants, 1999) it is likely that this made it more difficult to correctly anticipate and adapt to upcoming pulses (Konvalinka et al., 2010; Repp & Su, 2013). Both players had comparable errors across levels of partner realism but were more accurate when playing with the polyrhythmic backing track and improved their playing over time. Learning effects had significant linear and 5th order terms with the latter illustrating the fact that trials were bundled (two musical backgrounds for each level of partner realism) due to the randomization constraints. Participants thus learned to perform the task better across as well as within levels of partner realism.

Embodied coregulation: Participants moved more energetically with the polyrhythmic backing track and when seeing the other as avatar. Movement analysis did reveal complex relationships across tasks, partner realism and musical backgrounds as illustrated by their interaction effects. Specifically, when playing with the metronome background, there was more movement energy when seeing as avatar for the ternary task, while less so for the binary task. When seeing the other as real, this effect was reversed. Given that the metronome background was more difficult to perform, the player might have compensated for this increased challenge using intensified bodily movements, yet only in the seeing as real condition. Visual coupling might have served as a compensatory or communicative mechanism to deal with the lower information density in the metronome background when partner realism was sufficiently rich in detail. Analysis of coherence values indicated that participants moved less in synchrony when playing with the polyrhythmic backing track. Bodily coordination or coherence as synchrony between players' movements only showed convincing levels at the ternary rhythm. The absence of a clear coherence at the common pulse might reflect the difficulty of synchronizing with the slowest pulse (van Noorden & Moelants, 1999) while the lack of coherence at the binary rhythm might be because it was more difficult to perform (Møller et al., 2021).

Coherence at the ternary rhythm decreased over trials and was lower when playing with a polyrhythmic backing track. While the hypothesis was that players would coregulate their movements more when playing in a richer auditory context and when getting more familiar with each other and the task, this decreased coherence and the improvement of prediction errors over time suggest that each player increasingly embodied their binary or ternary rhythm individually (Burger et al., 2014; Toiviainen et al., 2010). While players had a shared goal (create the polyrhythmic pattern) with a shared pulse, this goal might have been too complex to achieve, motivating players to focus on their individual parts first. Further research could investigate whether this decrease in coherence also holds for expert musicians as musical training had a positive effect on coherence levels.

Subjective experience: Visual coupling of players led to higher feelings of shared agency and SOI. These effects held for both seeing as avatar as for seeing as real with a more outspoken effect for the latter. It appeared that seeing the other modulated player's perceived sense of shared control, possibly because partner's actions became more predictable (Bolt & Loehr, 2017). Agency and flow increased when playing with a polyrhythmic backing track. Music, as an information and expressively rich mediator, was found to be capable of modulating player's perceived sense of self control, possibly through the increased movement related to the richer musical background (Dell'Anna et al., 2020). An interesting role was played by musical expertise in which more musical training and training in percussion led to a higher sense of SOI and shared agency respectively. These findings align with the prosocial effects of music found in other research (Stupacher et al., 2017) and might suggest the lesser role played by visual coupling for musicians to induce shared agency and SOI as compared to novices. As active investigations into the different forms of agency are underrepresented in virtual settings (Loehr, 2022), this study demonstrated an effective way to disentangle their relations and provided first steps towards leveraging the potential of extended reality for research into agency.

Circling back to the hypotheses in this study, the results indicate how a richer musical background led to the expected increase in performative precision between players' individual and joint performance. Given that the polyrhythmic backing track included beats at all levels (tactus, binary and ternary), the increased information density offered a rhythmic template on which to minimize asynchronies (Repp, 2012) and might have improved motor timing and stability through an entrainment effect (Carrer et al., 2022; Rose et al., 2021) and stronger groove (Leow et al., 2014). Regarding the second hypothesis, it was shown how increased partner realism and a richer musical background led to an increase in positive prosocial effects. These findings correspond to other research highlighting the social power of music (Stupacher et al., 2017; Tarr et al., 2014) and of synchronising with virtual humans (Hale & Hamilton, 2016; Tarr et al., 2018). It thus seems that the performative and experiential aspects of the dyadic interactions did not deteriorate in extended reality. Regarding the final hypothesis, different levels of partner realism did not have a significant impact on movement coherence between players while players did move more energetically when seeing each other as avatar and playing with a richer musical background. Novelty of the extended reality medium might have led to exaggerated movements to compensate expressive limitations (Chang et al., 2017) as well as enhanced social contagion (Dotov et al., 2021).

While players were allowed to improvise to induce more play in the musical interactions, this aspect was not analysed in this study. Future work could make improvisation and the role of visual coupling a research question on itself (Eerola et al., 2018) or include other aspects such as expressivity and underlying (shared) intentions (Goupil et al., 2021) in the layered analysis framework. While the polyrhythmic task was ideally suited for improvisation and playfulness, a synchronisation/continuation task such as in (Schultz & Palmer, 2019) might be an interesting alternative as it would more directly relate movement synchrony or coherence with task performance. While we opted to perform a cross-sectional study to attract a sample population with diverse backgrounds and characteristics, future work could opt to perform a longitudinal study (with novices or experts) to avoid large variability between dyads and focus on the observed learning effects. Finally, while the findings here result from a dyadic interaction between two virtual humans or avatars, future work could introduce additional human or computer-controlled players to investigate group dynamics (Dotov et al., 2022).

The analysis presented here has shown how virtual humans can be used to test specific hypotheses in a controlled experimental paradigm. Given the flexible nature of extended reality stimuli, building an extended reality platform such as the one in this study offers a solid basis on which to continue further investigations. For example, one could relatively easily replace the human-controlled avatars by computer-controlled agents (Tarr et al., 2018; Van Kerrebroeck et al., 2021) to test and refine dynamic models of control. By moving the research closer to where the action is (Engel et al., 2016), a paradigm such as the one presented in this study promises to achieve a more detailed view on the dynamics underlying social interaction and cognition.

4.6 Conclusion

Extended reality technologies have tremendous benefits to offer for scientific research as they allow us to test specific hypotheses in controlled, experimental scenarios in ever more realistic models of the world. However, a crucial need for music, and by extension joint-action or action-oriented, research is to assure that coregulation in human coordination does not break down in an extended reality context. This study evaluated this potential of experimental settings in extended reality by investigating to what extent visual coupling and an auditory context between musical players engaged in a joint-action polyrhythmic task influences performative, embodied coregulatory dynamics, and induces positive prosocial effects. While an informationally richer auditory context proved to be most beneficial for lower-level sensorimotor performative dynamics and favoured individual over joint performance, visual coupling in the form of an avatar or as in real life positively impacted experiential qualities of the interaction and induced prosocial effects. Embodied coregulation through bodily coordination was found to play an important but complex role illustrated by interaction effects of task complexity, visual coupling, and auditory context on movement energy. With the analysis presented in this study and its relevance for a broad range of research themes ranging from sensorimotor synchronisation, embodied coordination to experiential aspects in musical interactions, this study has demonstrated the value of state-of-the-art extended reality technologies in action-oriented research. A key point has been to show how the introduction of networked, (human-) controlled virtual humans capable of a realistic level of real-time expressive (bodily) interaction offer new experimental paradigms with more ecological validity while retaining high levels of experimental control.

5.

An auditory biofeedback system

Published as: Van Kerrebroeck, B., & Maes, P. J. (2021). A Breathing Sonification System to Reduce Stress During the COVID-19 Pandemic. Frontiers in Psychology, 12, 623110.

5.1 Introduction

The COVID-19 pandemic has impacted mental health globally (Torales et al., 2020). Studies of previous pandemics show that during times of a pandemic, people exhibit fear and anxiety-related distress responses (Bakshi et al., 2021). Although the consequences of the recent pandemic are yet to be understood, data has indicated widespread emotional distress leading to the definition of the COVID stress syndrome and stress scales (S. Taylor et al., 2020). While one area of research deals with the understanding of such consequences, another lies in the analysis of proper means to mitigate them. Research has already stressed the importance of interventions that can be delivered under pandemic conditions to reduce mental health issues and boost wellbeing (Holmes et al., 2020).

One avenue towards improved health and well-being is through music. Music can have physiological effects independent from individual preferences (Bernardi et al., 2009) and has shown benefits in therapeutic interventions. Specifically, music has shown its potential to reduce stress and anxiety in patients under hospital care through its facility to modulate arousal levels, regulate moods and by distracting patients from the experience of pain (Hunter & Schellenberg, 2010; Nilsson, 2011). Such music therapy has shown significant benefits as a support intervention to reduce stress and improve wellbeing in clinical staff working with COVID-19 patients (Giordano et al., 2020).

Another way of reducing stress and anxiety is through the practice of breath regulation (Clark & Hirschman, 1990; V. A. Harris et al., 1976). Breath pattern training can have a positive impact on patients with chronic obstructive pulmonary disease (Estève et al., 1996), a condition that increases the risk of developing a severe COVID-19 disease (Zhao et al., 2020). A goal of breath regulation is the modulation of (para)sympathetic activity, the so-called rest-or-digest and fight-or-flight response, resulting in increased relaxation or awareness and energy (Van Dixhoorn, 1998). Breathing has a strong influence on heart rate variability and has been described as an interface for voluntary control of the autonomic nervous system (Jerath et al., 2006; Kox et al., 2014). This coupling between breathing and hearth rate is the core principle behind the technique of resonance breathing in which people regularly exercise the slowing of their respiratory rate to a resonance frequency of around 6 beats per minute, causing high oscillations in heart rate (Lehrer et al., 2000). Other research has shown the benefits of modulating the retention time (Jafari et al., 2016; Kox et al., 2014) or the inhale/exhale duration to a ratio of around 1:2 (Adhana et al., 2013; Van Diest et al., 2014).

In practice, breath regulation comes in two forms: through conscious control or by stimulus entrainment. When breath is guided by an external stimulus, it is often through some form of biofeedback (Lehrer et al., 2000). These biofeedback systems differ by the type of stimulus (Bergstrom et al., 2014), whether users are instructed or not (Moraveji et al., 2011), whether some behavioural or physiological measures serve as input to the system (Herath et al., 2018), the type of sensors (Ayoola et al., 2018), the data processing and representations (Feijs et al., 2010) or whether the biofeedback is individually presented or in group (S. Moran et al., 2016). While a lot of research has been done, many, if not all, of these systems lack a close and fluent adaptation to the user. While they often incorporate some form of tempo-alignment, they fail to take the breathing phase into account. Another shortcoming with instructed breathing is the fact that users have reported hyperventilation, which can influence their state of relaxation (Van Diest et al., 2014).

An effective way of providing biofeedback is through the use of sonification. Sonification is defined as “the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation” (Kramer et al., 2010). Although sonification emphasizes the conversion of data relations into sound, music can also be used as a tool for sonification (Bergstrom et al., 2014). Importantly, biofeedback can then leverage the reinforcing and rewarding aspects of music (Maes et al., 2016). Other studies have shown how musical stimuli can help to control breathing patterns (Fried, 1990; J. Harris et al., 2014; Reza Namazi, 2017). Leveraging the rewarding aspects of music and the informationally transparent features of sonification can thus help to both motivate, monitor and modify physiological and physical processes (Maes et al., 2016). However, this requires an analysis of causes and effects of biofeedback systems in order to formulate appropriate alignment strategies leading to the proper modification of behaviour and physiology (Moens & Leman, 2015; van Dyck et al., 2017). This paper could support such an analysis by proposing a distance-based experimental protocol and biofeedback system that uses sonification to modify breathing patterns and induce states of relaxation. The aim of this study is the validation of this protocol and biofeedback system and evaluation of three auditory stimuli in modulating respiration rate, ratio, relaxation, and perceived pleasantness.

5.2 Method

5.2.1 Participants

Eight females and 11 males were recruited using an online banner in the public Facebook group of the Institute of Psychoacoustics and Electronic Music (IPEM). One male was excluded due to a misunderstanding of the experimental instruction. There were no participants that reported any respiratory disorders, 10 participants had musical training (Mean=11.7 years, SD=10.3) and 15 participants performed physical activity of which 7 did this more than three times a week. No exclusion criteria were applied, participants were not compensated for their time and average age of participants was 34 years (SD=8.61).

5.2.2 Procedure

The study protocol was reviewed and approved by the ethical commission of the University of Ghent. The experiment involved a within-subjects, repeated-measures design including three conditions of a breathing exercise in which different audio stimuli were tested.

Participants performed audio-guided breathing exercises of 8 minutes while indicating their breath cycles using arrow keys on their computer keyboards. The goal of each breathing exercise was to decrease the respiration rate to an optimal resonance frequency of approximately 6 beats per minute (Lehrer et al., 2000) and to increase the respiration ratio towards a healthier 1:2 inhale/exhale ratio (Adhana et al., 2013; Van Diest et al., 2014).

Each participant performed three trials of every condition and a baseline divided over six days which they were free to spread over a maximum of two weeks. Trial order was randomized between participants and each participant performed three trials on day 1, one trial on day 2 to 4, three trials on day 5 and the baseline trial on day 6. The baseline condition had no sound and was performed at the end of the experiment to incorporate learning effects and assure participants were used to breathing while indicating breathing onsets.

Experimental trials were conducted individually by participants at home for which they downloaded an application beforehand. At the start of the experiment, they met with the experimenter through a video call to verify sound playback, assure the use of headphones and receive instructions. During this first session, the experimenter stressed the importance of performing conditions in a comfortable position at around the same time in uniform contexts (same lighting, room, chair). The participant’s task during each trial was to breath naturally and comfortably and indicate breath inhalation and exhalation using the left and right arrow keys on their keyboard. All participants were asked to set the volume to comfortable levels while excluding background sounds. They sat in front of their computer while looking at the application that showed an indication when they pressed the arrow keys on their keyboard.

Each trial lasted 8 minutes, consisting consecutively of 1 minute of silence, followed by 7 minutes of an audio-stimulus. The stimulus was tempo-aligned to the participant’s breathing cycles for 1 minute, phase- and tempo-aligned for 3 minutes and then phase-delayed for 3 minutes (see section 5.2.3 and Figure 5-1 for more details).

Participants were asked to fill in a questionnaire before and after each trial that evaluated their relaxation state, hyperventilation, and pleasantness of the stimulus. The experiment concluded with a questionnaire probing for COVID-19 related stress levels and appreciation of the breathing exercises.

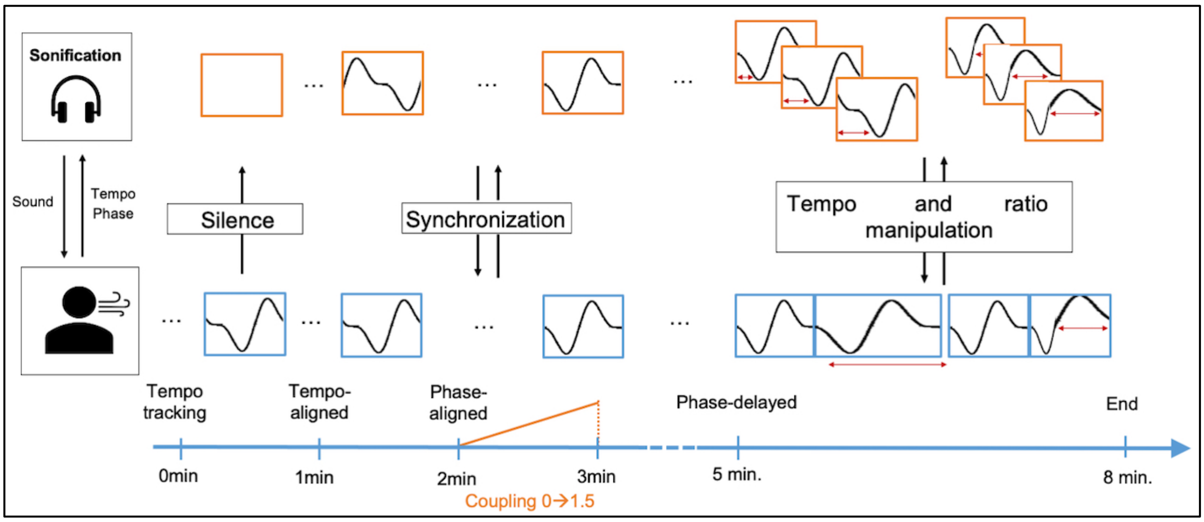


Figure 5-1: Biofeedback alignment strategy of breathing tempo and breathing phase

5.2.3 Stimuli

Each condition had a different stimulus that was modified using the same alignment strategy. The first condition, "noise" (No), consisted of amplitude and low-pass filter frequency modulated pink noise mimicking a human breath sound. The breathing sound was designed for a breathing cycle with an inhale/exhale ratio of 1:2 using smooth ramps toward the in- or exhalation onset. The second condition, "nature" (Na), aimed at providing a relaxing, ambient, and acoustically immersive environment. It emulated a natural environment with wind-blown leaves, windchimes, soft rain and bird song sound samples and binaural playback of which the wind-blown leaves sound sample was spatially contracted and expanded based on the breath cycle phase. Specifically, using an Ambisonics plugin (see section 5.2.4), eight duplications of the wind-blown leaves sound sample were placed at equidistant positions on a circle whose radius fluctuated with the breathing phase and was placed at anterior position of the listener. These samples were not time-aligned, were independently frequency filtered and amplitude modulated to prevent phase-cancellations and assure sounds differed enough to allow the spatial effect. The remaining sound samples were set at fixed, asymmetrical spatial positions around the listener. The third condition, "music" (M), played the stimulus from the noise condition on top of a decreasing arpeggio and 2-5-1 chord progression of a marimba-like sound, timed and played for an exhalation duration twice the length of inhalation. The noise sound was included to assure a continuous sound throughout the breathing cycle and because it created a sea-like sound in an ambient musical piece. When participants would indicate the start of an inhalation before the end of the phrase, the arpeggio and progression would quickly fade out. The rationale behind this design was that people would be spontaneously stimulated to wait for the harmonic resolution at the end of the phrase, and accordingly, to perform an optimal inhale/exhale ratio.

Tempo and phase-alignment between sonification and breathing as indicated by keypress onsets was done using an implementation of the Kuramoto model (Acebrón et al., 2005). The Kuramoto model is a mathematical model that models synchronization for coupled oscillators. It takes frequency, phase, and a coupling constant of oscillators as inputs, and outputs a frequency for each oscillator that moves them towards synchronization. In our application, tempo was calculated by taking the median of the last five inter-onset-intervals of the participant's keypresses. Phase was set to 0 and 240 degrees at keypresses indicating respectively the in- and exhalation. This ratio was chosen to ensure a fluent adaptation when the participant's respiration ratio approached the ideal 1:2 ratio. Between keypresses, phase was linearly interpolated based on the current breathing tempo. Phase-delay in the last 3 minutes of each session was varied between 0 and 50 degrees, based on the amount of synchronization between participant's breathing and the sonification stimulus measured using the mean vector length between the two phase vectors. We used this model as it has proven to be an effective means to adapt music to human behaviour (Moens et al., 2010) and because the model allows a fluent adaptation of the stimuli's tempo to the respiration rate. An overview of the different phases in the alignment strategy is given in Figure 5-1.

5.2.4 Materials

The biofeedback system was developed using the Max MSP framework[7],compiled for the MacOS and Windows operating systems and distributed using a Github respository[8]. Samples for the "nature" conditions were downloaded from the Freesound database (see Github for more details). Condition "noise" and condition "music" used stereo recordings. Ambisonics for binaural playback was used for condition "nature" and was created using the ICST Ambisonic spatialization tools with a standard head-related transfer function and without head-tracking (Schacher & Kocher, 2006). Time stamps of arrow key presses on participants' keyboards were recorded locally in .txt files, which were sent to the experimenter after conclusion of the experiment. Behavioural data consisted out of the recorded timestamps of these keyboard presses during the trials and the baseline.

Psychometric data consisted of measures of relaxation captured using 12 questions from the Smith Relaxation State Inventory Quick Test (SRSIqt) (J. C. Smith, 2001). We report

---

[7] https://cycling74.com/

[8] The application can be downloaded here: https://github.com/ArtScienceLab/SonicBreathing

four affective states from the questionnaire (basic relaxation, quiet focus, transcendence, positive emotion) and two stress dimensions (somatic and cognitive stress). Participants filled in the Nijmegen hyperventilation questionnaire (Van Doorn et al., 1982) to confirm natural breathing. A post-trial questionnaire asked participants how pleasant the sounds were (1=not at all, 3-4=moderately, 6=a lot). A follow-up questionnaire asked for COVID-19 related stress effects, whether they would like to continue doing breathing exercises, when they would do them, and whether they would prefer them with or without music. Stress effects were measured using a rating scale with four questions developed for this study asking about participants' feeling of stress and social isolation in general and during the pandemic (two factors with Cronbach alphas > 0.8 and ratings between 1=not at all, 3-4=average, 6=a lot).

5.2.5 Analysis

Two-way ANOVAs, with position (pre/post) and condition (No/Na/M/B) as within-subject factors are reported below, together with a Kruskal-Wallis test on the follow-up questionnaire. Bonferonni correction were applied for multiple follow-up comparisons of significant interactions. Êta-squared effect sizes are reported below as well.

Respiration rate and ratio timeseries were calculated by taking the median Inter-Onset-Time (IOT) interval in a 6-keypress sliding window and timeseries were interpolated and resampled at 0.1Hz (48 samples for each 8-minute trial). IOTs and ratios that exceeded 5 times the standard deviation of the previous window were treated as erroneous presses and discarded ([Mean, SD] of the percentage of trial skipped: rates=[1.13, 1.92], ratios=[1.04, 1.69]).

As we were interested in evaluating the effects of the stimuli in decreasing the respiration rate and increasing the respiration ratio over time, we modelled timeseries with linear-mixed effects models using the lme4 package (Bates et al., 2015). Timeseries were log-transformed and modelled using a fourth-order orthogonal polynomial with a fixed effect of the condition on linear and all time terms for rate and ratio respectively. Random effects consisted of participants and participants-by-condition-by-day on all time terms. Statistical significance for individual parameters was estimated using the normal approximation.

5.3 Results

One participant did not perform day 6 of the experimental procedure. Two participants performed the wrong condition once, two other participants missed one trial on day 5. In total, 18 participants performed 175 trials.

Regarding the psychometric data, 6 participants filled in fewer questionnaires than prescribed by the protocol. One participant did not fill in the follow-up questionnaire. Two participants answered the pre- and post-trial questionnaires on day 5 only once, two other participants made the same mistake but on day 1, the remaining participants had 5 missing pre- or post-trial questionnaires on days 2, 3, or 4. In total, questionnaire data was received for 162 of the 175 trials

5.3.1 Psychometric data

Hyperventilation measures from the Nijmegen questionnaire assured a natural behaviour of participants during each trial (mean=5.52, SD=4.28). Concerning stress and relaxation, we compared the six considered subcomponents of the SRSIqt (J. C. Smith, 2001) before and after each trial for all days (see Figure 5-2). Pre- and post-trial scores showed a significant effect for the relaxation component "transcendence" (F(1,301)=5.756, *p*=.017, η^2=.018, 15 outliers removed) with a significant interaction effect between condition and pre/post trial scores (F(1,301)=4.703, *p*=.003, η^2=.044). Post-hoc tests showed significant increases for "transcendence" in post-trial questionnaires for the noise and nature conditions and decreased scores for the music condition ([pre-post, No-M]: *p*<.001, [pre-post, Na–M]: *p*=.010). There were significant effects for pre/post trial scores for the components "somatic stress" (F(1,306)=9.428, *p*=.002, η^2=.03, 10 outliers removed) and "cognitive stress" (F(1,298)=7.091, *p*=.008, η^2=.022, 18 outliers removed) with decreased scores across conditions. Finally, there was a significant effect across conditions on the "cognitive stress" component (F(3,298)=4.307, *p*=.006, η^2=.040) with post-hoc tests indicating a significant decrease for the noise condition (*p*< .003).

Stimuli from the different conditions received the following average scores on perceived pleasantness ([Mean, SD]: B=[2.24, 1.30]; M=[3.43, 1.17]; No=[4.16, 1.36]; Na=[4.88, 0.89]). A Kruskal-Wallis test showed significant differences between scores (χ^2(3, 162)=49.307, *p*<.001). Post-hoc pairwise Wilcoxon Rank Sum tests indicated significant differences between all stimuli (*p*(No-Na)=.014, *p*(No-M)=.005, *p*(No-B)<.001, *p*(Na-M)<.001, *p*(Na-B)<.001, *p*(M-B)=.004).

Finally, in the follow-up questionnaire, participants did not indicate significantly higher stress levels during the COVID-19 pandemic as before (*p*=.29) yet did show a significant increase in terms of feeling of social isolation (*p*=.029). 13 out of 17 participants indicated they would continue doing the breathing exercises after the experiment, of which 10 answered to prefer them with as compared to without music. A majority would do them at stressful or lonely moments and despite the COVID-19 pandemic, two participants indicated they would do them because of the pandemic.

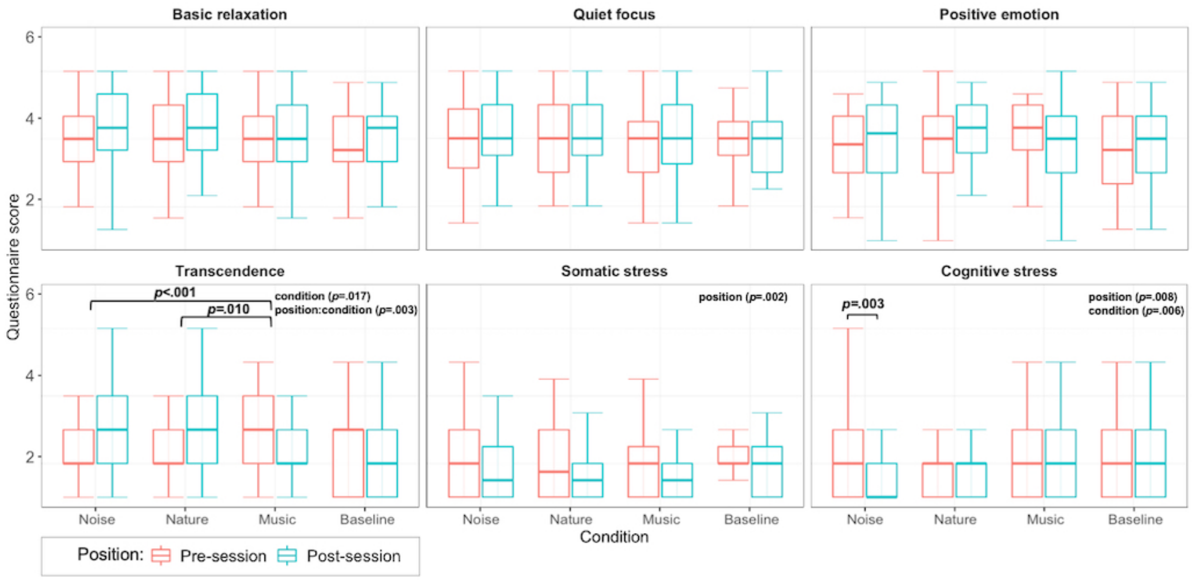


Figure 5-2: Boxplot of relaxation questionnaire data (subcategories taken from (J. C. Smith, 2001)*)*

5.3.2 Behavioural data

Respiration rate timeseries were analysed for all participants for 173 trials. Two trials with sudden respiration rate peaks of 18.64 and 26.70 beats per minute were removed because of outliers in the timeseries and problems with model fitting (residual outliers).

Mean respiration rate curves and model fits are shown in Figure 5-3. There were small significant effects of the noise and nature conditions on the intercepts ([Estimate, SE, *p*]: No=[-0.193, 0.099, 0.052]; Na=[-0.205, 0.099, 0.039]), indicating overall higher odds on a lower respiration rate at the start of a trial. Significant effects were found for these conditions on the linear slope time term indicating an overall significant decrease in respiration rate for the trial with as compared to without sounds ([Estimate, SE, *p*]: No=[-0.341, 0.155, 0.029]; Na=[-0.460, 0.155, 0.003]) .

Respiration ratio timeseries were analysed for 169 trials. Six trials with large respiration ratio fluctuations were removed because of outliers in the timeseries and problems with model fitting (residual outliers).

Mean respiration ratio curves and model fits are shown in Figure 5-4. Large standard error bands are explained by the individual differences captured by the random effects in the model and one participant with a stable high ratio for the baseline condition ([Mean, SD]=[2.24, 0.17]). There were significant effects of condition on most time terms for the noise condition ([Estimate, SE, *p*]: linear=[0.363, 0.102, 0.0004]; quadratic=[-0.222, 0.074, 0.003]; quartic=[0.096, 0.045, 0.035]). As illustrated in Figure 5-4, the effects indicate an increasing ratio over time with stable plateaus at the beginning and end of the trial for the noise condition.

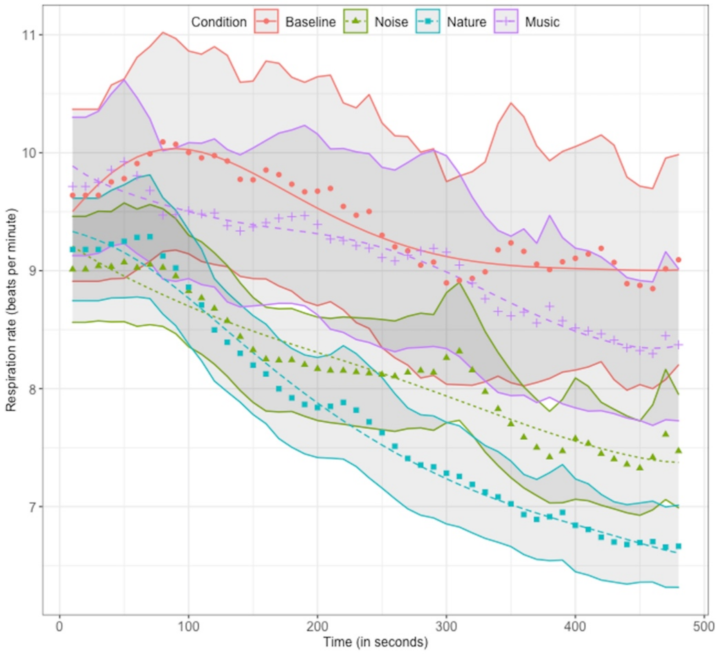


Figure 5-3: Mean, standard error, and model fitted curves for the respiration rate timeseries for all conditions

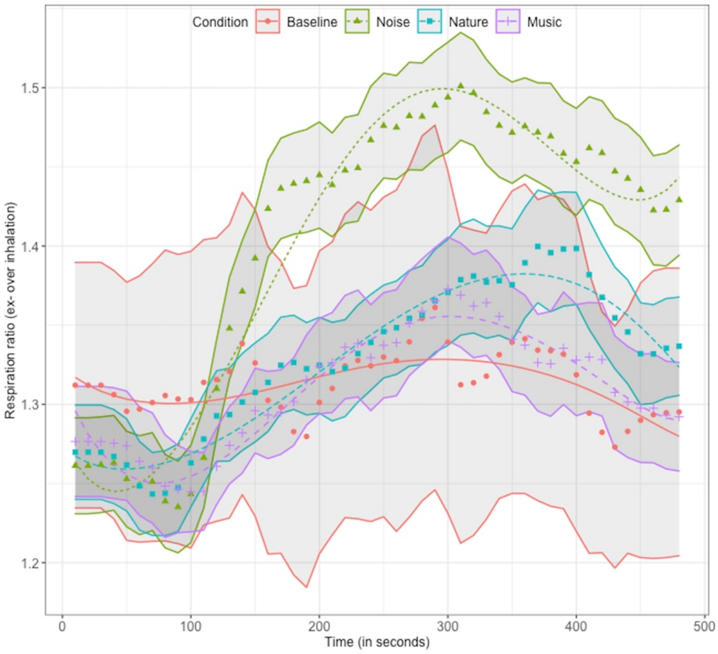


Figure 5-4: Mean, standard error, and model fitted curves for the respiration ratio timeseries for all conditions

5.4 Discussion

The aim of the present study was to assess the effects of a newly developed breathing sonification system on stress levels in persons during the COVID-19 pandemic. Our starting point thereby was the idea that stress can be reduced, and relaxation increased by modulating breathing patterns demonstrated by earlier research. We proposed the use of music and sound to effectuate spontaneous breathing pattern adaptation towards these goal states rooted in their combined arousal (Bergstrom et al., 2014; Hunter & Schellenberg, 2010; Nilsson, 2011) and (spontaneous) synchronization effects (Maes et al., 2019). The present study applied these principles of music and sound into a sonification system for spontaneous breathing adaptation. We specifically tested the effect of three types of sonification: (1) a noise sonification simulating a natural breathing sound; (2) a nature environmental sonification with ambient and spatially expanding/contracting sounds; (3) a musical sonification with tension and resolution in chord progressions and arpeggios. We assessed effects on breathing behaviour as well as on perceived pleasantness, relaxation, and stress levels.

Concerning the results on breathing behaviour, it was found that the nature sonification led to the most prominent decrease in respiration rate. Interestingly however, it was the noise sonification that had the most beneficial effect on the respiration ratio. This differentiated effect on breathing behaviour, depending on the sonification type, is likely to be explained by the specific auditory quality of the sonification. Potentially, the nature sonification with an ambient atmosphere may prompt arousal mechanisms, while the noise sonification with human breath-like sounds may rather spur synchronization mechanisms. This explanation is supported by findings from earlier research that has indicated the positive physiological effects of being immersed in natural soundscapes following stress (Aletta et al., 2018; Medvedev et al., 2015) and the existing link between perceived pleasantness and arousal from listening to soundscapes (Aletta et al., 2018; Hume & Ahtamad, 2013). On the other hand, sensorimotor synchronization mechanisms might have facilitated the changing respiration ratio resulting from the noise sonification. Since this sonification mimicked human breathing, in addition to rhythmical entrainment (Bardy et al., 2015), participants might have "mirrored" these sounds as in the use of ecological breathing sounds (Murgia et al., 2016). Though the underlying mechanisms may not be clarified, the finding urges us to consider a combination of both sonification types in future research to maximize breathing adaptation effects.

Next to the physical breathing measures, we collected psychometric data to assess subjectively perceived pleasantness, relaxation, and stress levels. Participants reported, on average, higher feelings of social isolation during the COVID-19 pandemic, but no significant higher stress level. However, across the different sonification types, the intervention led to significant reductions in stress and increased relaxation. Looking at the specific effects of the sonification type, we found that both the noise and nature sonification led to significantly higher increase of the experienced "transcendence" dimension compared to the musical sonification. We found "cognitive stress" to be significantly lower using the noise sonification and the nature sonification as the most pleasant. Finally, a large portion of participants (13 out of 17) reported that they would be willing to continue using our breathing sonification system in the future. From that portion, 10 participants pointed out that they preferred using it with music demonstrating the added value of the intervention.

The main outcome of the present study is that it offers an effective distance-based protocol and biofeedback system to improve interventions for breathing adaptation and stress regulation. Results indicate that a combination of the noise and nature sonification is the proper candidate for future research and application. Both types of sonification contribute positively to distinct aspects of experience and adaptation in physical breathing behaviour. Further research with larger sample sizes from heterogeneous populations with different backgrounds in for example music or meditation, studied at home and in the lab could confirm the expectation that their combination will lead to optimal results as well as reveal discriminatory effects resulting from individual profiles.

To generalize these preliminary findings, further research should compare results with those from post-pandemic experiments performed both at participant's home and in the lab as well as investigate the influence of exposure conditions. While performing the experiments at participants' home increased the validity of the results, it also increased variability. Further research could control for this variability by an in-depth characterization of exposure conditions such as ambient sounds, headphone type, lighting, and psychosomatic factors during the introductory session with the participant. Another complementary way would be to incorporate automated controls such as for example headphone screening tests (Milne et al., 2021; Woods et al., 2017), a gold standard (Nelson & Allen, 2019) and background level recordings (Murphy & King, 2016). In future research, we plan to use a portable, custom-made sensor system that we developed to automatically track breathing cycles which allows to free the cognitive resources required to manually indicate breathing onsets. These studies will explicitly target physiological measures that indicate stress levels, such as heart rate variability and skin conductance, providing a better view on effects realized by our intervention application. In addition, while we induced a spatial effect in this study on manually positioned sound sources using Ambisonics and a standard head-related transfer function, a possible improvement could be to incorporate head-tracking and individualized head-related transfer functions as it has shown to enhance binaural playback (Hong et al., 2019; Katz & Parseihian, 2012).

The COVID-19 pandemic poses substantial challenges to empirical research, in particular the data collection aspect, given the requirement of self-isolation and social distancing. One of the important contributions of our breathing intervention application is that it allows the collection of quantitative data in a reliable and accurate manner, while participants can stay at home. This allows not only the coping with COVID-19 regulations, but equally to collect data within the participants' familiar, real-life environment, making them feel more at ease as compared to a laboratory environment.

6.

A selection of experimental music research in extended reality

6.1 Introduction

This chapter presents four studies that used XR technologies to create musical spaces for experimentation. They demonstrate several strengths of using XR in experimental music research and highlight opportunities that could lead to an increased understanding of the social, cognitive, affective, and behavioural dynamics inherent to music making. Abstracts will briefly introduce the context and research questions of each study, followed by a discussion of the XR elements that render their investigation possible.

The first two studies are characterized by a high level of experimental control and focus on fundamental questions around interpersonal coordination. They leverage XR's potential to deliver impossible stimuli, induce body illusions, and devise novel experimental scenarios using virtual humans and perspective manipulations. By integrating a combination of neurophysiological and behavioural measurements in a single user and dyadic immersive setting, these studies demonstrate a comprehensive and structured approach on how to perform cognitive (neuro)psychology research into music making using XR.

The next two studies are characterized by a high ecological value, investigating music making in "live" educational and performance contexts, and can be considered a form of applied research. Both studies evaluate the impact of immersion, one by presenting a virtual violin teacher in a 2D and 3D display, and the other by livestreaming music performances using 2D or 360 video. Using questionnaires in both studies, as well as motion capture in the virtual violin teacher study, they offer concrete suggestions on how to improve educational applications or online concert experiences.

6.2 Tapswap

Title: Embodied perspective taking in dyadic entrainment

Authors: Mattia Rosso, Canan Gener, Bavo Van Kerrebroeck, Pieter-Jan Maes, Marc Leman

Abstract: Humans exhibit a compelling tendency to synchronize their movements with one another. As soon as two individuals exchange information via one or multiple sensory channels, such phenomenon may occur spontaneously and even against the intention to ignore the other. Ecologically, dyadic interactions take place in settings where partners perceive each other from a face-to-face 2nd person perspective. In the present work, we intended to explore the role of visual perspective in interpersonal coordination, under the hypothesis that perceiving the movements of a partner from their 1st person perspective would further strengthen their entrainment. Such perspective-taking can be induced in a bottom-up fashion by experimentally transposing the visual scenes perceived by two individuals into the partner's egocentric frame of reference.

Twenty pairs of participants were equipped with VR headsets and engaged in a joint finger-tapping task. We video recorded their hands with low-latency cameras from different angles and streamed the visual scene to different headsets across conditions. In a 2 x 2 factorial design, participants performed a finger-tapping task while looking at their partner's hand in 2nd person (Coupled, 2P), while looking at their own hand in 1st person (Uncoupled, 1P), while looking at the partner's hand in 1st person (Coupled, 1P), and while looking at their own hand in 2nd person (Uncoupled, 1P). The drifting metronomes paradigm for dyadic entrainment (Rosso et al., 2021) was adopted to quantify overall strength of dyadic entrainment and attractor dynamics, while subjective ratings of embodiment and agency over the visually presented hand were measured by means of questionnaires.

We expected that perceiving the other's hand in 1P would result in stronger dyadic entrainment, due to the tendency of the brain to correct for temporal mismatches (i.e., asynchronies with the partner's finger-taps) in an otherwise visuo-spatially congruent percept (i.e., the partner's hand matching the location of the own hand). We also expected this effect to be mediated by a higher subjective feeling of embodiment and agency over the partner's hand.

Although the overall strength of entrainment did not significantly differ across levels of perspective, we found a significant difference at the level of local dynamics. Whilst in 2P coupling participants managed to maintain more sustained periods of decoupling, in condition of 1P coupling partners were not capable of pursuing independent individual trajectories. We conclude that visual perspective influences coordination dynamics of dyadic interactions, and we address the potential of deploying the technology hereby presented in the context of motor rehabilitation to support interpersonal coordination between patient and therapist.

XR design and relevance: A prominent feature of XR is its capability to induce virtual embodiment and offer control over the perspectives of its users. By exchanging perspectives of people in real-time, one can induce body illusions, causing an illusory sense of ownership over prosthetic or virtual limbs (Slater, Perez-Marcos, et al., 2009), potentially leading to a sense of agency over their movements (Kalckert & Ehrsson, 2014). Here, these capabilities are leveraged in a controlled experimental study for fundamental research illustrating the potential of XR as a new methodological paradigm. The study manipulates 1st and 2nd person perspectives of participants engaged in a tapping task to investigate the impact on their sensorimotor synchronization. As research has shown the cognitive load of perspective switching and the benefits of learning from a 1st person perspective (Fiorella et al., 2017; A. K. Martin et al., 2019), facilitating this process with XR technology could free up cognitive resources, potentially leading to improved applications for learning, training, and rehabilitation of sensorimotor functions (De Oliveira et al., 2016).

The study succeeded in capturing behavioural and neural dynamics of sensorimotor synchronization through a modular XR design of hardware and software for data capture, mediation, and display. It integrated pipelines for visual stimuli using Unity, HTC Vive Pro HMDs, and low-latency Logitech Brio cameras, for audio stimuli using Ableton and custom-made drum pads, together with an Antneuro EEG brain monitoring setup (see Figure 6-1 and Figure 6-2). This modular design of the study, together with its experimentally controlled nature, lends itself well for further research into interpersonal coordination. For instance, it would be relatively easy to introduce controlled latencies in the video feeds or replace video feeds of "real" hands with human- or algorithmically-controlled virtual hands to improve coordination models (Tognoli et al., 2020). Another avenue could be to introduce additional participants, opening the door to investigations into group dynamics (Dotov et al., 2022; Gordon et al., 2020; Shahal et al., 2020).

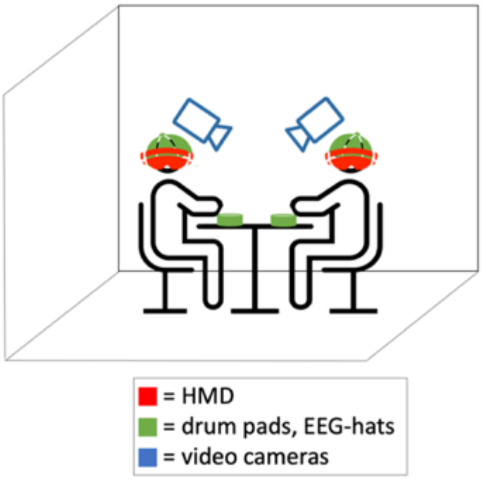


Figure 6-1: Schematic XR design

Figure 6-2: Snapshot taken during data collection showing two participants wearing HMDs and matching gloves, a separating screen, the video cameras, and EEG system

6.3 Virtual partners

Title: Expressive virtual agent during music performance induces critical levels in brain systems of arousal

Authors: Marc Vidal, Bavo Van Kerrebroeck, Ana M. Aguilera, Pieter-Jan Maes, Thomas Hans Fritz, Marc Leman

Abstract: Using EEG and pupillometry recordings while moving and singing with virtual humans in VR, we provide a novel perspective on the relationship between movement

and levels of emotionality. Arousal mechanisms of the nervous system are evident during music performance practices, not only to mitigate external stressors but to enhance the performance as well. In this paper, we hypothesize that a virtual avatar and agents with expressive motor abilities can engage and intensify the emotional experience of music. Participants with musical training were recruited to perform vocal tasks in an immersive VR experience. Participants saw a virtual avatar or agent in the experience, that moved along to the music by mimicking participant's movements or conducting with various degrees of expressiveness. To decode levels of emotional arousal, a model is currently being developed that relates power components of brain signals to pupillometry signatures, which have been previously associated with the activity of specific neurotransmitters.

XR design and relevance: Virtual humans enabled by XR technology promise to be a useful methodological tool for social cognitive psychology research by offering experimental control, replicability, and the elimination of confounds of suggestion (Pan & Hamilton, 2018). An argument to use them in experimental research comes from research showing how people respond realistically to virtual humans when experiencing high presence in virtual environments (de Borst & de Gelder, 2015; Sanchez-Vives & Slater, 2005). Consequently, they have been a useful tool to investigate the prosocial effects of social closeness resulting from synchrony (Tarr et al., 2018), and of increased liking and trust from mimicry (Hale & Hamilton, 2016; Verberne et al., 2013).

However, care must be taken in their use as experimental stimuli, given the concerns raised regarding their potential to elicit different behavioural and neuronal responses as compared to real humans (de Borst & de Gelder, 2015). An often-mentioned cause for this differential is the Uncanny Valley hypothesis, introduced earlier in section 1.4.2, stating how virtual humans can evoke a feeling of repulsion or eeriness when they behave or appear almost like real humans (Carter & Pollick, 2014; Mori et al., 2012). Appearance and behavioural realism of virtual humans are thus important aspects to take into consideration, given they can both impact emotional responses (Jun et al., 2018; Zibrek et al., 2019), and potentially mutually influence their effects (de Borst & de Gelder, 2015). This study addresses this issue by using virtual humans with a non-realistic, uniform appearance (see Figure 6-4), together with realistic movements with varying degrees of expressiveness. It investigates emotional responses in participants singing and moving along with a non-expressively and expressively moving virtual agent, as well as a human-controlled avatar mimicking participant's movements based on the virtual mirror paradigm (González-Franco et al., 2010).

Comparable to the Tapswap study in section 6.2, this study has a modular XR design (see Figure 6-3), allowing for flexibility and control in the delivery of stimuli. The main differences with the previously presented study are the single-user setup and the full immersion in a virtual environment populated with a virtual human. Neurophysiological data obtained from a study such as the one presented here, could strengthen behavioural findings of mimicry and synchrony with virtual humans (Hale & Hamilton, 2016; Jun et al., 2018; Tarr et al., 2018; Verberne et al., 2013), potentially enriched with additional manipulations, such as the integration of facial expressions (Jun et al., 2018) or behavioural models in the agent (Tognoli et al., 2020). Furthermore, manipulating appearance realism (Zibrek et al., 2019) and integrating a condition with real humans would help to establish baselines and disentangle the interactions of behaviour and appearance realism. By introducing virtual agents and a virtual mirror in a controlled neurophysiological experimental setup, this study thus provides a comprehensive approach to allow a better understanding of the emotional responses in (musically) moving along with others.

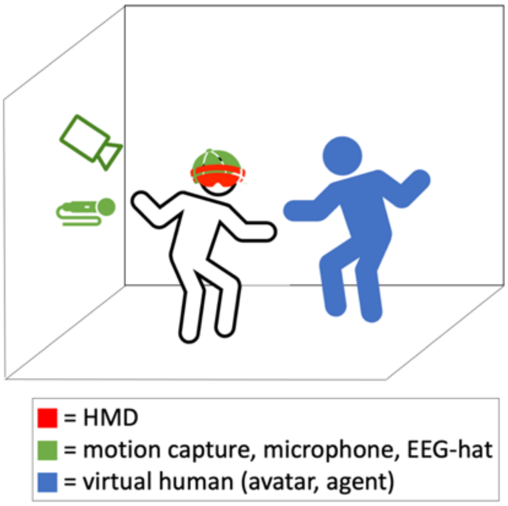


Figure 6-3: Schematic XR design

Figure 6-4: Data collection with a view of the participant's back (left) and a 1st person view of the virtual agent (right)

6.4 Virtual teacher

Title: The assessment of presence and performance in an AR environment for motor imitation learning: a case-study on violinists

Authors: Adriaan Campo, Aleksandra Michałko, Bavo Van Kerrebroeck, Boris Stajic, Maja Pokrić, Marc Leman

Abstract: The teaching of refined gestures presents an educational challenge in several domains of skilled sensorimotor performance. This study investigated whether AR learning for violin practice is more efficient in a 2D environment than in a 3D environment, with the latter rendered as a violin playing avatar visible through a HoloLens 2 HMD.

Eleven participants practiced two musical pieces with this avatar in four trials, spread evenly over a month. Each participant practiced in one of two conditions, either with a 2D rendering of the avatar or with a stereoscopic 3D rendering of the avatar. Like in real orchestral playing, participants were asked to mimic the avatar as closely as possible when it came to using the bow, i.e., mimic gestures related to bowings, articulations, and dynamics. Gestures of violin playing were recorded and analysed in terms of kinematic metrics, while questionnaires examined subjective experiences of presence. Based on hierarchical regression modelling, results revealed an overall better similarity in gestures and a higher experience of presence with the 3D avatar as compared to the 2D avatar, while a learning effect was observed in the 2D avatar condition but not in the 3D avatar condition. Findings suggest that performance in the 3D condition started from a better baseline, leaving less room for improvement than in the 2D condition. Overall, the effect of the 3D environment was greater on performance quality than on learning. This work is concluded with suggestions for future work on AR-based advanced gesture training based on participant's comments, together with suggestions to overcome identified challenges in measurement methodology.

XR design and relevance: This study illustrates how XR can be used to develop and evaluate musical applications based on a user-centred design. By having students use an AR music education application while evaluating their learning progress and collecting their feedback, this applied form of research demonstrates how to quickly prototype XR features and test specific hypotheses about 2D and 3D learning (McIntire et al., 2014), presence and learning (Ochs & Sonderegger, 2022), and presence and performance (Grassini et al., 2020). Further insights on the underlying mechanisms of learning in immersive education and the role of presence are valuable given the advantages of AR for learning (Wojciechowski & Cellary, 2013) and its operationalisation in musical contexts (Martin-Gutierrez et al., 2020; Trujano et al., 2018).

The application offered a 2D and 3D virtual teaching agent (see Figure 6-5 and Figure 6-6) together with virtual affordances in the form of a button and slider, allowing the participant to play and pause the audiovisual animation or move playback for- or backwards. Several methodological challenges to be addressed in future research are raised in the study, such as the use of questionnaires to measure presence, the novelty effect of XR applications, and the difficulty in evaluating multimodal education applications due to individual differences. Incorporation of individual differences such as personal backgrounds, communication preferences, and age have been suggested as a way to improve XR music education applications (Michałko et al., 2022), as well as to inspire informationally-rich and adaptive (bio)feedback applications (Maes et al., 2018). While the sample population in this study was small (11 participants), placing XR music making applications “in-the-wild” (Kothgassner & Felnhofer, 2020), or “there where the action is” (Engel et al., 2016), and evaluating their (learning) dynamics in a structured way, can lead to the introduction of an iterative, user-centred, and co-creative cycle between researcher, designer, student, and (virtual) teacher in, and for, the development of flexible, playful, and stimulating learning environments (Michałko et al., 2022).

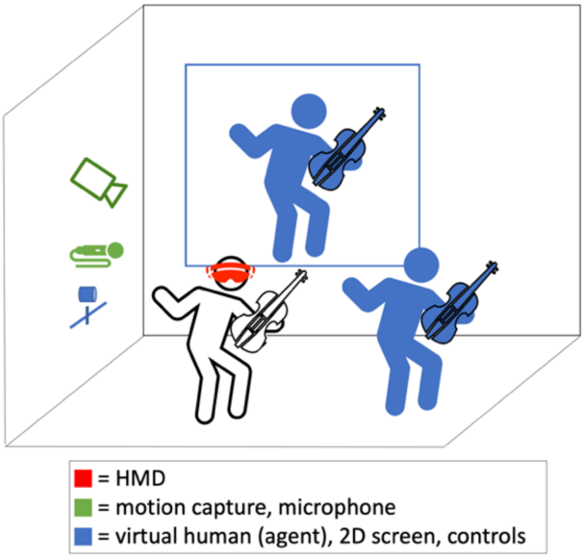


Figure 6-5: Schematic XR design

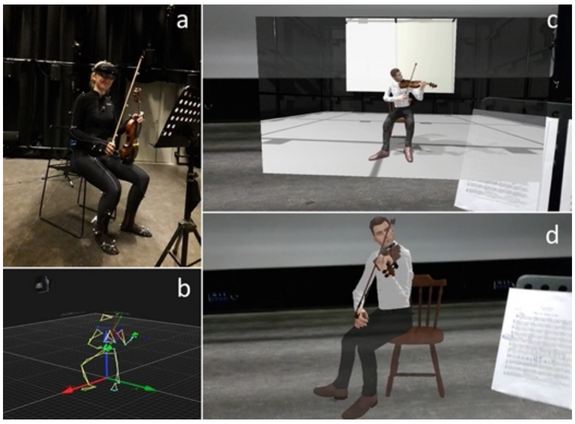


Figure 6-6: Virtual teacher (a. Motion capture, b. Animation skeleton, c. Agent as 2D stimulus, d. Agent as 3D stimulus, c. and d. show the participant’s 1st person view through the Hololens)

6.5 Livestream experiments

Title: The role of COVID-19, agency, presence, and social context in facilitating social connectedness

Authors: Kelsey E. Onderdijk, Dana Swarbrick, Bavo Van Kerrebroeck, Maximillian Mantei, Jonna K. Vuoskoski, Pieter-Jan Maes, Marc Leman

Abstract: Musical life became disrupted in 2020 due to the COVID-19 pandemic. Many musicians and venues turned to online alternatives, such as livestreaming. In this study, three livestreamed concerts were organized to examine separate, yet interconnected concepts—agency, presence, and social context—to ascertain which components of livestreamed concerts facilitate social connectedness. Hierarchical Bayesian modelling was conducted on 83 complete responses to examine the effects of the manipulations on feelings of social connectedness with the artist and the audience. Results showed that in concert 1, where half of the participants were allowed to vote for the final song to be played, this option did not result in the experience of more agency. Instead, if their preferred song was played (regardless of voting ability) participants experienced greater connectedness to the artist. In concert 2, participants who attended the concert with virtual reality headsets experienced greater feelings of physical presence, as well as greater feelings of connectedness with the artist, than those that viewed a normal YouTube livestream. In concert 3, attendance through Zoom led to greater experience of social presence, but predicted less connectedness with the artist, compared to a normal YouTube livestream. Crucially, a greater negative impact of COVID-19 (e.g., loneliness) predicted feelings of connectedness with the artist, possibly because participants fulfilled their social needs with this parasocial interaction. Examining data from all concerts suggested that physical presence was a predictor of connectedness with both the artist and the audience, while social presence only predicted connectedness with the audience. Correlational analyses revealed that reductions in loneliness and isolation were associated with feelings of shared agency, physical and social presence, and connectedness to the audience. Overall, the findings suggest that in order to reduce feelings of loneliness and increase connectedness, concert organizers and musicians could tune elements of their livestreams to facilitate feelings of physical and social presence.

XR design and relevance: This study, rooted in the domain of cognitive psychology, controlled aspects of agency, presence, and the social context of audiences in live concert settings to investigate the facilitation of social connectedness (see Figure 6-7 for a schematic overview and Figure 6-8 for screenshots of the three livestreamed concerts). Its XR design consisted of affordances offered to the audience such as the voting on a song and its control of the degree of immersion.

Immersion only involved live-streamed video content and did not involve the design or modelling of detailed virtual environments. This however raised its ecological value by investigating the live concert experience the way we were used to during the COVID-19 pandemic (Vandenberg et al., 2021). Users watched live concerts from a fixed vantage point and only chose their point of view in a 360 video on YouTube or a cardboard VR (see Figure 6-7). While the XR design might have been relatively simple, a fixed vantage point and control of the point of view do represent the classic concert experiences in “real” life and are thus important elements to evaluate and potentially safeguard, import, or carefully redesign in virtual experiences.

The study’s finding that physical and social presence facilitate connectedness and can reduce feelings of loneliness are important for the design of livestreamed concert experiences and would be interesting to further investigate when additional XR capabilities, such as spatiotemporal navigation, the choice of vantage point, and virtual embodiment are introduced. Another pathway for further research would be to modulate immersion in terms of audio and video resolutions. Given the plethora of mobile, 2D and

3D screens, and audiovisual playback today, more insights on the impact of audiovisual quality could have direct implications for the future organisation of livestreamed concerts. Findings that 3D sound can increase social presence (Shin et al., 2019), and that audio quality influences physical presence (Larsson et al., 2010; Nordahl & Nilsson, 2014), represent first indicators of the relevance and value of improved audio playback for virtual concert experiences. This study, close to the musical activities taking place in real-life, illustrates how XR technology and resulting insights can improve the understanding and design of live musical experiences, clearing the way for its beneficial social and experiential effects.

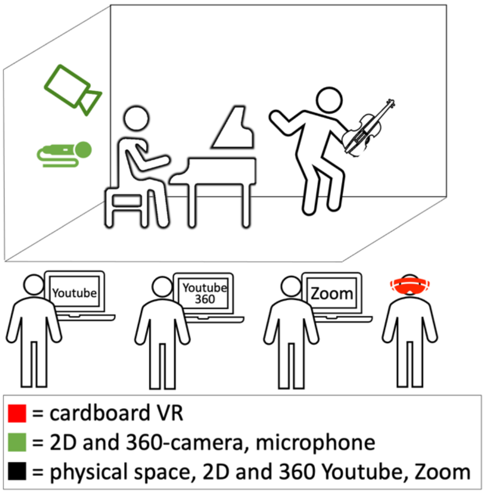


Figure 6-7: Schematic XR design

Figure 6-8: The three livestreamed concerts (a. 2D video stream, b. 360 video stream, c. 2D video stream + social interaction (not shown))

7.

A selection of artistic music research in extended reality

7.1 Introduction

This chapter presents six artistic studies grounded in the musical spaces for performance, composition, and narration. Their aim is to highlight the opportunities of XR for artistic expression and the creation of novel musical experiences. An abstract of each study will present artistic goals and their realisation in public demonstrations. They are followed by a discussion on the underlying XR design together with implications for the embedded musical roles and experiential aspects arising from presence, perspectives, body illusions, and affordances enabled in XR. Video material of the artistic studies can be found online[9].

7.2 Scientific case-studies as performance

Musical performances in XR offer new ways for performers to communicate expressive intentions and share them with hybrid, online and offline audiences. Virtual embodiment allows performers to interact expressively using fine-grained gestures as well as afford new ways to present and appear on stage. Networking these performers online allows easier access for the public as well as new ways of immersion through mobile, screen-based displays or HMDs. This section presents two musical XR performances taking place in and around ASIL with performers interacting as virtual humans in front of an audience. The Piano Phase study focuses on the interaction of a pianist with a dynamic, algorithmically-controlled virtual agent, presented to an audience with different levels of immersion and degrees of agency. The Music Moves study presents a networked, embodied, polyrhythmic interaction between two performers presented to a hybrid, online and offline audience.

7.2.1 Piano phase

Authors: Bavo Van Kerrebroeck, Giusy Caruso, Pieter-Jan Maes

Abstract: This performance explored the transfer of musical interactions from a scientific setting into public space. It built on the work done for the scientific case-study presented in chapter 3 and involved the piano duet performance Piano Phase composed by Steve Reich. The performance took place on the 5th of November 2019 in the ASIL of the University of Ghent and was performed several times to groups of around 6 to 10 people.

The piano duet was performed by a professional pianist together with the algorithmically-controlled agent introduced in chapter 3. The performer wore a motion capture suit that allowed the rendering of her movements as an avatar in a virtual environment shared

---

[9] https://youtube.com/playlist?list=PL_9jB_2tw40xQW2VQt-juZlkW4cBQIJPs (Accessed on the 20th of November 2022)

with the virtual agent. Avatar and agent were rendered as futuristic humanoids and played a point cloud rendering of a piano captured a few days before the performance. The point cloud moved along with the music, oscillating more energetically when piano playing of avatar and agent was desynchronized. Sounds were played back as simple stereo sources, positioned at the physical positions of both performers. The audience was standing in the same space as the performer, with one spectator wearing a wireless HMD navigating the virtual space (see Figure 7-1 and Figure 7-2). The 1st person view of this spectator was projected on a 2D screen in the physical performance space, to allow the larger audience to follow the virtual exploration of the active spectator. No HMD was given to the performer. While this excluded any embodied coordination between performers, it did allow closer connection of the performer with the larger audience.

In contrast to the scientific case-study in chapter 3, only one condition was performed here, namely those with the physically-absent, yet virtually-present, dynamic agent. The purpose of the performance was to draw attention of the audience on the subtle shifts in phase inherent to the piece, and highlight the process of continuous adaptation and (auditory) coregulation between performer and the virtual agent. As the agent was physically absent, yet auditorily present, the purpose was to provoke reflective questions in the audience as to what caused the music to change and how. The visual rendering of the agent in the virtual environment, the exploratory nature of the active spectator, and the oscillating piano point clouds responding to the music were all meant to stimulate this reflective stance, and to guide the audience's attention towards the hidden (musical) causes in the performance space.

XR design and relevance: This project had a relatively simple XR design that succeeded, with minimal means, to demonstrate an experimental and scientific musical scenario in a public setting. The public setting involved an audience with mixed degrees of immersion. While the entire audience listened to the music performed in the physical space, only one spectator used a HMD to visually explore the virtual environment. Multi-user VR could have increased levels of social presence or feelings of unity and togetherness in the public performance (Onderdijk et al., 2021), but would have required a significantly more complex setup dealing with challenges of synchronisation, latency (Ruan & Xie, 2021), and questions of virtual embodiment.

An interesting observation in the performance was that spectators wearing the HMD walked around the virtual point-cloud pianos even though they were not physically present in the space. This might have been an indicator of physical presence, the fact that despite the minimal and futuristic XR design (a plane, two pianos, an avatar, an agent, and a starry night sky), the spectator did feel and behave as if being in the virtual environment. This could be confirmed by inducing and observing breaks in presence (Slater & Steed, 2000), for example by having the spectator bump into someone from the passive audience or into virtual objects. Another interesting question is whether the physical co-location of spectator and performer when entering the virtual environment might have helped to induce social presence and consequently strengthened the physical presence in VR.

Finally, a contribution of this project is the successful realisation of a real-time performance between a fully embodied virtual avatar and agent. Fully embodied agents have great potential as virtual musicians, both for scientific research of the underlying behavioural models, as for learning, education, and artistic performances (for example, see section 6.4 or 7.4.1). They may provide rich (bio)feedback on human players themselves, as a "digital mirror" for improvement of playing techniques (Caruso et al., 2016), or as challenger and stimulator of creativity (N. Collins, 2011; Impett et al., 2015). The refinement of the algorithms that control human-agent interactions remains a significant research challenge (Iqbal & Riek, 2018; Tognoli et al., 2020), and placing these agents in public contexts could provide valuable feedback on the strengths and weaknesses of the underlying behavioural algorithms.

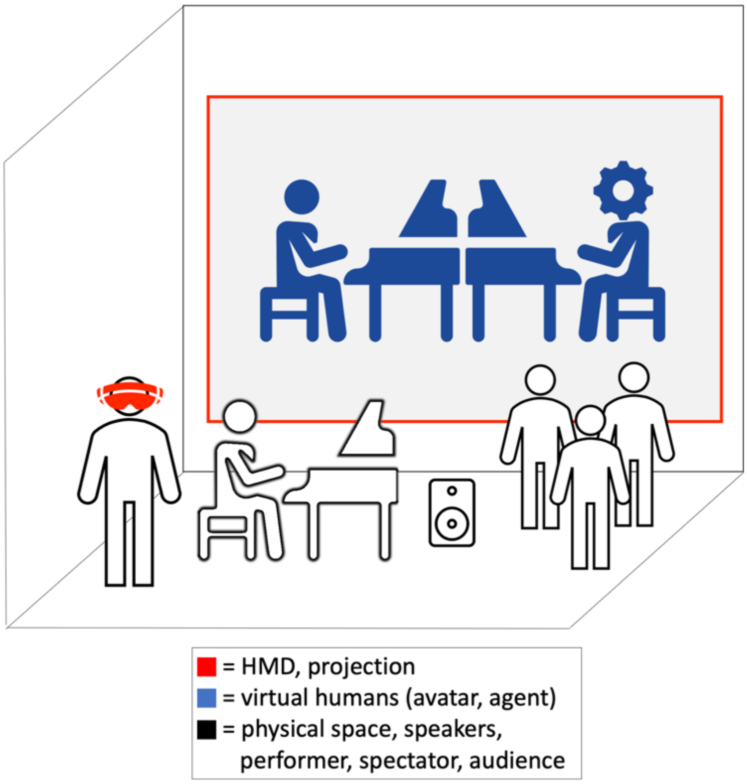


Figure 7-1: Schematic XR design

Figure 7-2: The live performance of Piano Phase in ASIL (performer, passive audience, active spectator wearing a HMD, and projection of the spectator's 1st person view)

7.2.2 Music moves

Authors: Bavo Van Kerrebroeck, Niels Van Kets, Bart Moens, Glenn Van Wallendael, Pieter-Jan Maes

Abstract: XR technologies offer numerous promising avenues for the enhancement and expansion of musical performance. They can bring performers and audience together in networked, virtual performance spaces and offer tools for the design and modelling of these spaces, thus shaping existing or creating new ways of musical expression. However, several aspects of such musical performances in XR would benefit from closer examination. First, musical interactions in XR could still benefit from improved controls to improve expressivity and nuance (Berthaut, 2020). Second, while non-verbal communication such as hand gestures and facial expressions are possible, they are often isolated from full-body interactions which are needed to facilitate an embodied virtual presence (Kilteni, Groten, et al., 2013). Offering a virtual body to humans for their interactions in XR is important, as it has consequences on humans' experience and behaviour (Gonçalves et al., 2022; Kilteni, Bergstrom, et al., 2013; Sanchez-Vives & Slater, 2005). Finally, fine-grained social interaction between avatars in real-time remains a challenge due to technical constraints in computer power, networking requirements, and latency (Ruan & Xie, 2021).

This study addressed these challenges by building on the XR implementation developed for the experimental study presented in chapter 4 to enable real-time, networked, and multi-user interactions using human-controlled, full-body avatars. These were then demonstrated in a public music performance that was presented to a hybrid, online and offline audience. The performance was part of the symposium "Onze (on)bekabelde cultuur"[10] and involved two physically separated professional drummers that engaged in a polyrhythmic musical interaction. Drummers were situated approximately 130 meters from each other, one in a public theatre hall and one in ASIL. The performance had 16 people attending physically in the theatre space and 16 people watching the performance streamed online (see Figure 7-3 and Figure 7-4 below).

Performers improvised over a binary-ternary polyrhythm with a backing track and were supported in their performance by a virtual drum circle with rotation spheres. Performers were motion tracked in both rooms and wore Hololens HMDs to interact with each other as full-body avatars in AR. In addition to the individual visualization in the Hololenses, avatars were also placed in a virtual environment designed as a digital twin of ASIL. Video feeds from both locations, the virtual environment, and visual renderings in the Hololens HMDs were placed in a video collage demonstrating the different perspectives present in the performance. Collaged perspectives included the 1st person performers' perspectives, 3rd person real world perspectives, and 3rd person perspective on the virtual environment (See Figure 7-4). The video collage was streamed using Zoom software for the online audience and projected behind the theatre stage for the offline audience. This allowed both audiences to see both players in the virtual environment, with the offline audience seated in the theatre also seeing one player physically present in the same space. A schematic overview of the performance space can be seen in Figure 7-3.

XR design and relevance: This performance demonstrated the feasibility of XR technology to enable real-time, networked, and multi-user musical interactions using full-body movements of human-controlled avatars. It also demonstrated the use of a virtual musical instrument to facilitate interpersonal music coordination and a transparent approach to share perspectives in an XR performance to a hybrid, online and offline audience. While individual XR components of the performance were relatively simple to implement, integrating the whole system while keeping data streams synchronized and within latency and bandwidth bounds represented a bigger challenge.

The XR networking setup and visuals were built in the Unity game-engine and synchronised using the Photon Unity Networking framework. A networking switch was added in the theatre space to connect both buildings over a 10Gbps fibre optic cable. Switches were configured such that the local area network of ASIL encompassed the

---

[10] The symposium took place on the 30th of January 2022, more information can be found here: https://www.ugent.be/lw/kunstwetenschappen/ipem/en/research/projects/ongoing-projects/asilminardproject (Accessed on the 20th of November 2022)

theatre space, ensuring all devices were on the same network and had the lowest latency possible. To produce online video streaming to an audience, two Panasonic N-series PTZ cameras were installed at both locations, capable of streaming uncompressed video over Network Device Interface, a low-latency realtime video transmission protocol. The video collage seen in Figure 7-4 was created using the OBS software suite.

There were several crucial latencies of audio and visual data streams in the XR system. Latency from drum pad to speaker audio output for both performers was 20ms. Latency from the performer’s movement to visualisation in the shared environment was 80ms. Latency from performer’s movement to visualisation in the other performer’s Hololens was 120ms. Although latencies were relatively high, performers reported these did not impair the performance. Third person video feeds of performers on the top-left and top-right of the collage had 180ms latency from movement of a drummer to broadcast in Zoom. Visual streams from the Hololens HMDs were collected from the Microsoft Hololens companion app which increased their latencies to 205ms. This did not impact the interaction between both performers as these latencies were only visible for the audience.

With social activity increasingly taking place in hybrid online-offline, real-virtual forms and environments, the design as proposed in this project offers an approach to facilitate musical dynamics, as well as present them to, and share them with the broader public. In addition to optimizing the current system, several improvements could lead to diverse experiences of performers and audience, and offer new avenues for scientific and artistic experimentation and exploration.

First, one could further adapt the appearance and representation of the participants in the system (Zibrek et al., 2019). For instance, one could increase realism working towards increasingly photorealistic representations, avoid the use of generic avatars with real-time volumetric captures[11], or intentionally move towards more abstract representations. While the system presented here focused on representing movement, one could also include other expressive cues using for example facial expression.

Second, the XR performance could be extended beyond dyadic interactions. When working with more than two users, one could place multiple participants in the same room, avoiding the need to duplicate the motion tracking setup, and simultaneously creating rich blends of human-human, human-avatar, and extended-reality group

---

[11] See for example Microsoft Mesh: https://www.microsoft.com/en-us/mesh?rtc=1 (Accessed on the 20th of November 2022)

interactions. Inserting computer-controlled agents would allow for an even more diverse set of possible interactions in the virtual space (Kostrubiec et al., 2015).

Finally, while our focus in this paper lay on the design and facilitation of expressive interaction, future implementations will pay more attention to spatial sound design to induce attention as well as modulate feelings of presence (Kobayashi et al., 2015; Salselas et al., 2021; Shin et al., 2019). In addition, future work will continue the development with new instruments, such as shared virtual objects that participants can throw to each other (see the end of the "MusicMoves: additional features" video online[9]).

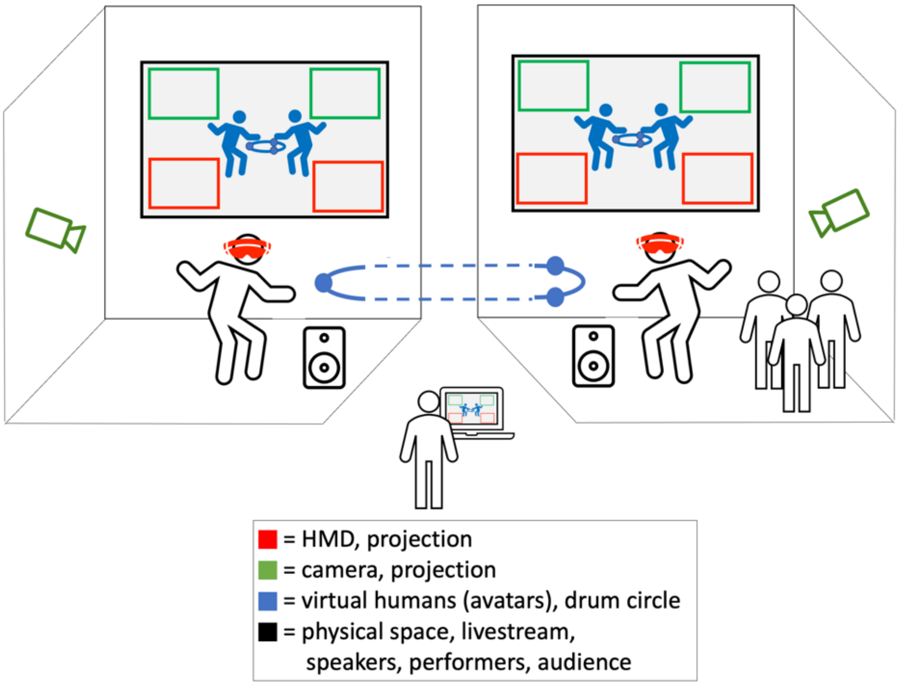


Figure 7-3: Schematic XR view

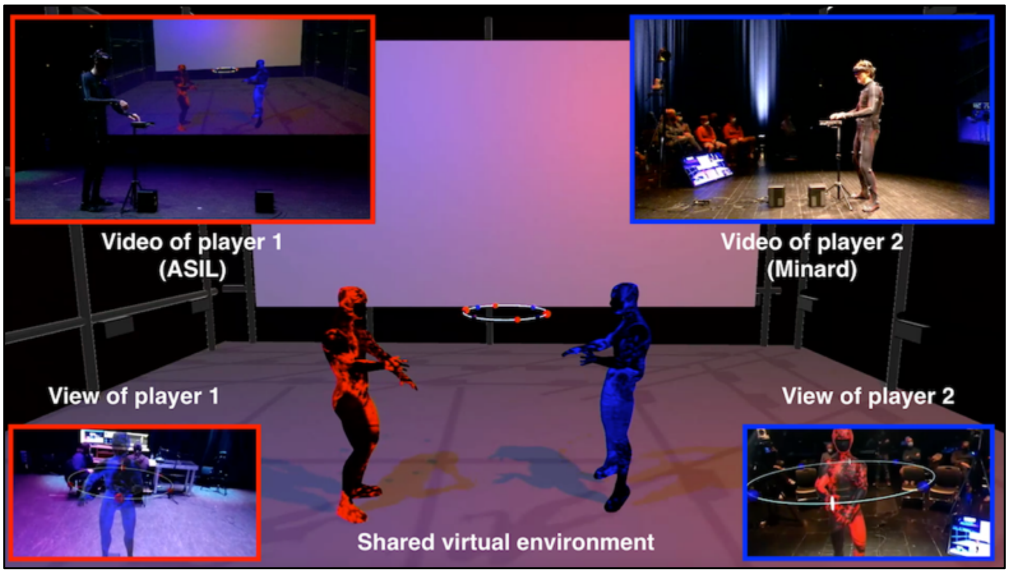


Figure 7-4: Annotated video collage that was projected and streamed for the offline and online audience

7.3 Embodied, expressive composition

XR can be seen as an ideal medium for the spatial composition of music. By immersing its user in a virtual environment, it allows the merging of audio and visual renderings of expressive musical actions, and can offer instructive feedback through tight action-perception couplings. Moreover, by placing new and enriching affordances in the virtual environment, it can provide mental support, suggestions, and inspiration in the creative process through algorithmic-aided composition. Ideally, design of the virtual environment and composition of the musical work are carried out in close collaboration, as innovations in one can inform and shape the other. Through a modular design, such as the one proposed in the study presented below, interpretation of the musical material and mental imagery provoked by it is designed, developed, integrated, and finally enacted in new affordances of the virtual environment.

7.3.1 Lines and swarms

Authors: Bavo Van Kerrebroeck, Pieter-Jan Maes

Abstract: Since the advent of electronic and electro-acoustic music, audio playback modes evolved from mono, stereo, multichannel, to 3D sound. These advances led to the development of compositional and performance practices in which the "spatiality of sound" functions as a primary musical parameter, alongside pitch, tone duration, timbre, and dynamics (Schmele & Finney, 2011). Sound spatialization offers rich, additional ways for the expression of emotions, creative narratives, and imaginative thoughts in music (Hagan, 2008). A core question pertains to the arrangement and control of sound trajectories in 3D space. In line with the embodied music cognition theory, a key role can be attributed to bodily gestures in musical expression and sense-making (Leman, 2007; Lesaffre et al., 2017). Existing interfaces for the control of sound trajectories in space have included various gestural spatialization controllers (Pysiewicz & Weinzierl, 2016).

This study aimed at complementing the predominantly technological focus of these gestural interfaces with (1) artistic knowledge of the music composition, and (2) scientific knowledge in the domain of embodied music cognition and interaction. The aim was to fully exploit the artistic and creative potential hidden in the musical material by rendering gestural, imaginative expressions and repertoires into corresponding sound trajectories in 3D space. In addition, a core goal of the project was to explore the complementarity of VR visual displays in the arrangement and experience of gesture-based 3D sound trajectories. VR visual displays can capture volatile gestures and their sound traces into corresponding visual traces in 3D space, experienced from an immersive, first-person perspective. This extension of 3D sound arrangement into interactive and immersive visual VR environments can then incite new sound art experiences, practices, and interactions (see section 7.4.1 for an example).

The outcome of the study was a spatial music composition tool in which the user or composer could draw, edit, enrich, capture, design, and embody expressive gestural trajectories in VR (see Figure 7-5 and Figure 7-6). Composition took place in ASIL to allow spatial audio playback using its 80-speaker system. The composition performance started when the user put on a HMD and opened the “Lines and Swarms” application. Users entered a “digital twin” of ASIL and chose an affordance using a hand-held VR controller from a user interface (UI). Users would choose an audio sample on which to work, that started playing (binaurally) as soon as users started to create a trajectory. Users could choose between several options.

- **Draw:** An intuitive option in which the spatiotemporal position of the controller is followed to create audiovisual trajectories.
- **Enrich:** An algorithmically predefined trajectory (for example through random walk algorithms) is visualised in space and users add a velocity profile by aligning and moving the controller onto it.
- **Capture:** A swarm of birds is controlled using a swarm algorithm[12] of which one bird draws a visual trajectory in its flight path. Users choose trajectory parameters such as speed and trajectory length, as well as parameters governing global swarm behaviour such as alignment and repulsion forces. Users capture and store trajectories they are interested in for later spatial audio playback.
- **Structure:** Several parameters are offered in the UI to create a geometric shape of which corners represent spatial sound sources and trajectories. Users set the shape’s dimension and number of planes as well as 3D rotation velocities.
- **Embody:** An option to draw trajectories using hand tracking was added to alleviate the use of a hand-held controller and allow more natural, free movements. This did constrain hand tracking to the HMDs field of view.
- **Map**: This option only visualised up to 50.000 particles controlled by a 2D slime mould algorithm[13] that was transformed to work with 3D. Users could manipulate algorithmic parameters that allowed the emergence of patterns and spatiotemporal trajectories in the slime simulation.
- **Edit:** Users could always edit 3D trajectories expressed in space by scaling them using a UI slider or moving them by grabbing and displacing a virtual anchor.

---

[12] Swarm algorithm adapted from https://github.com/chris-kirkham/boids (Accessed on the 20th of November 2022)

[13] 2D version of the slime mould algorithm: https://github.com/SebLague/Slime-Simulation (Accessed on the 20th of November 2022)

- **Display:** After a trajectory was created, users could store it in the application. A display option was offered in the UI to visualize all stored trajectories.

XR design and relevance: The compositional tool was developed as a Unity stand-alone application and ran on the Oculus Quest 2. Audio samples in Ableton were played back as stereo sound while composing, and were played back as 3rd order Ambisonics spatial sound afterwards. Sound trajectories were stored in the application as text files and imported through a Python script into Ableton as automation. Communication between Unity and Ableton was done using the OSC protocol.

The compositional tool developed in this study leveraged expressive gestures to draw trajectories and offered users the options to evaluate and modify them using visual and auditory feedback. Future work could integrate algorithms and toolkits for gesture recognition and prediction to assist the user further in the compositional process (Caramiaux et al., 2022; Hilton et al., 2021). Doing so would reduce the barriers to musical expression for novices, assist experts, as well as improve the algorithms themselves.

Users could draw trajectories using a hand-held controller or free movement based on hand tracking of the HMD. For the latter, given that tracking only worked when hands were inside the HMDs field of view, users moved as if their hands were leading their bodies. Given these differences in expressive movement originate from imposed technological constraints, a tool such as the one proposed in this study could help to better understand the underlying impact of physical and virtual affordances, as well as lead to better compositional tools leveraging fluid movement for enhanced creativity (Slepian & Ambady, 2012).

Finally, while the tool in this study was for individual use only, future work could introduce a social component in the tool. One could allow multiple users to enter the compositional space and have them work collaboratively on the musical work (Barrass & Barrass, 2006; Men & Bryan-Kinns, 2019). Another way would be to introduce spectators or actors in the virtual space, having composers improvise their expressive musical trajectories for an audience, blending the musical spaces for composition and performance (Ciciliani, 2020; Wozniewski et al., 2018).

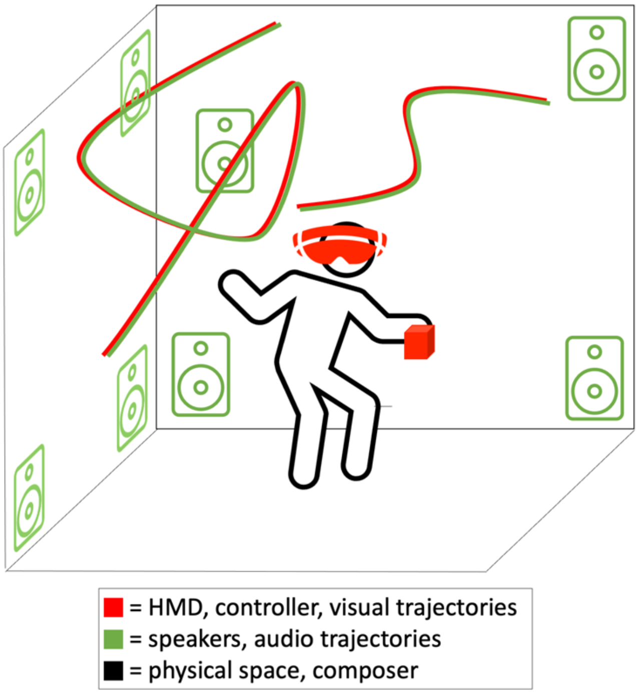


Figure 7-5: Schematic XR design

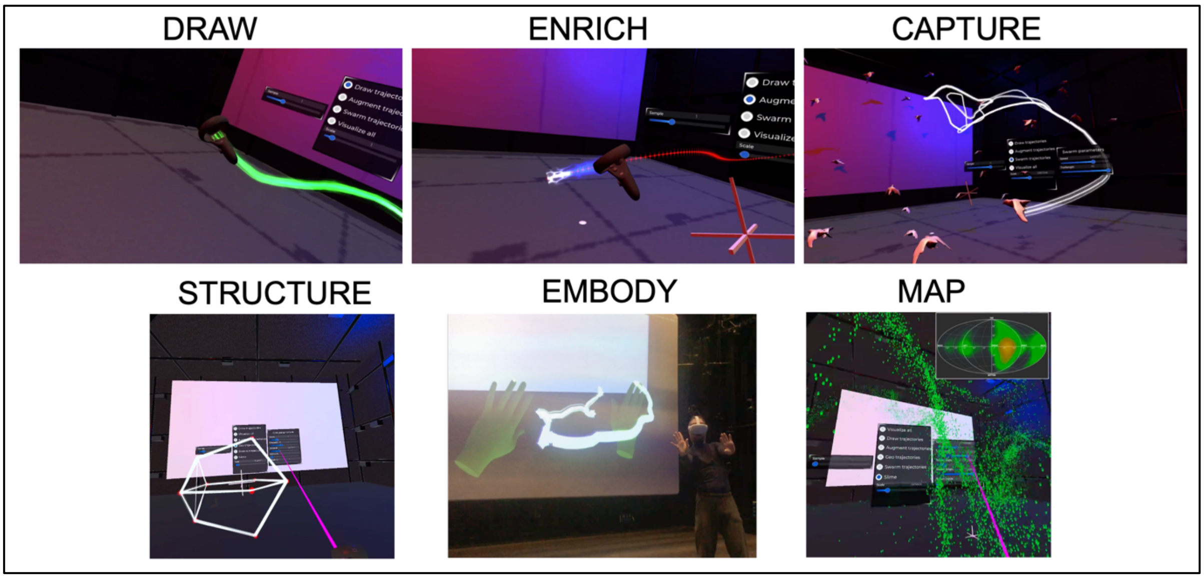


Figure 7-6: Overview of the different affordances offered in the application

7.4 Immersive, interactive narrations

Music is powerful as exemplified by its ability to modulate listeners' experience of space and time (Schäfer et al., 2013) and evoke strong emotions (Juslin & Sloboda, 2013). Narration or storytelling is widely considered a central element for the construction of individual and collective meaning, and a fundamental cognitive mode for organizing human experience (Bruni et al., 2022). Combining both in XR could allow the emergence of synergies that leverage the experiential powers of both. This section presents three artistic studies in which spatial music, soundscapes, and immersion are combined to illustrate key capabilities of XR for narration such as virtual embodiment, multimodal integration, and the creation of social, interactive spaces.

7.4.1 Being hungry

Authors: Bavo Van Kerrebroeck, Celien Hermans, Pieter-Jan Maes

Abstract: In 2020, Tineke De Meyer and Duncan Speakman released their album "The House was Alright"[14]. All musical works on their album were originally stereo-mixed. This project used the spatial composition tool presented in section 7.3.1 to recreate one particular work of the album, "Being Hungry", into a 3D spatialized audio version to be presented in ASIL and the Mozilla Hubs environment. The project encompassed two main parts: creation of the 3D audio work and presentation in an immersive (virtual) environment.

'Being Hungry' consists of 7 audio layers in a multitrack composition which were stereo-mixed in the original release. To recreate this work for a spatial composition, audio tracks were first segmented into individual sound entities to assign 3D audio trajectories. The design of audio trajectories symbolized the underlying inspiration of the work "Being Hungry":

> *"'Being Hungry' was written at the beginning of spring as some of us entered lockdown. Inspired by the book Jonathan Livingston Seagull, by Richard Bach, it imagines a small group of people that decided to devote themselves to the practice of just walking, a passion for moving around without a distinctive aim."*[14]

The resulting work was presented as part of an XR music call at the ICMC 2021 conference[15] and as a hybrid demo on the SysMus22 conference[16]. The online version of "Being Hungry" was presented as a multimodal room in the Mozilla Hubs environment[17] (for an impression, see Figure 7-8). Visitors entered the room using an internet browser and could then navigate into one of two virtual spaces, a digital twin of ASIL reflecting the compositional space, or a narration space in which to experience the spatialized audio work. The compositional space contained pictures of ASIL and a 360-video with stereo sound illustrating the use of the compositional tool presented in section 7.3.1. The

---

[14] https://dspk.bandcamp.com/album/the-house-was-alright (Accessed on the 20th of November 2022)

[15] XR music works accepted at ICMC are listed here: https://ccrma.stanford.edu/~rob/xrmusic/xr_programnotes.html (Accessed on the 20th of November 2022)

[16] The conference program can be found here: https://www.sysmus22.ugent.be/program/ (Accessed on the 20th of November 2022):

[17] The room can be visited through this link: https://hubs.mozilla.com/Z2qbhLA (Accessed on the 20th of November 2022)

narration space contained a video projection created to accompany the work, an animation of an agent engaged “to the practice of just walking” based on motion captures done in ASIL, a binaural down-mix of the spatialized work, and spheres following the 3D audio trajectories. For the SysMus22 conference, the work was presented simultaneously on the Mozilla Hubs environment as well as played back using 3rd order Ambisonics sound in ASIL (see Figure 7-7).

XR design and relevance: A central place in this project was given to the notion of a shared space that crossed physical and virtual boundaries. This space arose out of the underlying meaning of the musical material and was co-constructed by spectators fulfilling both the role of audience and actors, “moving around without a distinctive aim”[14]. Visitors moved around physically in ASIL, virtually as avatars in Mozilla Hubs, and alongside a virtual agent embodying pre-recorded walks. XR enabled this narration by immersing visitors in an auditory soundscape while offering a virtual meeting place inducing physical and social presence. XR technology was developed to allow an embodied interpretation of the musical material (see section 7.3.1). This then stimulated the design of new affordances in the compositional tool in an iterative and co-creative process of the technological mediators. This project thus illustrates how XR allows the integration of artistic knowledge of the underlying meaning in musical material and the realisation of an extended musical space for composition and narration.

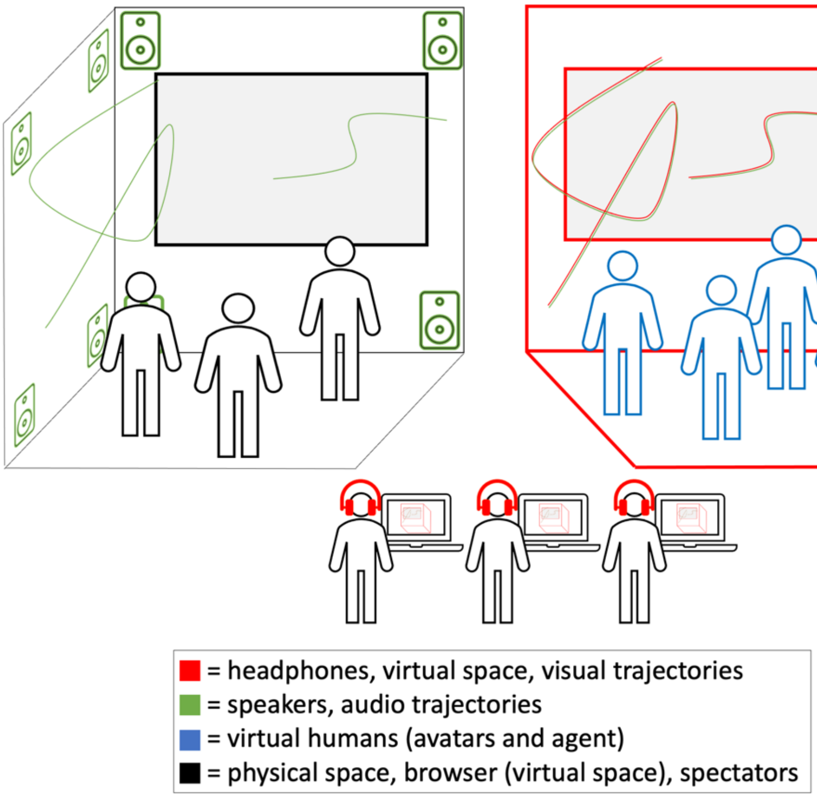


Figure 7-7: Schematic XR design

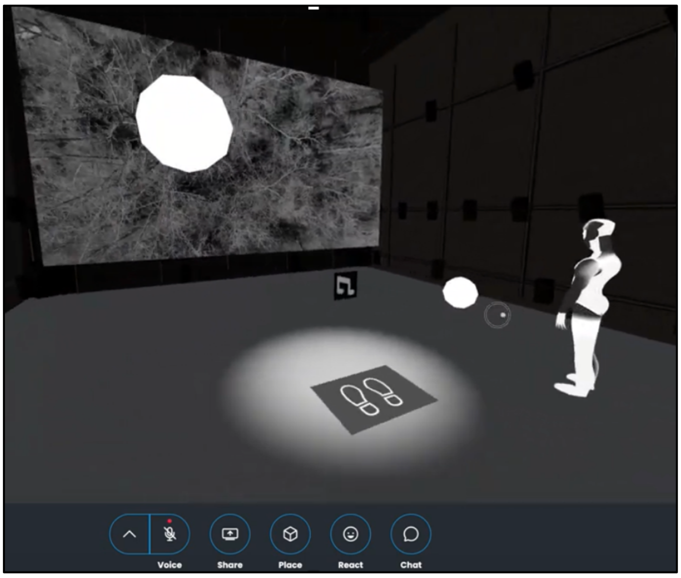


Figure 7-8: 1st person view of a visitor in the online Mozilla Hubs room

7.4.2 Searching borders

Authors: Bavo Van Kerrebroeck, Sarah De Mulder, Bras Lareu

Abstract: This study developed a multimodal, immersive installation with interactive storytelling in MR. The aim was to have a visitor explore the virtuality continuum by modulating virtual embodiment and presence. The project's title, "searching borders", alludes to the fictional division between opposites, and the mind's propensity to oscillate between dualistic terms such as the real-virtual division. The study was developed in the context of a 48-hour hackathon on immersive sound, organized by the record label Consouling Sounds[18] and the XR art collective XRT[19], both located in Ghent.

Inspired by the principles of coordination dynamics and complexity theory (De Wolf & Holvoet, 2004; Kelso, 2009), this project started from a view on conscious experience as emerging out of a dynamical process embedded in a landscape of emotional, social, and environmental forces and attractions. By immersing a person in an environment with a landscape of real and virtual impressions, affordances, and interactions, they are challenged to move to specific states or, as was the interest of this project, to linger between them. Concretely, it was the goal of this project to investigate how a person's presence and embodiment can be modulated by having a visitor shift between 1st and 3rd person perspective, by blending real with virtual elements in the environment, to play live

---

[18] https://consoulingpresents.be/tinnenpot, https://consouling.be/de-console (Accessed on the 20th of November 2022)

[19] https://xrtimmersive.com/ (Accessed on the 20th of November 2022)

and recorded sounds, to move from feeling localised in physical space to feeling present in virtual worlds, to feel embodied at certain times, and disembodied at another.

The following paragraph will describe the narration storyline, of which a schematic overview can be seen in Figure 7-9. A 1st person impression of the installation can be seen in Figure 7-10 as well as in the online video playlist[9].

Upon starting the experience, visitors first received a sound transducer to hold in their hands that converted sounds from a microphone in the physical space to haptic feedback. They were then asked to put on a HMD to enter the virtual world. Visitor's hands were tracked and visualized when in their field of view. Visitors entered a dark space, seeing a text with a button in the distance. When walking towards it and pressing the button, they would see themselves in 3rd person as captured by a video camera in the physical space. The text then asked visitors to press again, which removed the video projection and allowed visitors to step into another dark room with a text and button. When pressing this button, a speaker in the space started playing an ambient soundscape, and sounds captured by the microphone in the physical space were now visualised as real-time spectrograms in the virtual space, next to being haptically delivered by the transducer. Finally, when pressing the button again, visitors saw a virtual gate with a text inviting them to walk through it. When approaching the gate, a virtual fence was visualized. When stepped through, it enabled pass-through cameras that displayed the physical space in the HMD.

XR design and relevance: A first distinction of this study is the specific use of MR by offering real elements in the virtual space (i.e. augmented virtuality), placing the user in full virtual immersion (i.e. VR), as well as allowing the user to enable pass-through cameras to leave the virtual environment (i.e. AR). Several XR capabilities such as virtual embodiment, body illusions, and multimodal feedback to modulate physical presence were used as narrative devices in this installation to, as mentioned before, have a visitor linger on the virtuality continuum.

For instance, the visitor would see him or herself from a 3rd person perspective to hint at an out-of-body illusion. Although the required visuotactile stimulation was not administered (Petkova & Ehrsson, 2008), the perspective was skewed (see Figure 7-10) with the intention to disturb the viewer and provoke reflections on self and physical presence. Hand tracking and visualizations were also introduced to invoke a degree of self-presence as well as allow interaction with the environment. Presence might have been severely disrupted at the end of the narration in breaking the "4th wall", when pass-through cameras were enabled and visitors would see their real bodies and hands.

Multimodal feedback through haptic and visual rendering of "live" sounds were used to modulate the sense of physical presence. They were aimed at bringing the visitor back from a virtual immersion towards a physical grounding of the visitor in the real world. This was especially at work when visitors realised the sounds they were making themselves (coughs, footsteps, laughter) were immediately fed back into the virtual visual and haptic displays. Sound might be especially effective for this purpose as they play back in the same 3D environment as the listener and may thus more effectively blur the lines between real and virtual (Garner & Grimshaw, 2014). While there were no sounds placed in the virtual environment, the integration of headphones allowing the switch between open-back or noise-cancelling and working alongside speaker playback in the physical space, could have blurred the real-virtual line even further. They could have directed the attention of the visitor to the real or virtual space, highlighted (virtual) affordances, and alternated doing with being, introspection with action and perception (Atherton & Wang, 2020).

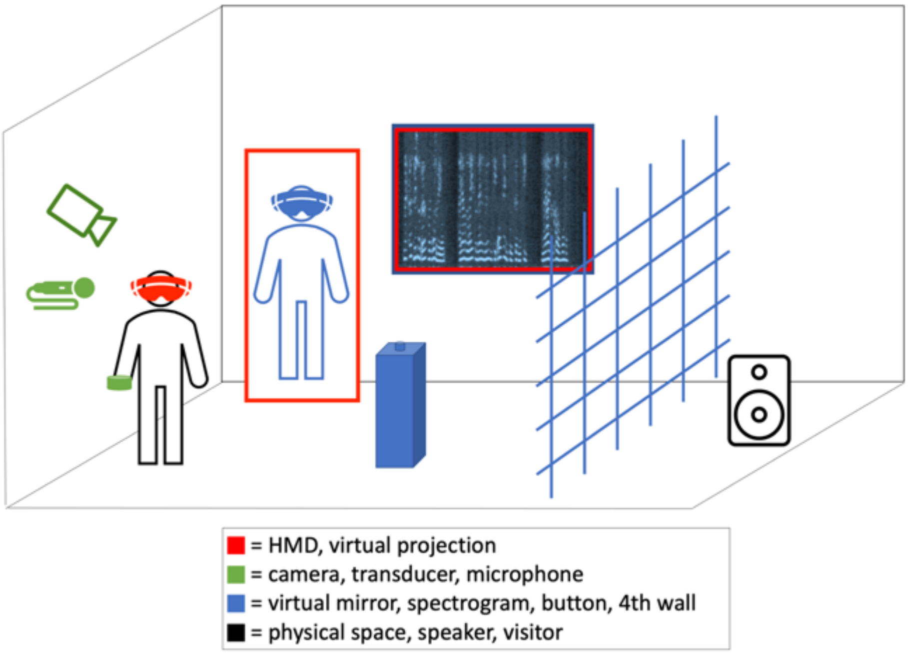


Figure 7-9: Schematic XR design

Figure 7-10: 1st person view of a visitor showing a 2rd person view of visitor on himself and white virtual hands

7.4.3 Be hear now

Authors: Lennert Carmen, Bavo Van Kerrebroeck, Xander Steenbrugge, Pieter-Jan Maes

Abstract: This study aimed at exploring the use of immersive audio and visuals to restage and recreate a meaningful experience and activity happening "in-the-wild", such as exploring the historical city of Ghent. For this purpose, a mixture of audio and visual material of the city of Ghent was captured, processed, and displayed as an interactive and immersive "cityscape" consisting of three experiential chapters: water, culture, and night. Visitors were motion tracked and stepped on floor-projected words "Here" and "Hear" to trigger visuals or spatially processed field recordings from the city of Ghent (see Figure 7-11 and Figure 7-12).

At the start of the narration, visitors entered the room and were immersed in a mix of Ambisonic field recordings. After a minute, they heard windows closing after which three words were projected on the floor. When a visitor stepped on one of the words, they triggered one of three cityscapes. Each cityscape lasted for about four minutes after which they were brought back to choose one of the remaining cityscapes. During a cityscape, visitors could again step on floor-projected words, this time to start a video or lift audio filters on spatially rendered audio. The latter was done to give the impression of windows opening and letting cityscape sounds enter the room. The final scene ended when visitors stepped towards the centre which then showed all floor-projected words in silence.

Field recordings were recorded at several places in Ghent as Ambisonic multichannel or as stereo sounds. They were mixed and mastered, and stereo sources were positioned spatially. Individual audio sources and narrative elements such as for example a bicycle sound or a bird call were carefully chosen and (spatially) mixed as inspired by the acoustic niche hypothesis (Krause, 1993). This hypothesis highlights the careful timbral structure of a soundscape and states how animal species avoid competition by occupying specific sound spectra at specific times. Video was rendered as transformations between pictures of the city of Ghent, generated using deep-learning generative adversarial networks (Aldausari et al., 2023). Visitors were asked to wear a motion-tracked hat to detect when they stepped on floor-projected words.

The narration was shown in two instances, once as an individual experience in ASIL and once as an online experience on the Mozilla Hubs platform. The individual experience had around 20 visitors. For the online experience, the narration was reduced to a minimal, non-interactive implementation and binaural mix. This experience was demonstrated to around 40 visitors during the SysMus22 conference. The online version is available here: https://hubs.mozilla.com/3ocpWXV

XR design and relevance: This study is relevant for XR research as it highlights the conceptual differences between immersion and presence in dealing with the auditory sense. Immersion as defined in this thesis, refers to the technical fidelity of the system, while presence refers to experiential aspects induced in its users, “as a reaction to immersion” (Slater, Lotto, et al., 2009). Auditory immersion is high in this study due to the use of high-fidelity spatial sound recordings and playback, while visual immersion can be considered low due to simple projection on a 2D screen. Auditory immersion, resulting from the playback of field recordings, might have induced mental imagery and mental transportation to pre-recorded physical spaces. It thus allowed us to work on the “place illusion” component of presence conceptualized as physical presence (K. M. Lee, 2004; Slater, Lotto, et al., 2009).

Self-presence, social presence, and agency of the user can be assumed to differ across the individual, offline and the shared, online experience. While self-presence and social presence might have been respectively very high and absent for the individual experience, visitors did receive a simple, virtual avatar and controlled it in the online, shared Mozilla Hubs experience. While self-presence might have been very low, the avatar had minimal levels of appearance and behavioural realism with control over the avatar done using the computer keyboard, social presence might have been higher as visitors could see each other’s avatars and communicate using their computer microphones. The physical and online experience also differed in the control and resulting sense of agency they gave visitors over the narrative. Visitors in the physical experience could choose when to start a new chapter or cityscape, but could only choose their point of view on a cityscape as passive spectators in the online experience. While the experience of narration in both physical and virtual places poses severe technical and conceptual challenges, XR offers the methodological and technical means to address them and retain or enable a musical narration’s meaning-laden form and content.

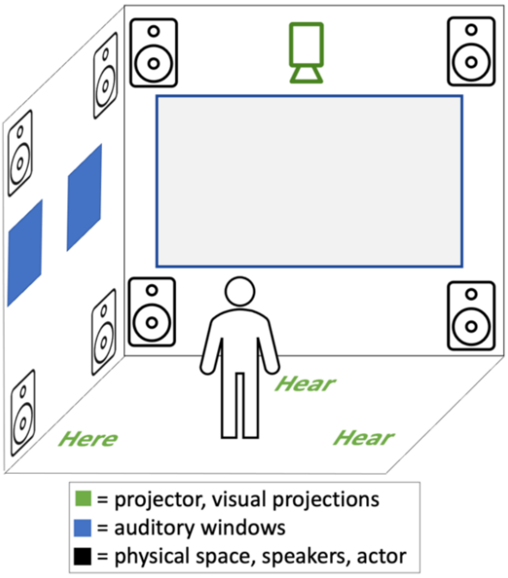


Figure 7-11: Schematic XR design

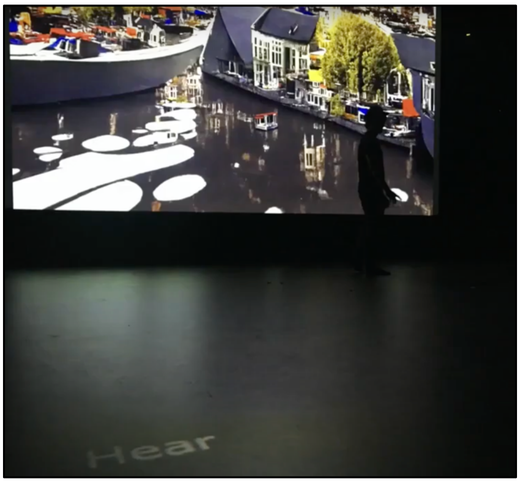


Figure 7-12: The narration showing a visitor, the video projection and floor-projected word

8.
General discussion

8.1 Contributions

This research has presented an integrated and comprehensive methodology of using XR for scientific and artistic music research and demonstrated its value in a selection of experimental studies. It has put forth the notion of the "extended musical space", the active, embodied, social, and digitally augmented musical environment in which composers, performers, spectators, and participants can meaningfully interact and express themselves (Cook, 2013; Leman, 2007). Four extended musical spaces were conceptualized: a space for experimentation, for performance, for composition, and for narration. The majority of this thesis has been concerned with the design, development, and empirical testing of these musical spaces in a pursuit to answer the research question and goals presented in Table 8-1 below.

Main question: How can XR technology contribute to the development of musical environments that enrich embodied and collective music making?	
Scientific goal	Explore XR's potential for an improved understanding of the dynamical processes and subjective experiences in music making
Artistic goal	Explore XR's potential for novel expressions and experiences in music making

Table 8-1: Main research question

Music making, with its requirement of fine-tuned and skilful co-regulation of bodily articulation (Leman, 2007; Leman & Maes, 2015), has been the ideal playground for investigations into XR's potential for embodied, social cognition and interaction research. XR technology, with its capability to immerse its users in computer-generated, multisensorial, and interactive environments while delivering a sense of presence (Sanchez-Vives & Slater, 2005; Slater, Lotto, et al., 2009), has allowed the "hacking" of continuous and dynamic action-perception cycles underlying co-regulation, effectively handing over experimental control of conscious experience to the researcher (T. K. Metzinger, 2018).

The extended musical spaces were enabled by XR technology and built on the theories of embodied music cognition, coordination dynamics, and human-computer interaction introduced in section 1.3. The embodied music cognition framework has allowed an improved understanding of multimodal integration and a view of the body as a source of expressive and intentional communication (Leman, 2007). It has thus proven to be a fertile basis for XR research, stimulating new experimental paradigms through body illusions induced by virtual embodiment (De Oliveira et al., 2016; Ehrsson, 2007; Petkova & Ehrsson, 2008). The framework of coordination dynamics has offered the methodological tools to deal with the spatiotemporal, non-stationary, multi-scaled and multi-layered co-regulation processes (Kelso, 2009), offered the terminology capable to deal with the surprising, playful, and metastable nature of music making (Borgo, 2005; Demos et al., 2014; A. E. Walton et al., 2018), and helped to develop mathematical models for behavioural control of virtual agents or other non-humanoid, adaptive (bio)feedback (Tognoli et al., 2020). Finally, the field of human-computer interaction has offered the interaction models and design principles for novel, useful, and inspiring virtual affordances and instruments (Berthaut, 2020; Serafin et al., 2016; Tanaka, 2019) as well as the engineering and technical baselines for developing networked, real-time, and ubiquitous interactions between humans and machines (Ruan & Xie, 2021; Turchet & Rottondi, 2022).

As mentioned before, four musical spaces were introduced in this thesis. The majority of this thesis was situated in the space for experimentation, with specific research questions and investigative outcomes presented below. A core focus was to improve our understanding of the coregulatory and experiential processes in music making using XR by developing the theory (chapter 2), methodology (chapter 3), and empirical analysis (chapter 4) to do so. This investigation started in chapter 2, which was dedicated to answer the following question.

Question chapter 2: how to improve our understanding of musical sense-making?

An answer to this question should address two main challenges central to social cognition and music research. First, it has to deal with the large variability in observed spatiotemporal (bodily and physiological) patterns, contexts, and personal characteristics. Second, as music making constitutes a meaning-laden and intense subjectively experienced activity, it has to deal with coordinative patterns and the emerging subjective qualities, that are methodologically hard to reach with quantitative, empirical methods. To tackle these challenges, chapter 2 proposed a theoretical paradigm consisting of a mixed-method approach that combined quantitative and qualitative methods and insights inspired from the embodied music cognition (Leman, 2007) and coordination dynamics (Kelso, 2009) frameworks. Regarding the coregulatory processes between musical performers, the paradigm emphasized the need for a search towards underlying organisational principles, as opposed to merely looking at observable patterns. It stressed the time varying nature of these processes and the need for a systems level perspective. This perspective would treat the interacting performers as a whole, allowing integration of the intersubjective, co-created, and emergent intentionalities and experiential states in the musical interaction such as, for example, shared agency and group flow (Cochrane, 2017; Loehr, 2022; Schiavio & De Jaegher, 2017). Importantly, relationships between the quantitative coordination patterns and the subjectively felt qualities are crucial to the paradigm, making it suited for an action-oriented and interdisciplinary stance in future social cognition, interaction, joint-action, and music research (De Jaegher et al., 2010; Engel et al., 2016; Sebanz & Knoblich, 2021).

Combining this theoretical paradigm with XR could have tremendous benefits for experimental music and social cognition research through the simulation of existing, or the creation of new musical contexts. A fundamental question that must be addressed first though, is whether these new contexts, in the form of digitally augmented environments or extended musical spaces, are capable of inducing social presence by creating the fertile conditions for successful co-regulation between performers. Given the challenges in the assessment of presence (Skarbez et al., 2017; Slater, Lotto, et al., 2009), chapter 3 aimed to answer the following question.

Question chapter 3: how can we evaluate social presence in musical interactions in XR?

The answer was given in the form of a methodological framework that conceptualized social presence as a multi-layered concept rooted in, and emerging from, the behavioural and experiential dynamics of music interactions. The framework captured coregulatory dynamics in a performative, embodied, and experiential layer, allowing assessment and analysis on multiple interaction levels, from low-level spontaneous coordination (Demos et al., 2019; Konvalinka et al., 2010), to higher-level learning, planning, and prediction (Koban et al., 2019; Sebanz & Knoblich, 2009). The chapter demonstrated the value of the framework by applying it to the evaluation of a musical interaction between two real and virtual, human- and algorithmically-controlled pianists. It was able to demonstrate successful, meaningful, and expressive music making between two virtual avatars by comparing the interaction dynamics to a real-life setting. In addition, it showed its potential for assessing human-agent music interactions by demonstrating their lack of expressive communication and adaptive coregulation. The proposed framework thus showed its value for the assessment and development of future virtual, embodied, dynamic, and creative human-avatar or human-agent interactions, which led to its use in an empirical study presented in chapter 4.

Question chapter 4: to what extent can XR technology facilitate adequate performance quality, embodied coregulation and subjective experiences in musical interactions?

This study used a virtual drum circle to guide and assist two performers in a polyrhythmic improvisation to investigate the impact of auditory context and visual coupling on music making. Using the framework introduced in chapter 3, it was able to demonstrate how performative and experiential dynamics did not deteriorate in XR, paving the way for further empirical music research in XR (Turchet et al., 2021). Furthermore, the framework integrated quantitative methods capable of dealing with the improvisational nature of the musical interaction, making it available for the metastable processes and adaptive flexibility present in meaningful music interactions (Borgo, 2005; Demos et al., 2014; A. E.

Walton et al., 2018). The study showed how performers moved more energetically and had stronger prosocial feelings of shared agency and self-other integration when seeing each other as fully embodied, virtual avatars as compared to not seeing each other. The embodied musical interaction might thus have benefited from social contagion and the prosocial effects of synchrony (Dotov et al., 2021; Tarr et al., 2018). A richer auditory context increased movement energy, performance quality, and feelings of agency and flow potentially stimulated by the motivating and rewarding aspects of engaging with music (Maes et al., 2016). While it was difficult to generalise results, given only 16 dyads with varying musical backgrounds were included in the study, the analysis did illustrate a comprehensive approach to obtain insights in social interaction and subjective experiences from the assessment of embodied, musical, and joint-action dynamics in XR.

Next to the simulation potential of XR and its use in fundamental research, it also has potential as biofeedback through its ability to deliver adaptive, multimodal stimuli based on coordination dynamics principles. Chapter 5 combined this potential with the power of music, leveraging its combined arousal and (spontaneous) synchronization effects (Bergstrom et al., 2014; Maes et al., 2019). This was done to investigate the use of an adaptive breathing sonification system to combat stress and induce relaxation states. Its main research question was as follows.

Question chapter 5: can we use a dynamic, immersive auditory environment to manipulate human behaviour and experience?

Using a distance-based, repeated measure experiment spread over multiple days, the study was able to show how an adaptive, spatial, natural soundscape worked best to slow down breathing. This successful behavioural manipulation was explained as working through experiential and arousal mechanisms corresponding to findings on the beneficial effects of immersive soundscapes (Aletta et al., 2018; Hume & Ahtamad, 2013; Medvedev et al., 2015). Results also showed how an adaptive, filtered pink noise sound was able to stimulate healthy breathing inhale-exhale ratios and reduce stress levels. Given that the noise stimulus resembled a human breathing sound, reduced breathing ratios might have resulted from spontaneous synchronization mechanisms (Murgia et al., 2016). While the study only involved an auditory component, an interesting extension would be to include a visual XR component, as immersive, natural scenes have already shown their beneficial effects on well-being (Naef et al., 2022). A strong outcome of the study was that the majority of the participants indicated they liked to continue the breathing exercises. Given that XR allows for easy replication and that the application developed in this study was downloaded online and executed at home, the study provides a ready basis for further research into musical XR used in biofeedback applications for well-being.

Taking a step back, chapter 6 and 7 aimed at demonstrating the broader potential of XR for scientific and artistic music research. This was done by presenting abstracts of four scientific and six artistic studies together with a discussion of their XR designs and relevance.

Question chapter 6: what are the opportunities for XR technology in experimental music research?

The four presented abstracts in chapter 6 used XR for an increased understanding of specific behavioural, affective, cognitive, and experiential aspects in music making. Two case studies took an integrated and modular approach, integrating neurophysiological and behavioural measurements in a controlled setting for fundamental music research of interpersonal coordination. One study leveraged XR's capability for body-illusions (Slater, Perez-Marcos, et al., 2009) to swap perspectives in a controlled tapping experiment. Its aim was to investigate cooperating and competitive tendencies in sensorimotor synchronization, with potential applications in rehabilitation and education (De Oliveira et al., 2016). Another study used (mimicked) movement and virtual embodiment in the form of agents and a virtual mirror paradigm (González-Franco et al., 2010) to model neurophysiological indicators. Its aim was to gain a deeper understanding of the link between emotional levels, music, and movement (Koelsch, 2015; Tarr et al., 2014). The two other studies focused on the use of XR in ecological, "live" contexts. The first study observed students in a virtual teacher paradigm to investigate the added value of AR for music learning (McIntire et al., 2014). The other study manipulated the immersion degree of a hybrid audience to investigate the links between presence, agency, and feelings of togetherness in livestreamed concerts. Together, these studies demonstrated the use of virtual embodiment and immersion as methodological tools of XR relevant for scientific music research.

Complementary to the scientific focus of chapter 6, chapter 7 focused on the combined artistic potential of XR and music for novel musical expressions and experiences. It presented six case studies in the extended musical spaces for performance, composition, and narration dedicated to the following research question.

Question chapter 7: what are opportunities of XR technology for artistic musical expression and experience?

These abstracts presented concrete case studies and demonstrated ways to leverage XR's capabilities for distributed, networked performances (Ruan & Xie, 2021; Turchet & Rottondi, 2022), novel virtual affordances (Berthaut, 2020), instruments (Serafin et al., 2016), scenography (Berthaut et al., 2015), and the introduction of novel roles for performer and audience (Catricalà & Eugeni, 2020; Lichty, 2014). A first section focused on the musical space for performance.

Question 7.1: Can we transfer a scientific musical scenario in XR to a public performance context?

Two case studies successfully demonstrated how to take musical scenarios such as the virtual agent from the Piano Phase study (chapter 3) and the virtual drum circle from the polyrhythmic study (chapter 4) out of their controlled lab setting and place them in a public performance context. Placing such experimentally developed models and instruments “into the wild” has the benefit of complementing knowledge gained in sterile and controlled settings with that from rich and noisy ones while welcoming feedback and reflections from (non-expert) participants. XR technology is ideally suited for this task, reflected in its capability to bridge internal and external validity (Kothgassner & Felnhofer, 2020).

Given XR’s capability to offer its users a sense of virtual embodiment and allow them to design their multisensorial environment spatially and corporeally, section 7.3.1 presented a study aimed at answering the following question.

Question 7.2: What are opportunities of XR for spatial audio composition?

Building on scientific insights from the embodied music cognition framework (Leman, 2007) and artistic insights of the meanings hidden in musical material, the “lines and swarms” application designed and developed several novel compositional affordances to draw, enrich, capture, and edit spatial music trajectories in VR. While a few other studies have developed spatial music composition tools in XR (Santini, 2019; Wozniewski et al., 2018), this case study illustrated an approach on how to transfer meaning from artistic intents and metaphors hidden in a musical work into virtual, spatial, and audiovisual affordances. These affordances consequently enabled the design and development of an immersive narration that was experienced by a hybrid online-offline audience (section 7.4.1) and served as a response to the final question tackled in chapter 7.

Question 7.3: What are opportunities of XR for musical narrations?

The case studies presented in section 7.4 integrated several aspects of XR relevant for narration. For instance, XR allows the flexible introduction of non-linear, micro-narratives (Spampinato & Carticalà, 2021) and the design of user roles balancing interactivity of the medium with the control over a storyline (Bruni et al., 2022). The use-cases combined the experiential powers of narration with those of music by immersing its users in spatial soundscapes and having them explore and discover the narration space. The studies played with users’ sense of presence by taking advantage of XR’s capability for out of body illusions (Ehrsson, 2007), multimodal integration (D. Martin et al., 2022), and immersive soundscapes (Reyes-Lecuona et al., 2022).

The use of XR in scientific and artistic music research is now beginning to be widespread (Çamcı & Hamilton, 2020; Turchet et al., 2021). This research is being done across a broad range of disciplines and has allowed for the bridging of social, (neuro)psychological, and philosophical (Cipresso et al., 2018; T. K. Metzinger, 2018; Pan & Hamilton, 2018; Parsons, Gaggioli, et al., 2017) as well as engineering, design, and creative fields (Benford & Giannachi, 2011; Bruni et al., 2022; Cairns et al., 2022; Çamcı & Hamilton, 2020; Loveridge, 2020; Spampinato & Carticalà, 2021; Turchet et al., 2021). Given this amalgamation of insights, theories, and frameworks from such varied sources, it remains a challenge to properly coordinate these efforts and to develop working terminology that allows for mutual exchange and the emergence of synergies. This thesis addressed this challenge by offering a comprehensive methodology and framework to perform fundamental and applied music research in XR together with its demonstration in concrete scientific and artistic experiments. It has shown how XR can be used in experimental paradigms by immersing its users and offering them a virtual, embodied sense of presence. Further, it has shown how XR technology can be integrated in experimental methodologies to capture, mediate, and display fine-grained spatiotemporal coordination patterns and create multimodal and dynamic experimental stimuli. A varied selection of scientific and artistic studies illustrated how XR allowed for the mediation of expressive musical intentions and invocation of new experiences in the extended musical spaces for experimentation, performance, composition, and narration.

8.2 Limitations

This thesis has been devoted to the potential of XR for music research by designing, developing, and experimentally testing the emerging performative, behavioural, cognitive, and experiential dynamics in a selection of extended musical spaces. While the promise of this potential must now be clear, several limitations and identified challenges of this work must be addressed.

A first limitation in this work comes from the many sources of variability present in empirical observations. While chapter 2 argued for the methodological tools from dynamical systems theory to deal with the variability in observed spatiotemporal patterns, other sources of variability remain difficult to address using analytical tools. For instance, the novelty of XR and its immersive, virtually embodied environments must have enhanced variability in participant's responses. In addition, while XR technology, such as HMDs and spatial sound playback, is now available commercially, wearing a HMD is known to induce motion sickness for some and might still represent a relatively invasive setup for participants, let alone when combined with motion capture suits and (neuro)physiological sensors. While the experimental studies presented in this thesis always involved familiarization time with XR environments and stimuli, significant learning effects could still be observed (see chapter 4) and thus remain difficult to causally link to either musical training or increased familiarization. One workaround has been to place users in AR instead of VR, allowing them to remain physically grounded in space. Another has been to couple befriended participants in dyadic experiments, exchanging the variability coming from musically engaging with a virtual stranger for the variability in personal backgrounds. Future work could move from cross-sectional to longitudinal studies, more capable of capturing variability in the experiential evolutions of music making in XR.

A second limitation arises from technical constraints inherent to working with the digitally augmented environments enabled by XR. For instance, when performing music research with XR, one naturally has to deal with aspects of networking, bandwidth, latency, and computing power. Much, if not all, of this work resulted from interdisciplinary approaches that involved engineering and computer science expertise working with state-of-the-art hardware integrated by custom-made software. While time-consuming work, custom software, and expensive equipment might not count as limitations for scientific and artistic outcomes per se, they do make replicability of results more difficult. While XR has been lauded for its promise of replicability that originates from its digital nature, dealing with full-body, multi-user musical interactions in real-time, while safeguarding the capture of fine-grained and synchronous data, remains challenging outside of high-end lab environments (Pan & Hamilton, 2018; Parsons, Gaggioli, et al., 2017).

Finally, while this thesis aimed to offer a comprehensive view on XR's potential for music research by proposing a methodology and its operationalization in scientific and artistic studies, more can be done to assess experimental outcomes. For instance, experimental studies would have profited from larger sample sizes, especially given the variability mentioned before. While the artistic studies have successfully demonstrated the operationalization of experimental developments and insights in public contexts, evaluation of their impact could have inspired improved designs and invoked new scientific reflections in turn. Future work could incorporate semi-structured interviews or questionnaires to gather audience, performer, and composer feedback on communicated intentions and immersion degrees, as well as experiential aspects such as the sense of presence, agency, and feeling of togetherness.

8.3 Extended humanities

Given the technological nature of our personal and social activities today, observations and insights resulting from the extended musical spaces presented in this work might help with the needed "proactive" stance of the humanities (Lesaffre & Leman, 2020) and further highlight the pharmacological nature of the current technoculture (Frankel & Krebs, 2021; Stiegler, 2019). The virtual, with its capacity for self-deployment and awareness, simultaneously offers a route for psychological escape and mood modification as well as contains the dangers of addiction (Nabi & Charlton, 2014). Our online lives, while having the potential to “make us intensely aware of what it is to be human in the physical world” (Lanier, 1988), can simultaneously offer the means for narcissistic self-reflection as well as fracture our scattered selves through forms of “distributed embodiment” (Nabi & Charlton, 2014; Waterworth & Waterworth, 2014). While our presence and engagements in creatively imagined worlds allow us to meaningfully connect to otherwise out of reach virtual others, the multimodal overload of content, choices, and interactions might lead to difficulties in discerning the ontological boundaries between fiction and reality (Bruni et al., 2022). Given the commercial and political forces currently governing a large part of our collective virtual spaces (Ball, 2022; Kreps, 2014), it remains important to remain aware of powers hidden in technological configurations as well as inform and include the general public in their designs.

The argument here is for an “extended humanities”, characterized by critical reflections on the technoculture and the use of XR to create social, participatory, and meaning-laden contexts to share, inspire, and enrich these reflections. Research in such spaces would go beyond a mere “augmentation” of the humanities field with digital technologies and interaction with cultural artefacts, towards an “extension” of humanities problems, questions, and practices that involve the public arena together with the design, engineering, scientific, and artistic fields (Oxman, 2016). The extended humanities would use XR to create contexts such as the extended musical spaces presented in this thesis, to immerse its users in rich and meaning-laden contexts and offer them a 1st person, embodied perspective on their social, (techno)cultural, and in our case musical, engagements. Ideally, these spaces would be persistent, meaning the (inter)actions and creations of its users would continue to exist in their absence and allow for continued critical reflection. While categories of experimentation, performance, composition, and narration were defined in this work, the extended humanities can be expected to work with practices spanning multiple categorical domains and disciplinary boundaries as it would be driven by questions originating out of both scientific and artistic, fundamental and applied, cultural and societal problems.

8.4 Conclusion

With so many of our social interactions co-evolving alongside digital means, and with music composed, performed, and shared in such variety of contexts, one can only begin to imagine the plethora of experiences and creative outbursts awaiting fruitful realisation. XR technology, as only the latest instantiation of our escape into the virtual, has now reached a point from which we are no longer bound to our physical bodies, substituting conscious experience of our physical environments for experimentally controlled, simulated and multisensorial contexts. While this evolution offers great potential for research into the dynamic, coregulatory processes underlying music making, fundamental questions regarding the technologically-mediated nature of musical expression and experience in XR remain.

This thesis addressed these questions by developing a theoretical paradigm and methodological framework which proposed a multi-layered approach to capture, analyse, and interpret performative, embodied, and experiential musical dynamics. By operationalizing this framework in an empirical study, it was able to show how music making between two human-controlled virtual humans did not deteriorate in XR, paving the way for future XR-enabled music research. Further, using a virtual agent or an auditory stimulus in respectively a musical interaction and a breathing exercise as (bio)feedback, this thesis demonstrated XR’s capacity to integrate dynamic models of coordination, effectively able to manipulate behaviour and induce specific experiential states. Other capacities of XR, such as immersion and virtual embodiment, were demonstrated in a selection of scientific and artistic studies aimed at uncovering fundamental principles in sensorimotor coordination, learning, and feelings of togetherness or at revealing the potential for hybrid performances, enriched compositions, and multimodal, interactive narrations.

Given the complexities and promises arising from our hybrid presence and (musical) extension into the virtual, the theory, methodology, scientific and artistic demonstrations of this thesis offered the basis necessary for an improved understanding of, and better dealing with, the combined potential of music making and XR. Such a basis should allow for the emergence of the extended musical space, in which composers, performers, spectators, and participants can explore, discover, and co-create the fertile dynamics and meaningful engagements in the practice of making music.

Appendix A: Schematics of XR implementations

This appendix presents a high-level overview of all XR implementations presented in this thesis. All implementations and their use in the presented studies were done in the ASIL lab except for the “Sonic Breathing” study (chapter 5) and “Searching Borders” case study (section 7.4.2) which were desktop-based applications. The “Music Moves” performance of section 7.2.2 had one performer placed in ASIL, while the other performer was situated 130 meter further in a public theatre hall.

The Art and Science Interaction Lab (ASIL) is a highly flexible and modular research lab, able to create immersive contexts ideally suited for experimental and empirical research. The lab is equipped with state-of-the-art visual, auditory, motion-tracking, and sensor technology, integrated in a high-speed and high-bandwidth networking infrastructure, fully synchronized, and connected to a central backend (for more details, see (Kets et al., 2021). Given ASIL’s central role in the studies presented in this thesis, a list of ASIL’s main components relevant for the XR implementations is presented in Table A-1 below.

Each XR implementation only used a subset of the ASIL infrastructure which can be seen from their high-level schematics presented in Figure A-1 to Figure A-13. Several crucial end-to-end latencies measured in the different implementations can be found in Table A-2.

Visual latencies and latencies to OBS were measured using a 240fps camera, other latencies using audio recordings. Latencies were averaged over 5 measurements with standard deviations or uncertainties due to camera frame rates reported in milliseconds in Table A-2 below. Latencies for the MusicMoves performance consisted of single measurements. Latencies for the “Searching Borders” case study was measured using an audio PC listed in Table A-1, while latency for the “Sonic Breathing” study was measured using a Macbook pro (2018) with I/O vector Size and signal vector size in Max MSP set to 128 and 64, respectively.

Category	Device	Description
networking	networking switch HP5406	16 x 10Gbps and 64 x 1Gbps, all compliant with the CAT6 standard
networking	2 wireless routers Ubiquiti Unify6 U6-Pro	Wi-Fi 6 supported
motion capture	18 Qualisys Oqus 700+ infrared cameras	spread over 2 labs (12 and 6 cameras)
motion capture	5 Qualisys Miqus RGB cameras	spread over 2 labs (4 and 1 cameras)

motion capture	motion capture suits	collection of S, M, L, XL with 42 passive markers each
visuals	Barco F80-4K12 beamer	used for projection on 7x4m acoustically transparent projection screen
visuals	XR HMDs	multiple HTC Vive Pro Eye, Hololens 2, Oculus Quest 2
visuals	USB cameras	Logitech Brio, Logitech HD Pro, Ricoh Theta S 360
audio	Dante audio networking protocol	audio networking in ASIL uses Dante for low-latency audio-over-IP, Rednet PCI Express cards on PCs (default buffer size=128, sample rate=48kHz), default Dante latency = 1ms
audio	acoustically treated lab	reverberation time of 0.5seconds (studio-level)
audio	multiple Focusrite Rednet audio interfaces	audio-over-IP analog-to-digital, digital-to-analog converters
audio	multiple Powersoft Ottocanali 8K4-DSP+D amplifiers	/
audio	72 Martin audio CDD6 speakers with 8 matching mid-bass and sub- woofers	Positioned spatially in a cube missing its bottom square
audio	4 KRK Systems Rokit 4 monitor speakers	to place a sound source at custom locations (eg. under a digital piano)
audio	microphones	2 x Rode NT1-A, 2 x Shure SM58
audio	headphones	multiple closed-back AKG K92
instruments	Yamaha P-60 digital piano keyboard	fully weighted keys, sending out MIDI data
instruments	two custom-made drum pads	A Teensy 3.2 micro-controller is used to convert an analog signal from a pressure sensor into MIDI data (sample rate = 120Hz)

instruments	piano chair	with integrated pressure sensors (see drum pads description)
synchronisation	Rhosendahl Nanosyncs HD clock generator	central unit in the synchronisation infrastructure, creates a SMPTE timestamp that is distributed throughout the lab (over coax, audio, and digitally over OSC)
computers	two audio PCs	CPU: i7-8700K @ 3.7GHz, RAM: 64GB, GPU: NVIDIA GeForce GTX 1080, Storage: 512GB + 1TB SSD 970 PRO, OS: W10e
computers	two VR PCs	CPU: Intel Xeon 6136 @ 3GHz, RAM: 128GB, GPU: NVIDIA GeForce RTX 2080 Ti, Storage: 512GB Samsung 970, OS: W10e
computers	two motion capture PCs	CPU: Intel Xeon 4110 @ 2.1GHz, RAM: 64GB, GPU: NVIDIA GeForce RTX 2070, Storage: 2TB SSD Samsung 970, OS: W10e
computers	one stand-alone PC	CPU: i7-5820K @ 3.3GHz, RAM: 64GB, GPU: NVIDIA GeForce GTX TITAN X, Storage: 1TB SSD Intel 750, OS: W10e

Table A-1: A list of materials comprising ASIL's technical infrastructure

Studies	**Modality**	**Input**	**Output**	**Latency (ms)**
All	(data)	Max for Live	Unity	1.46 ± 0.45
Piano Phase	visual	marker	HTC Vive Pro	42 ± 8
Piano Phase	audio	piano keyboard	Rokit 4 speaker	16 ± 3
Piano Phase	(physical)	pressure sensor	(into Max MSP)	< 5
Piano Phase	(timing)	piano keyboard	(tempo OSC out)	< 5
Drum Circle	visual	marker	Hololens 2 (wireless)	58 ± 4
Drum Circle	audio	drum pad	Rokit 4 speaker	17 ± 2
Drum Circle	(timing)	drum pad	(accuracy OSC out)	< 5
Sonic Breathing	audio	PC keyboard	headphone	42 ± 4
Sonic Breathing	(timing)	PC keyboard	(Kuramoto out)	< 5
Tapswap	visual	Logitech Brio	HTC Vive Pro	96 ± 16

Tapswap	audio	drum pad	headphone	17 ± 2
Virtual Partners	visual	marker	HTC Vive Pro	42 ± 8
Virtual Partners	audio	HTC Vive Pro	(into Ableton)	430 ± 6
Livestreams	visual	Logitech HD Pro	OBS	175 ± 4
Livestreams	audio	Rode NT1-1A	OBS	255 ± 4
Piano Phase performance	visual	marker	HTC Vive Pro (wireless)	54 ± 8
Piano Phase performance	audio	piano keyboard	Rokit 4 speakers	16 ± 3
Music Moves	visual	marker	Hololens 2 (wireless)	120
Music Moves	visual	marker	Unity	80
Music Moves	visual	marker	Zoom (3rd person)	180
Music Moves	visual	marker	Zoom (1st person)	205
Music Moves	audio	drum pad	Rokit 4 speaker	20
Lines and Swarms	audio	Quest controller	ASIL speakers	122 ± 6
Searching Borders	visual	Logitech Brio	Quest 2 (wireless)	93 ± 18
Searching Borders	visual	Rode NT1-1A	Quest 2 (wireless)	200 ± 16

Table A-2: Overview of end-to-end latencies in the various XR implementations

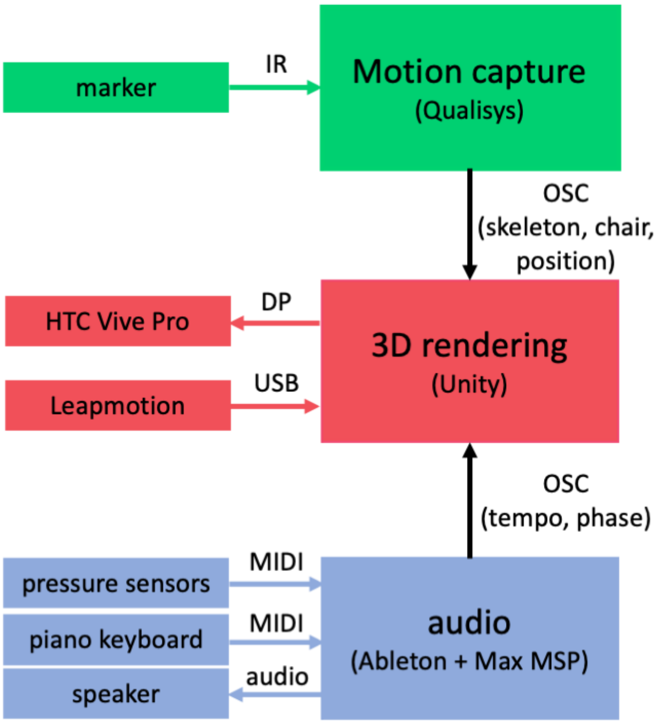


Figure A-1: Schematic for the XR implementation in the "Piano phase" study of chapter 3 (green = motion capture, red = visual, blue = audio)

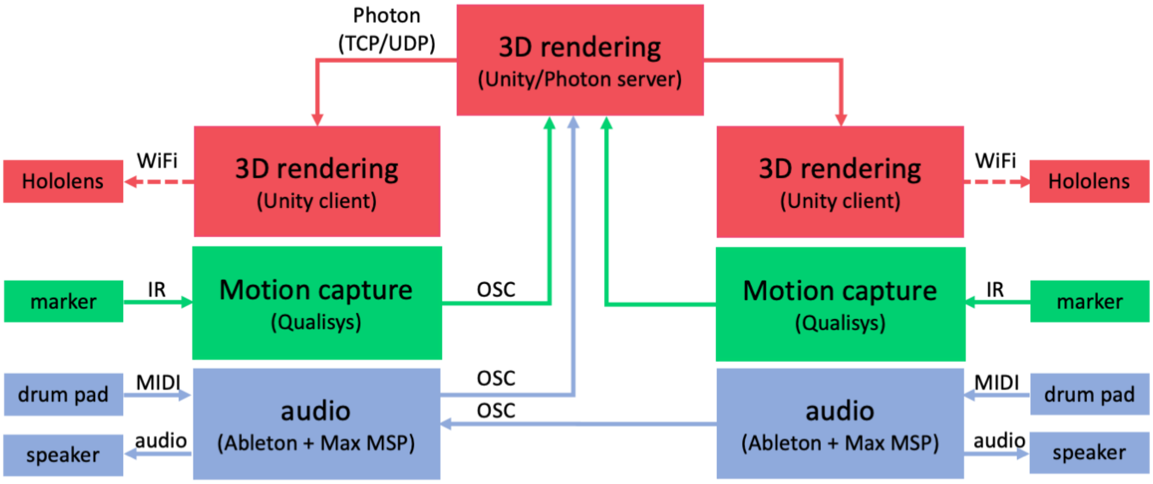


Figure A-2: Schematic for the XR implementation in the "Drum circle" study of chapter 4 (green = motion capture, red = visual, blue = audio)

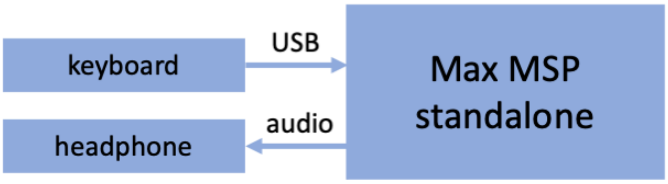


Figure A-3: Schematic for the XR implementation in the breathing biofeedback study of chapter 5 (green = motion capture, red = visual, blue = audio)

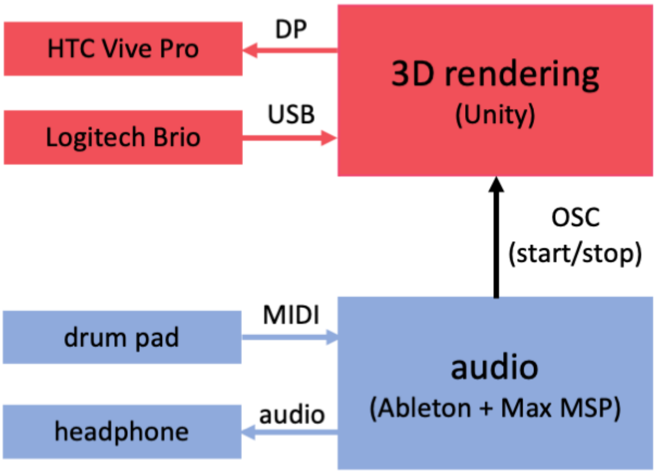


Figure A-4: Schematic for the XR implementation in the "Tapswap" study of section 6.2 (EEG setup not shown, green = motion capture, red = visual, blue = audio)

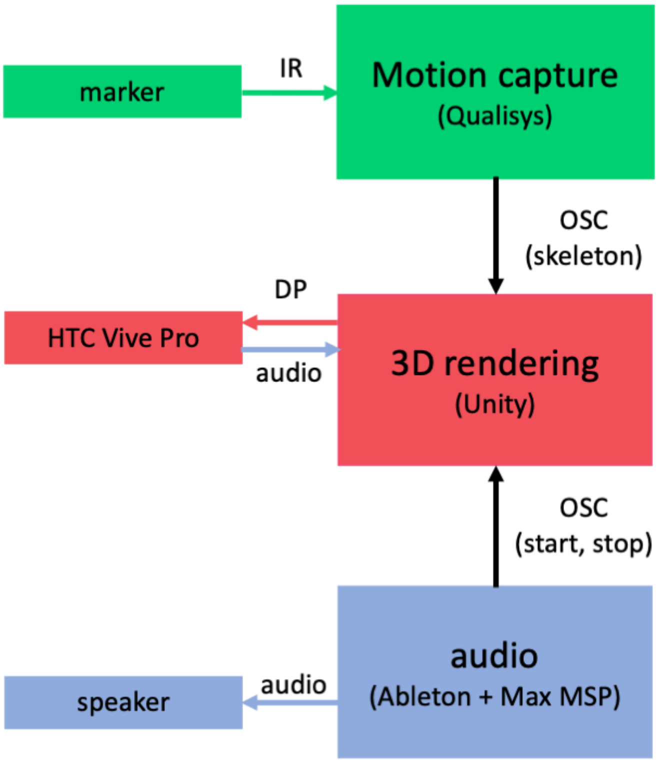


Figure A-5: Schematic for the XR implementation in the "Virtual partners" study of section 6.3 (EEG setup not shown, green = motion capture, red = visual, blue = audio)

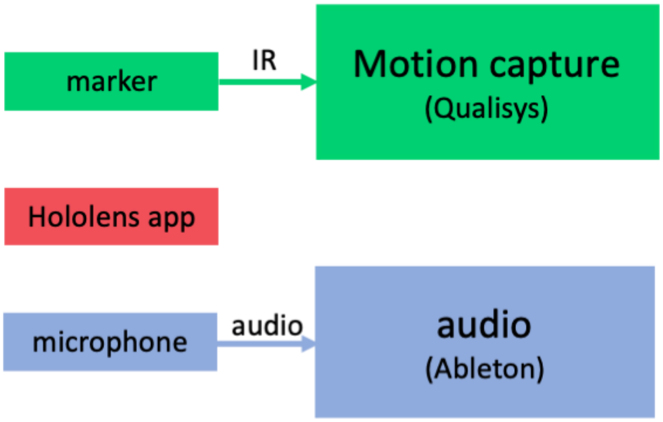


Figure A-6: Schematic for the XR implementation in the "Virtual teacher" study of section 6.4 (green = motion capture, red = visual, blue = audio)

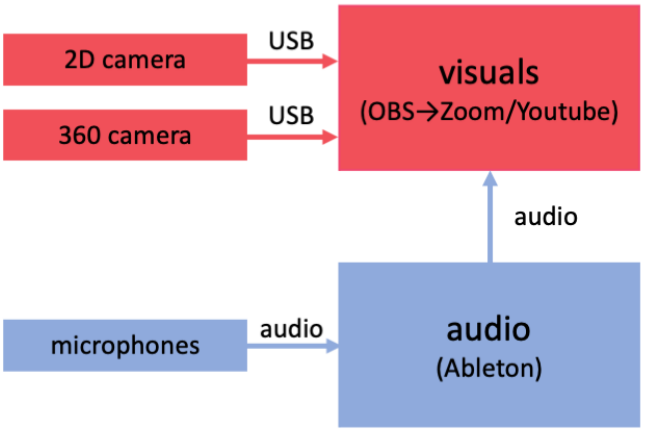


Figure A-7: Schematic for the XR implementation in the "Livestreams" study of section 6.5 (for an overview of cameras and microphones, see (Onderdijk et al., 2021). *Audio and video streams were synchronized using OBS, green = motion capture, red = visual, blue = audio)*

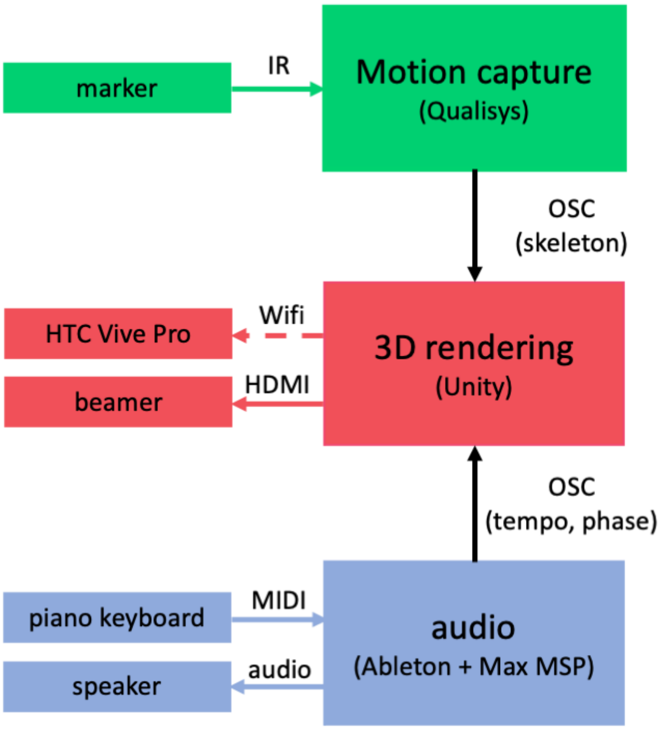


Figure A-8: Schematic for the XR implementation in the "Piano phase" performance of section 7.2.1 (identical setup as in the "Piano phase" study except for a wireless XR HMD, beamer projection, and only one pianist, green = motion capture, red = visual, blue = audio)

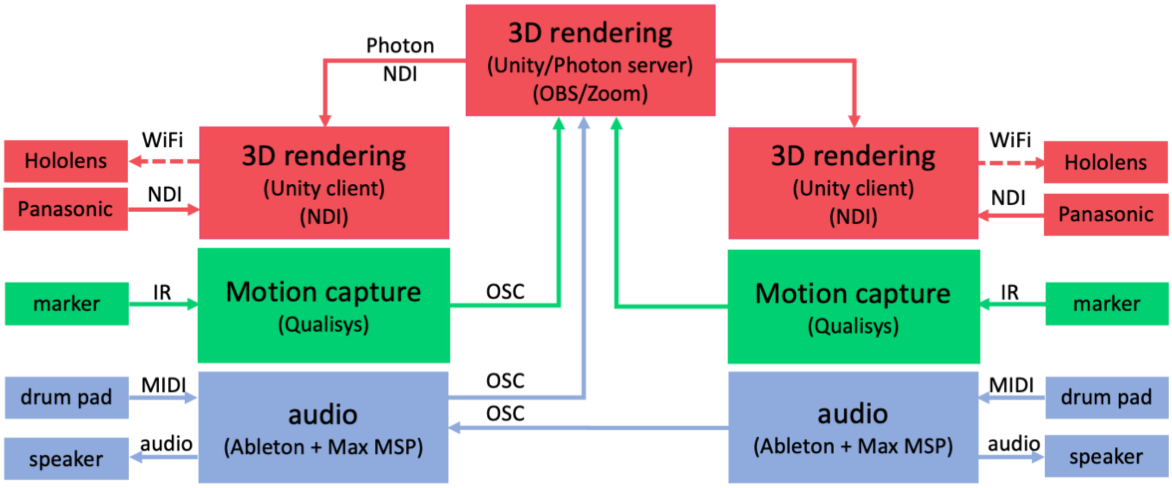


Figure A-9: Schematic for the XR implementation in the "Music moves" performance of section 7.2.2 (identical setup as in the "Drum circle" study, with an added OBS and Zoom instance, two Panasonic N-series PTZ cameras with NDI support, a 10Gbps networking switch and fiber cable, and a Soundcraft SI Compact Dante-compliant mixing console, green = motion capture, red = visual, blue = audio)

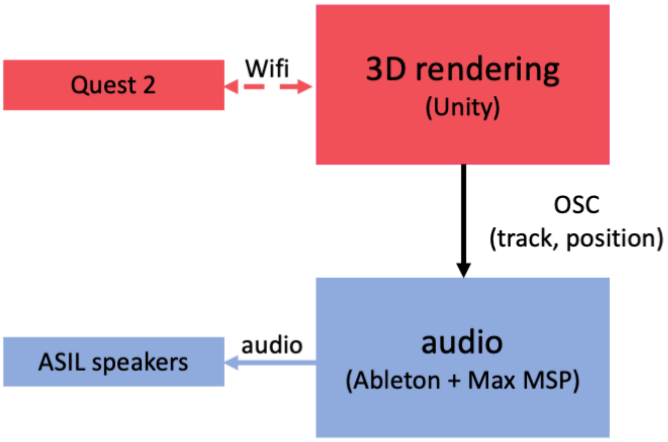


Figure A-10: Schematic for the XR implementation in the "Lines and swarms" narration of section 7.3.1 (green = motion capture, red = visual, blue = audio)

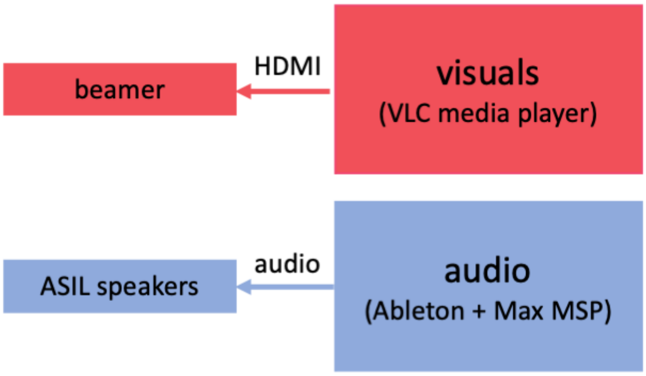


Figure A-11: Schematic for the XR implementation in the "Being hungry" narration of section 7.4.1 (green = motion capture, red = visual, blue = audio)

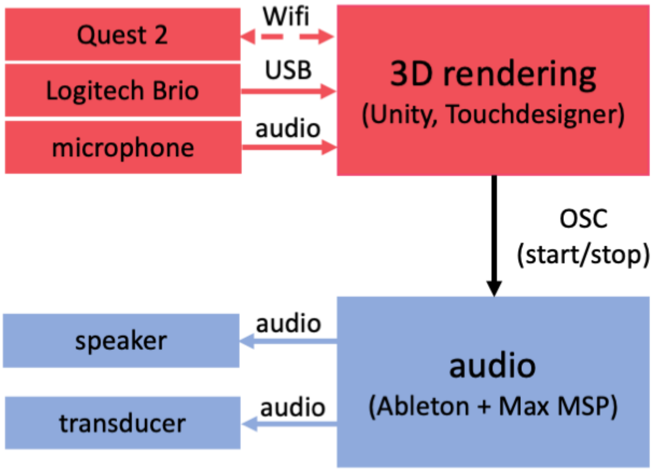


Figure A-12: Schematic for the XR implementation in the "Searching borders" narration of section 7.4.2 (green = motion capture, red = visual, blue = audio)

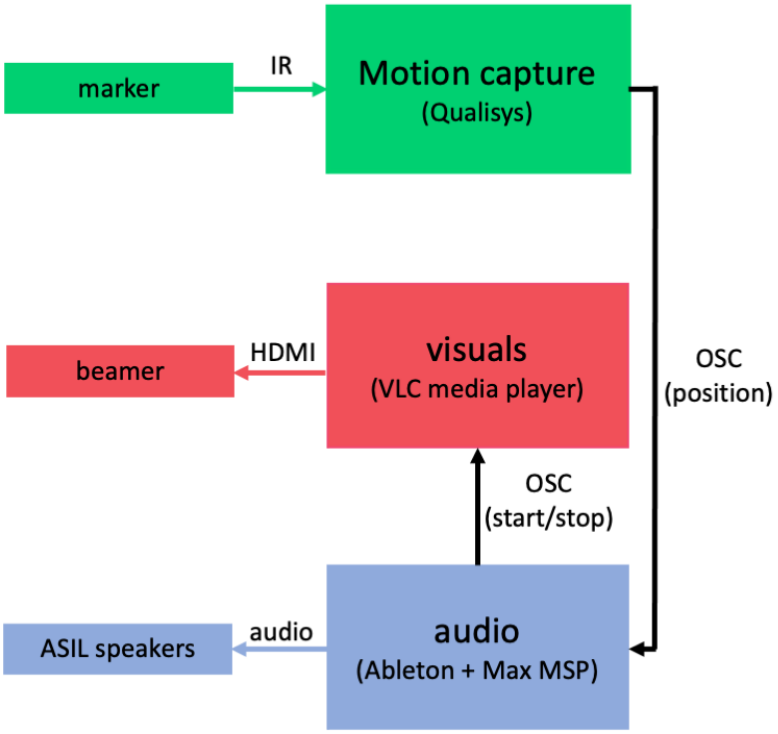


Figure A-13: Schematic for the XR implementation in the "Be hear now" narration of section 7.4.3 (green = motion capture, red = visual, blue = audio)

Appendix B: Supplementary material of chapter 3

Post-experimental questionnaire

1. How many years of ensemble piano playing do you have?
    a. *18 years of professional ensemble playing.*

2. Did you already have experience in performing the piece "Piano Phase" before the experiment? If not, how did you practice the piece before the experiment?
    a. *Alone and with the robotic recordings forwarded before the experiment.*

3. If you used the audio recording for your practicing, did you find this technology useful for the result of your performance during the experiment?
    a. *Working with a recorded performance is always helpful and can make one aware of some of the difficulties that ensemble playing will pose. However, it is by no means capable of giving you the full understanding of the challenges that an encounter with a real player will bring. Practicing too much with the same recording could also lead to learning too well the patterns of response to that particular recording. This might render the encounter with another player more difficult in the end.*

4. Describe how your experience was (feelings, reflections, motivation, performative strategies etc.) in playing with the real partner in VR in terms of interaction and rewards (First Condition).
    a. *Being the first time in a VR environment had little impact on my concentration. Although I was fascinated in the beginning, this did not prevent me to give full attention to the task at hand. The VR environment added more of a novelty element but did not bring a fundamental difference from any other performative experience.*

5. Describe how your experience was (feelings, reflections, motivation, performative strategies etc.) with the real partner animated in VR in terms of interaction and rewards (Second Condition).
    a. *As in the first case, the VR environment added solely an interesting, fun element, that was almost discarded when the actual playing took place. Because I had a good interaction with my partner, the feeling was very close to the one of a performance that happens in real conditions. The fact that the other player was responsive to me was enough to convince me that the situation was real and made me enjoy it thoroughly. Seeing a virtual representation of my partner made the whole situation seem more fun, albeit just as serious as any other performance. I had sometimes the feeling that I was in an amusement park. The fact that the second condition had more realistic visual elements than the first had little influence on my general perception and enjoyment of the performance.*

The enhancing element was that I was playing it for the second time, therefore getting more familiar with the piece, and more enthusiastic because of the efficient and pleasant development of the collaboration with the other pianist.

6. Describe how your experience was (feelings, reflections, motivation, performative strategies etc.) in playing with the virtual partner animated in VR in terms of interaction and rewards (Third Condition).
    a. *The third instance was very demanding. Some parts of it I have found quite frustrating, to the point where I would feel physically and psychologically uncomfortable. It took me about one to two minutes to realize that my virtual partner was, in fact, not present and that I was playing with a recording. My main focus was on trying to understand the intentions of something that was quite obviously not going to follow mine. To my ears at the time, the performance was disastrous. To the virtual pianist in front of me I have referred in my mind, throughout the entire performance, as 'that thing'. I had to make sustained efforts to maintain focus and complete the task. I was relieved when it was over. It is interesting to note that, as opposed to the second condition, where the fact that I felt the presence of a real person made me connect immediately to an image that was obviously not real, in the third condition I felt almost repulsed by the visual element. The fact that I was immersed in a VR environment made the whole experience feel quite grotesque.*

Appendix C: Supplementary material of chapter 4

Post-experimental questionnaire

1. Please indicate the picture below which best describes your relationship with your partner during the musical interaction:

1	2	3	4	5	6	7
O	O	O	O	O	O	O

2. How would you score
   (1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = very good, 6 = excellent, 7 = exceptional)

	1	2	3	4	5	6	7
The musical performance	O	O	O	O	O	O	O
The contribution of your musical partner	O	O	O	O	O	O	O
The contribution of yourself	O	O	O	O	O	O	O

3. To what extent did you feel “agency” over the musical performance?
   (Feelings of agency are described as having a subjective feeling of control. In other words, agency is feeling a general sense of control over what you are doing or over the situation you are in.)

1.	2.	3.	4.	5.	6.	7.
Not at all			Partly			Very much
O	O	O	O	O	O	O

4. To what extent did you feel “shared agency” over the musical performance?
   (When two of more people partake in something together, they can feel a sense of shared agency over what they are doing or the situation they are in. When you feel this shared agency, you feel as if you are going something together (your experience is the product of something done together). When you do not feel shared agency, you feel as if others did not have any control over your experience, and that you acted completely independently.)

1.	2.	3.	4.	5.	6.	7.
Not at all			Partly			Very much
O	O	O	O	O	O	O

5. Place yourself back in the activity and indicate the most appropriate category

	1. Not at all	2.	3.	4. Partly	5.	6.	7. Very much
I feel just the right amount of challenge	○	○	○	○	○	○	○
I don’t notice the time passing	○	○	○	○	○	○	○
I am totally absorbed in what I am doing	○	○	○	○	○	○	○
I am completely lost in thought	○	○	○	○	○	○	○

Data analysis

Parameter	**Value**
Metervalue	3/2, 2, 3, 4, 6, 8, 9
Tgvalue	1000 * 60/170
Gravity	0.01
Outscope	0.05
System noise	1e-05
Observation noise	1e-03
Sampling rate	100
Number of cycles	10
Number of blocks	4

Table C-1: Blistener parameters

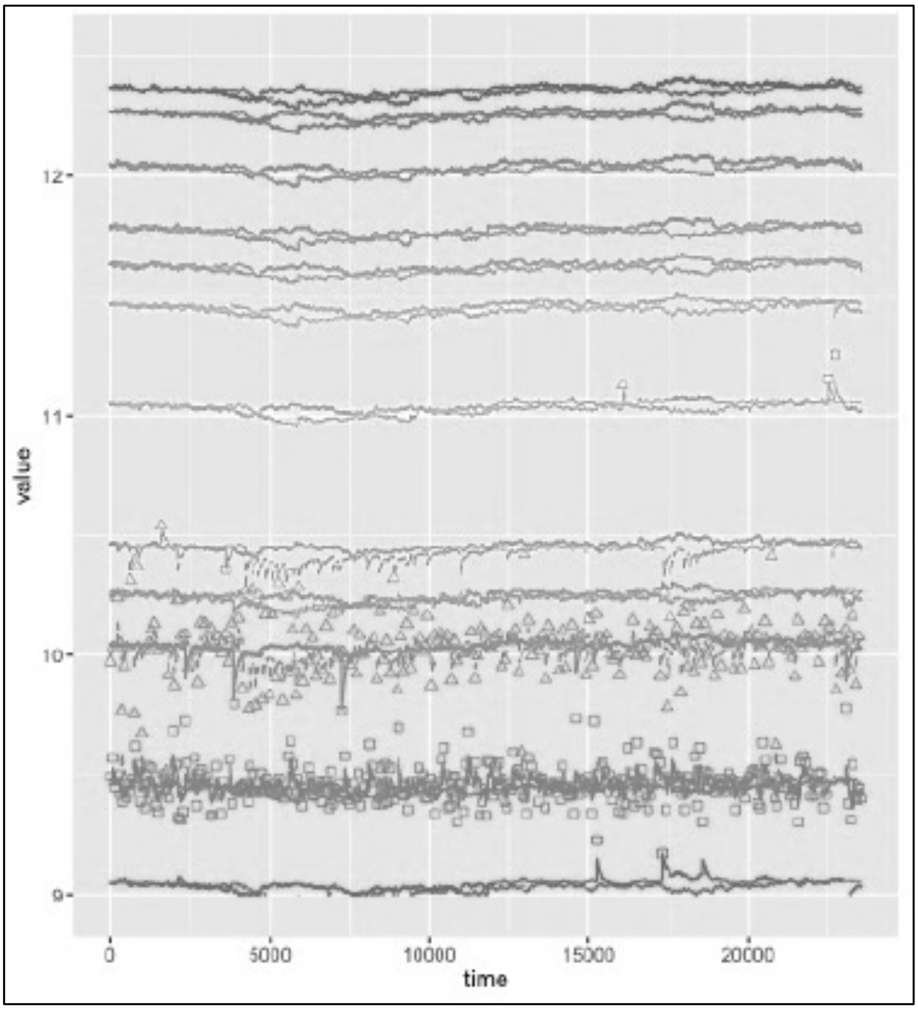


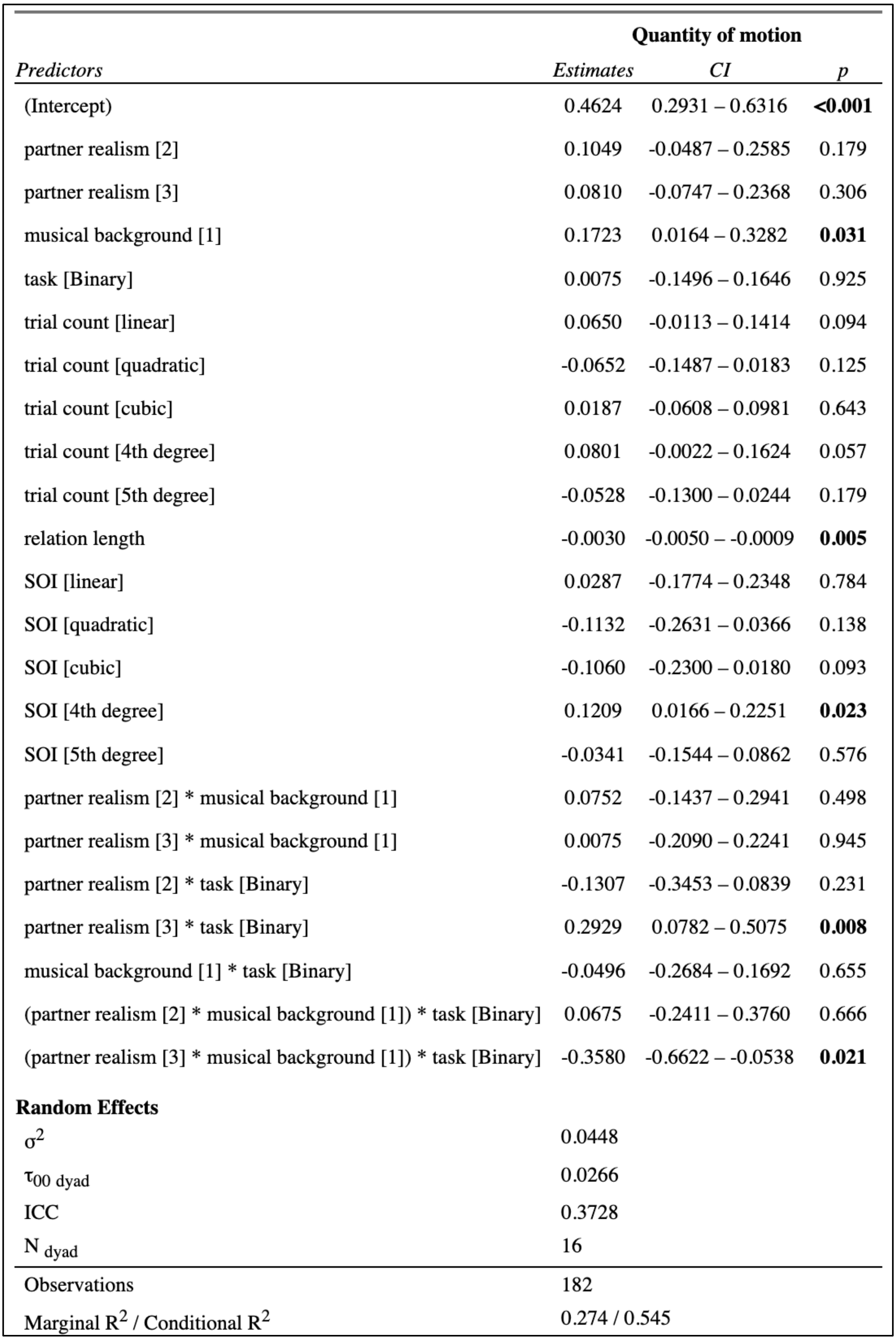
Figure C-1: IOI-predictions of 2 participants using 12 trackers per participant

	Prediction error (individual performance)		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.0194	0.0132 – 0.0284	**<0.001**
partner realism [2]	1.0550	0.8755 – 1.2713	0.572
partner realism [3]	1.1056	0.9091 – 1.3445	0.312
musical background [1]	0.8717	0.7493 – 1.0140	0.075
task [Binary]	1.7845	1.4334 – 2.2217	**<0.001**
trial count [linear]	0.7990	0.7000 – 0.9122	**0.001**
trial count [quadratic]	0.9928	0.8577 – 1.1492	0.922
trial count [cubic]	0.9478	0.8315 – 1.0805	0.420
trial count [4th degree]	1.0316	0.8969 – 1.1865	0.661
trial count [5th degree]	1.2265	1.0755 – 1.3987	**0.003**
training	0.9408	0.8853 – 0.9999	**0.050**
relation frequency [linear]	1.9290	1.3620 – 2.7321	**<0.001**
relation frequency [quadratic]	1.3547	1.0919 – 1.6808	**0.006**
relation frequency [cubic]	1.0412	0.8185 – 1.3244	0.741
relation frequency [4th degree]	0.8340	0.6576 – 1.0577	0.133
partner realism [2] * task [Binary]	0.8421	0.6485 – 1.0935	0.196
partner realism [3] * task [Binary]	0.9179	0.7058 – 1.1937	0.520
musical background [1] * task [Binary]	0.8316	0.6721 – 1.0290	0.089
Random Effects			
σ^2	0.1234		
$\tau_{00\ \text{dyad}}$	0.0614		
ICC	0.3323		
N $_{\text{dyad}}$	16		
Observations	171		
Marginal R^2 / Conditional R^2	0.442 / 0.628		

Table C-2: Overview of the LME model of the prediction error for the individual performances between players and (binary and ternary) task (estimates, confidence intervals, p-values, random effects, and residual variances)

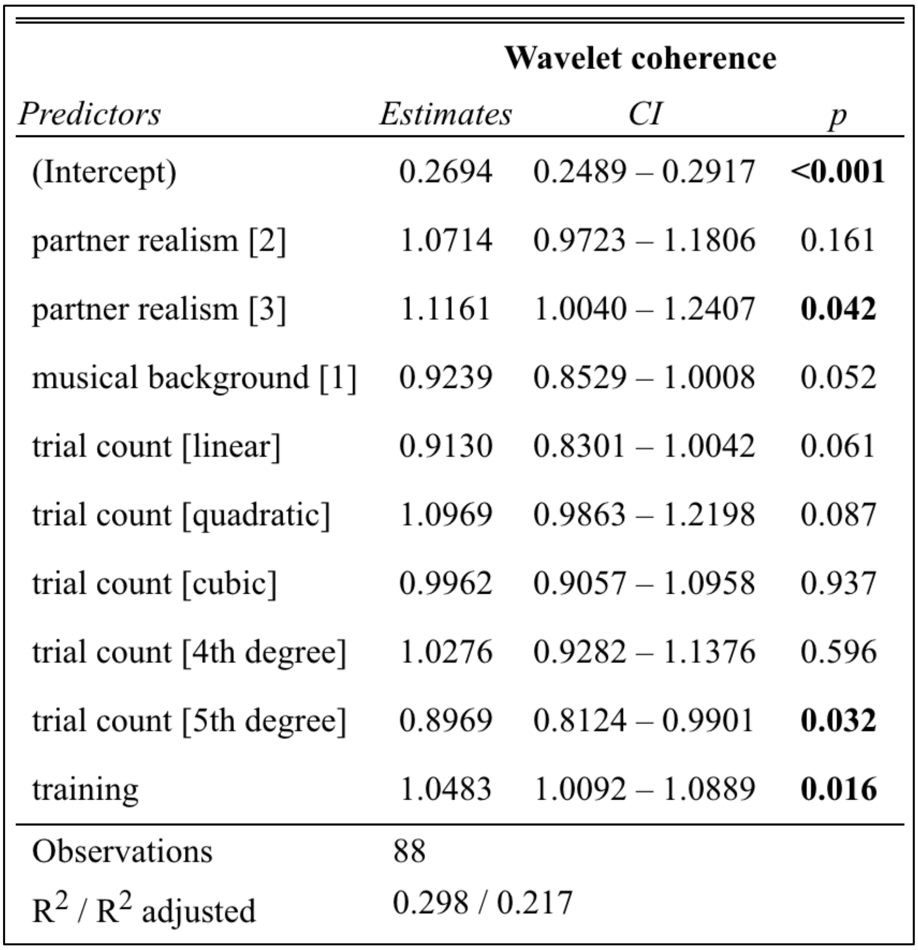
	Prediction error (joint performance)		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.0228	0.0176 – 0.0294	**<0.001**
partner realism [2]	1.0366	0.8387 – 1.2813	0.736
partner realism [3]	1.2673	0.9998 – 1.6064	0.050
musical background [1]	0.8007	0.6739 – 0.9513	**0.012**
trial count [linear]	0.7616	0.6168 – 0.9404	**0.012**
trial count [quadratic]	0.9414	0.7463 – 1.1876	0.606
trial count [cubic]	1.1065	0.8986 – 1.3626	0.336
trial count [4th degree]	0.9679	0.7724 – 1.2128	0.774
trial count [5th degree]	1.3482	1.0961 – 1.6583	**0.005**
SOI	0.8374	0.6690 – 1.0481	0.119
engagement	0.7686	0.6124 – 0.9645	**0.024**
training	0.8089	0.6518 – 1.0039	0.054
Random Effects			
σ^2	0.1578		
$\tau_{00\ \text{dyad}}$	0.1398		
ICC	0.4696		
N_{dyad}	16		
Observations	88		
Marginal R^2 / Conditional R^2	0.375 / 0.669		

Table C-3: Overview of the LME model of the prediction error for the joint performance between players (estimates, confidence intervals, p-values, random effects, and residual variances)



	Quantity of motion		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.4624	0.2931 – 0.6316	**<0.001**
partner realism [2]	0.1049	-0.0487 – 0.2585	0.179
partner realism [3]	0.0810	-0.0747 – 0.2368	0.306
musical background [1]	0.1723	0.0164 – 0.3282	**0.031**
task [Binary]	0.0075	-0.1496 – 0.1646	0.925
trial count [linear]	0.0650	-0.0113 – 0.1414	0.094
trial count [quadratic]	-0.0652	-0.1487 – 0.0183	0.125
trial count [cubic]	0.0187	-0.0608 – 0.0981	0.643
trial count [4th degree]	0.0801	-0.0022 – 0.1624	0.057
trial count [5th degree]	-0.0528	-0.1300 – 0.0244	0.179
relation length	-0.0030	-0.0050 – -0.0009	**0.005**
SOI [linear]	0.0287	-0.1774 – 0.2348	0.784
SOI [quadratic]	-0.1132	-0.2631 – 0.0366	0.138
SOI [cubic]	-0.1060	-0.2300 – 0.0180	0.093
SOI [4th degree]	0.1209	0.0166 – 0.2251	**0.023**
SOI [5th degree]	-0.0341	-0.1544 – 0.0862	0.576
partner realism [2] * musical background [1]	0.0752	-0.1437 – 0.2941	0.498
partner realism [3] * musical background [1]	0.0075	-0.2090 – 0.2241	0.945
partner realism [2] * task [Binary]	-0.1307	-0.3453 – 0.0839	0.231
partner realism [3] * task [Binary]	0.2929	0.0782 – 0.5075	**0.008**
musical background [1] * task [Binary]	-0.0496	-0.2684 – 0.1692	0.655
(partner realism [2] * musical background [1]) * task [Binary]	0.0675	-0.2411 – 0.3760	0.666
(partner realism [3] * musical background [1]) * task [Binary]	-0.3580	-0.6622 – -0.0538	**0.021**
Random Effects			
σ^2	0.0448		
$\tau_{00\ \text{dyad}}$	0.0266		
ICC	0.3728		
N_{dyad}	16		
Observations	182		
Marginal R^2 / Conditional R^2	0.274 / 0.545		

Table C-4: Overview of the LME model of the quantity of motion of the ternary and binary task (estimates, confidence intervals, p-values, random effects, and residual variances)



	Wavelet coherence		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.2694	0.2489 – 0.2917	**<0.001**
partner realism [2]	1.0714	0.9723 – 1.1806	0.161
partner realism [3]	1.1161	1.0040 – 1.2407	**0.042**
musical background [1]	0.9239	0.8529 – 1.0008	0.052
trial count [linear]	0.9130	0.8301 – 1.0042	0.061
trial count [quadratic]	1.0969	0.9863 – 1.2198	0.087
trial count [cubic]	0.9962	0.9057 – 1.0958	0.937
trial count [4th degree]	1.0276	0.9282 – 1.1376	0.596
trial count [5th degree]	0.8969	0.8124 – 0.9901	**0.032**
training	1.0483	1.0092 – 1.0889	**0.016**
Observations	88		
R^2 / R^2 adjusted	0.298 / 0.217		

Table C-5: Overview of the LM model of the wavelet coherence of the ternary and binary task (estimates, confidence intervals, p-values, random effects, and residual variances)

		Agency	
Predictors	*Odds Ratios*	*CI*	*p*
1|2	0.0035	0.0004 – 0.0292	**<0.001**
2|3	0.0295	0.0101 – 0.0865	**<0.001**
3|4	0.1377	0.0586 – 0.3236	**<0.001**
4|5	0.4910	0.2227 – 1.0825	0.078
5|6	3.4340	1.5267 – 7.7239	**0.003**
6|7	64.6045	22.3641 – 186.6272	**<0.001**
partner realism [2]	1.2060	0.6060 – 2.4001	0.594
partner realism [3]	1.8815	0.8788 – 4.0283	0.104
musical background [1]	2.7497	1.5369 – 4.9198	**0.001**
trial count [linear]	5.8608	2.8441 – 12.0773	**<0.001**
trial count [quadratic]	0.5796	0.2730 – 1.2305	0.156
trial count [cubic]	0.9719	0.4957 – 1.9054	0.934
trial count [4th degree]	1.1194	0.5411 – 2.3159	0.761
trial count [5th degree]	0.3373	0.1675 – 0.6795	**0.002**
Random Effects			
σ^2	3.2899		
$\tau_{00\ \text{individual}}$	1.8782		
ICC	0.3634		
$N_{\text{individual}}$	32		
Observations	182		
Marginal R^2 / Conditional R^2	0.165 / 0.469		

Table C-6: Overview of the CLMM model of the agency scores of the ternary and binary task (estimates, confidence intervals, p-values, random effects, and residual variances)

	Shared agency		
Predictors	*Odds Ratios*	*CI*	*p*
1|2	0.0069	0.0016 – 0.0292	**<0.001**
2|3	0.0299	0.0100 – 0.0894	**<0.001**
3|4	0.0630	0.0229 – 0.1729	**<0.001**
4|5	0.2094	0.0820 – 0.5350	**0.001**
5|6	0.9872	0.3982 – 2.4476	0.978
6|7	12.2679	4.4862 – 33.5476	**<0.001**
partner realism [2]	3.1164	1.5502 – 6.2651	**0.001**
partner realism [3]	4.0626	1.8951 – 8.7092	**<0.001**
trial count [linear]	4.8265	2.3717 – 9.8219	**<0.001**
trial count [quadratic]	0.6967	0.3314 – 1.4648	0.340
trial count [cubic]	0.7204	0.3728 – 1.3923	0.329
trial count [4th degree]	1.4078	0.6902 – 2.8713	0.347
trial count [5th degree]	0.4793	0.2437 – 0.9429	**0.033**
percussion [linear]	2.7711	1.2647 – 6.0716	**0.011**
relation length	0.9856	0.9723 – 0.9991	**0.036**
Random Effects			
σ^2	3.2899		
$\tau_{00\ \text{individual}}$	0.8593		
ICC	0.2071		
$N_{\text{individual}}$	32		
Observations	182		
Marginal R^2 / Conditional R^2	0.228 / 0.388		

Table C-7: Overview of the CLMM model of the shared agency scores of the ternary and binary task (estimates, confidence intervals, p-values, random effects, and residual variances)

	SOI		
Predictors	*Odds Ratios*	*CI*	*p*
1|2	0.0309	0.0039 – 0.2437	**0.001**
2|3	0.6623	0.1054 – 4.1623	0.660
3|4	2.6565	0.4208 – 16.7712	0.299
4|5	6.5190	1.0097 – 42.0894	**0.049**
5|6	31.7711	4.6055 – 219.1743	**<0.001**
6|7	417.1060	52.1456 – 3336.3761	**<0.001**
partner realism [2]	2.3158	1.1511 – 4.6591	**0.019**
partner realism [3]	6.4856	2.8325 – 14.8501	**<0.001**
trial count [linear]	10.4278	4.8326 – 22.5010	**<0.001**
trial count [quadratic]	0.5448	0.2548 – 1.1650	0.117
trial count [cubic]	0.6612	0.3342 – 1.3082	0.235
trial count [4th degree]	2.7126	1.2747 – 5.7725	**0.010**
trial count [5th degree]	0.4761	0.2410 – 0.9408	**0.033**
training	1.5310	1.0229 – 2.2913	**0.038**
Random Effects			
σ^2	3.2899		
τ_{00} individual	5.1488		
ICC	0.6101		
N individual	32		
Observations	182		
Marginal R^2 / Conditional R^2	0.215 / 0.694		

Table C-8: Overview of the CLMM model of the self-other integration scores of the ternary and binary task (estimates, confidence intervals, p-values, random effects, and residual variances)

	Flow (absorption)		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.6295	0.5782 – 0.6808	**<0.001**
musical background [1]	0.0351	0.0113 – 0.0590	**0.004**
SOI y [linear]	0.0647	-0.0264 – 0.1559	0.163
SOI y [quadratic]	0.1088	0.0333 – 0.1843	**0.005**
SOI y [cubic]	0.0765	0.0146 – 0.1384	**0.016**
SOI y [4th degree]	0.0147	-0.0476 – 0.0770	0.642
SOI y [5th degree]	0.0421	-0.0201 – 0.1044	0.183
relation frequency [linear]	-0.0110	-0.0994 – 0.0774	0.806
relation frequency [quadratic]	-0.0713	-0.1458 – 0.0033	0.061
relation frequency [cubic]	0.0253	-0.0524 – 0.1030	0.521
relation frequency [4th degree]	0.0457	-0.0289 – 0.1204	0.228
relation length	0.0012	0.0002 – 0.0022	**0.015**
Random Effects			
σ^2	0.0064		
$\tau_{00\ \text{individual}}$	0.0034		
ICC	0.3476		
$N_{\text{individual}}$	32		
Observations	177		
Marginal R^2 / Conditional R^2	0.297 / 0.541		

Table C-9: Overview of the CLMM model of the flow (absorption) scores of the ternary and binary task (estimates, confidence intervals, p-values, random effects, and residual variances)

Appendix D: List of artistic case studies and associated videos

Videos of the case studies listed below can be found here: https://youtube.com/playlist?list=PL_9jB_2tw40xQW2VQt-juZlkW4cBQIJPs

Case-study 1, section 7.2.1: Van Kerrebroeck, B., Caruso G., & Maes, P. J. (2020). Piano Phase with virtual pianist. De Krook, Ghent, Belgium. *Public performance of Piano Phase (Steve Reich, 1967) for pianist with virtual partner*

- Title: "Piano Phase with virtual pianist"
- Description: Technical setup and excerpt from a public performance of Piano Phase (Steve Reich, 1967) with virtual algorithmically controlled pianist at De Krook, Ghent, Belgium.
- Uploaded: 27th of February, 2021
- Link: https://youtu.be/GlVaMPCotzM

Case-study 2, section 7.2.2: Van Kerrebroeck, B., Van Kets, N., Moens, B., Van Wallendael, G., & Maes, P. J. (2022). MusicMoves. Onze (on)bekabelde cultuur symposium, Ghent, Belgium, *Hybrid (online-offline) XR music performance*

- Title:
  - "MRMoves: artistic case-study, public performance (online audience perspective)"
  - "MusicMoves: artistic case-study, public performance (offline audience perspective)"
  - "MusicMoves: additional features"
- Description: Three videos demonstrating a public performance of the MusicMoves platform.
- Uploaded: 21st of September, 2022
- Link:
  - https://youtu.be/gL_PbnShF9k
  - https://youtu.be/j9ZE2cwLyvA
  - https://youtu.be/7NYdFb64IKE

Case-study 3, section 7.3.1: Van Kerrebroeck, B., & Maes, P. J. (2021). Lines And Swarms. ASIL, Ghent, Belgium. A s*patial audio composition tool in virtual reality*

- Title: "LinesAndSwarms: Spatialization of audio trajectories in virtual reality"
- Description: VR audio spatialisation tool walkthrough. A user can draw, augment, record and edit visual trajectories in a virtual space that are rendered as spatial audio trajectories using Ambisonics.
- Uploaded: 20th of August, 2021
- Link: https://youtu.be/YacYzAUl1M0

Case-study 3, section 7.4.1: Van Kerrebroeck, B., Hermans, C., & Maes, P. J., (2021). Being Hungry. ASIL lab of IPEM, Ghent, Belgium and online in Mozilla Hubs. *Spatial audio composition presented in IPEM's Art-Science-Interaction Lab and on the Mozilla Hubs platform*

- Title:
    - "Being Hungry (Mozilla Hubs video)"
    - "BeingHungry (Binaural audio)"
- Description: Two videos demonstrating the online and interactive version or binaural audio and video of the immersive interpretation of the "Being Hungry" composition by Duncan Speakman and Tineke De Meyer created using the "Lines and Swarms" VR application.
- Uploaded: 25th of July, 2021
- Link: https://youtu.be/ZGhWqs9JQho

Case-study 4, section 7.4.2: Van Kerrebroeck, B., De Mulder, S., & Lareu, B. (2021). Searching Borders. Best submission HacXathon 2021 during 24 Hours of Deep Listening at Tinnepot theatre, Ghent, Belgium. *Immersive virtual experience*

- Title: "SearchingBorders"
- Description: A walkthrough of the XR submission for the Hackathon "24 Hours of Deep Listening - HacXathon 2021" organised by Consouling records and XRT Gent.
- Uploaded: 21st of September, 2022
- Link: https://youtu.be/unazOep9p7U

Case-study 5, section 7.4.3: Carmen, L., Van Kerrebroeck, B., & Maes, P. J. (2022). Be Hear Now. ASIL lab of IPEM, Ghent, Belgium. *Immersive, interactive audiovisual installation at IPEM's Art-Science-Interaction Lab*

- Title: "Be hear Now trailer"
- Description: A trailer of "Be Hear Now", an immersive and interactive audiovisual installation created during a RITCS internship at IPEM in collaboration with LUCA School of Arts.
- Uploaded: 21st of September, 2022
- Link: https://youtu.be/dCsWT55DPaM

Appendix E: List of acronyms

ASIL = Art and science interaction lab
AR = Augmented reality
CI = Confidence interval
CLM = Cumulative link model
DP = DisplayPort cable
HMD = Head-mounted display
IOI = Inter-onset-interval
IPEM = Instituut voor psychoakoestiek en elektronische muziek
IR = Infrared
LME = Linear mixed-effect
MIDI = Musical instrument digital interface
MR = Mixed reality
OSC = Open sound control
QoM = Quantity of motion
SMPTE = Society of motion picture and television engineers
SOI = Self-other integration
UI = User interface
VR = Virtual reality
XR = Extended reality

Appendix F: List of figures

Appendix G: List of Tables

Appendix H: References

Acebrón, J. A., Bonilla, L. L., Vicente, C. J. P., Ritort, F., & Spigler, R. (2005). The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, *77*(1), 137–185.

Adhana, R., Gupta, R., Dvivedi, J., Ahmad, S., Dvivedii, J., & Ahmad, S. (2013). The influence of the 2: 1 yogic breathing technique on essential hypertension. *Indian Journal of Physiology and Pharmacology*, *57*(1), 38–44.

Agawu, K. (2006). Structural analysis or cultural analysis? Competing perspectives on the "standard pattern" of West African rhythm. *Journal of the American Musicological Society*, *59*(1), 1–46.

Agrawal, S., Simon, A., Bech, S., BÆrentsen, K., & Forchhammer, Sø. (2020). Defining immersion: Literature review and implications for research on audiovisual experiences. *Journal of the Audio Engineering Society*, *68*(6), 404–417.

Alcañiz, M., Bigné, E., & Guixeres, J. (2019). Virtual reality in marketing: A framework, review, and research agenda. *Frontiers in Psychology*, *10*.

Aldausari, N., Sowmya, A., Marcus, N., & Mohammadi, G. (2023). Video Generative Adversarial Networks: A Review. *ACM Computing Surveys*, *55*(2), 1–25.

Aletta, F., Oberman, T., & Kang, J. (2018). Associations between positive health-related effects and soundscapes perceptual constructs: A systematic review. *International Journal of Environmental Research and Public Health*, *15*(11), 2392.

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence*, *149*(1), 91–130.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612.

Atherton, J., & Wang, G. (2020). Doing vs. Being: A philosophy of design for artful VR. *Journal of New Music Research*, *49*(1), 35–59.

Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: Sharing tasks with real and imagined co-actors. *Experimental Brain Research*, *211*(3–4), 371–385.

Ayoola, I., Stienstra, J., & Feijs, L. (2018). The Affective Respiration Device. In M. Mokhtari, B. Abdulrazak, & H. Aloulou (Eds.), *International Conference on Smart Homes and Health Telematics* (pp. 277–284). Springer.

Baker, C. (2018). Virtual, artificial and mixed reality: New frontiers in performance. In L. Goodman & A. Addison (Eds.), *Proceedings of the 2017 23rd International Conference on Virtual Systems and Multimedia: Vol. JAN* (pp. 1–10). IEEE.

Bakshi, A. J., Deshmukh, J., & Kumar, S. (2021). The psychology of pandemics: preparing for the next global outbreak of infectious disease. *British Journal of Guidance & Counselling*, *49*(4), 633–636.

Ball, M. (2022). *The metaverse: and how it will revolutionize everything*. Liveright.

Banakou, D., & Slater, M. (2014). Body ownership causes illusory self-attribution of speaking and influences subsequent real speaking. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(49), 17678–17683.

Bardy, B. G., Hoffmann, C. P., Moens, B., Leman, M., & Dalla Bella, S. (2015). Sound-induced stabilization of breathing and moving. *Annals of the New York Academy of Sciences*, *1337*(1), 94–100.

Barone, T. (2000). *Aesthetics, politics, and educational inquiry: essays and examples*. Peter Lang Inc.

Barrass, S., & Barrass, T. (2006). Musical creativity in collaborative virtual environments. *Virtual Reality*, *10*(2), 149–157.

Bartlette, C., Headlam, D., Bocko, M., & Velikic, G. (2006). Effect of network latency on interactive musical performance. *Music Perception*, *24*(1), 49–62.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1).

Baudrillard, J. (2000). The Vital Illusion. In *The Vital Illusion*. Columbia University Press.

Baumgartner, T., Valko, L., Esslen, M., & Jäncke, L. (2006). Neural correlate of spatial presence in an arousing and noninteractive virtual reality: An EEG and psychophysiology study. *Cyberpsychology and Behavior*, *9*(1), 30–45.

Beaudouin-Lafon, M. (2004). Designing interaction, not interfaces. In M. F. Costabile (Ed.), *Proceedings of the Workshop on Advanced Visual Interfaces AVI* (pp. 15–22). ACM New York.

Bégel, V., Demos, A. P., Wang, M., & Palmer, C. (2022). Social Interaction and Rate Effects in Models of Musical Synchronization. *Frontiers in Psychology*, *13*.

Benedikt, M. (1991). *Cyberspace: First steps*. MIT press.

Benford, S., & Giannachi, G. (2011). *Performing mixed reality*. MIT press.

Berg, L. P., & Vance, J. M. (2017). Industry use of virtual reality in product design and manufacturing: a survey. *Virtual Reality*, *21*(1), 1–17.

Bergstrom, I., Azevedo, S., Papiotis, P., Saldanha, N., & Slater, M. (2017). The Plausibility of a String Quartet Performance in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, *23*(4), 1352–1359.

Bergstrom, I., Seinfeld, S., Arroyo-Palacios, J., Slater, M., & Sanchez-Vives, M. V. (2014). Using music as a signal for biofeedback. *International Journal of Psychophysiology*, *93*(1), 140–149.

Bernardi, L., Porta, C., Casucci, G., Balsamo, R., Bernardi, N. F., Fogari, R., & Sleight, P. (2009). Dynamic interactions between musical, cardiovascular, and cerebral rhythms in humans. *Circulation*, *119*(25), 3171–3180.

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

Berthaut, F., Desainte-Catherine, M., & Hachet, M. (2011). Interacting with 3D Reactive Widgets for Musical Performance. *Journal of New Music Research*, *40*(3), 253–263.

Berthaut, F., Zappi, V., & Mazzanti, D. (2015). Scenography of immersive virtual musical instruments. In S. Serafin, F. Avanzini, M. Geronazzo, C. Erkut, A. De Goetzen, & R. Nordahl (Eds.), *IEEE VR Workshop: Sonic Interaction in Virtual Environments* (pp. 19–24).

Bevilacqua, F., Boyer, E. O., Françoise, J., Houix, O., Susini, P., Roby-Brami, A., & Hanneton, S. (2016). Sensori-motor learning with movement sonification: Perspectives from recent interdisciplinary studies. *Frontiers in Neuroscience*, *10*.

Biocca, F. (1992). Communication Within Virtual Reality: Creating a Space for Research. *Journal of Communication*, *42*(4), 5–22.

Biocca, F., & Delaney, B. (1995). Immersive Virtual Reality Technology. In F. Biocca & M. R. Levy (Eds.), *Communication in the Age of Virtual Reality* (pp. 57–124). Routledge.

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

Bissonnette, J., Dubé, F., Provencher, M. D., & Moreno Sala, M. T. (2016). Evolution of music performance anxiety and quality of performance during virtual reality exposure training. *Virtual Reality*, *20*(1), 71–81.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology: Just another methodological tool for social psychology? *Psychological Inquiry*, *13*(2), 146–149.

Bloom, B. S. (1953). Thought-processes in lectures and discussions. *The Journal of General Education*, *7*(3), 160–169.

Blum, J., Rockstroh, C., & Göritz, A. S. (2019). Heart rate variability biofeedback based on slow-paced breathing with immersive virtual reality nature scenery. *Frontiers in Psychology*, *10*.

Bod, R. (2013). Who’s afraid of patterns?: The particular versus the universal and the meaning of Humanities 3.0. *BMGN-Low Countries Historical Review*, *128*(4), 171–180.

Bodie, G. D., & Fitch-Hauser, M. (2010). Quantitative Research in Listening: Explication and Overview. In *Listening and Human Communication in the 21st Century* (pp. 46–93). Wiley Online Library.

Bødker, S. (2015). Third-wave HCI, 10 years later - Participation and sharing. *Interactions*, *22*(5), 24–31.

Bolt, N. K., & Loehr, J. D. (2017). The predictability of a partner’s actions modulates the sense of joint agency. *Cognition*, *161*, 60–65.

Borgo, D. (2005). *Sync or swarm: Improvising music in a complex age*. A&C Black.

Braun, N., Debener, S., Spychala, N., Bongartz, E., Sörös, P., Müller, H. H. O., & Philipsen, A. (2018). The senses of agency and ownership: A review. *Frontiers in Psychology*, *9*.

Bresler, L. (1995). Ethnography, Phenomenology And Action Research In Music Education. *Quarterly Journal of Music Teaching and Learning*, *6*(3), 4–17.

Brown, A. R., Gifford, T., & Voltz, B. (2016). Stimulating creative partnerships in human-agent musical interaction. *Computers in Entertainment*, *14*(2), 1–17.

Brown, S. C., & Knox, D. (2017). Why go to pop concerts? The motivations behind live music attendance. *Musicae Scientiae*, *21*(3), 233–249.

Bruni, L. E., Kadastik, N., Pedersen, T. A., & Dini, H. (2022). Digital Narratives in Extended Realities. In *Roadmapping Extended Reality* (pp. 35–62). Wiley Online Library.

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S. H., & Toiviainen, P. (2014). Hunting for the beat in the body: On period and phase locking in music-induced movement. *Frontiers in Human Neuroscience*, *8*.

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., & Toiviainen, P. (2013). Influences of Rhythm- and Timbre-Related Musical Features on Characteristics of Music-Induced Movement. *Frontiers in Psychology*, *4*.

Cairns, P., Daffern, H., & Kearney, G. (2020). Immersive network music performance: Design and practical deployment of a system for immersive vocal performance. In A. Andreopoulou & B. Boren (Eds.), *149th Audio Engineering Society Convention*.

Cairns, P., Hunt, A., Cooper, J., Johnston, D., Lee, B., Daffern, H., & Kearney, G. (2022). Recording Music in the Metaverse: A case study of XR BBC Maida Vale Recording Studios. In S. Serafin, P. Samarasinghe, D. Lou Alon, & H. Gamper (Eds.), *AES Conference on Audio for Virtual and Augmented Reality*.

Çamcı, A. (2019). Exploring the effects of diegetic and non-diegetic audiovisual cues on decision-making in virtual reality. In S. Serafin, F. Avanzini, I. Barbancho, & L. Tardón (Eds.), *Proceedings* *of the Sound and Music Computing Conferences* (pp. 195–201).

Çamcı, A., & Hamilton, R. (2020). Audio-first VR: New perspectives on musical experiences in virtual environments. *Journal of New Music Research*, 1–7.

Camurri, A., Poli, G. De, Leman, M., & Volpe, G. (2001). A Multi-layered Conceptual Framework for Expressive Gesture Applications. In J. Arnspang & K. Jensen (Eds.), *Proceedings of MOSART: Workshop on Current Directions in Computer Music* (pp. 29–34).

Caramiaux, B., Altavilla, A., Françoise, J., & Bevilacqua, F. (2022). Gestural Sound Toolkit: Reflections on an Interactive Design Project. *International Conference on New Interfaces for Musical Expression*.

Caramiaux, B., Montecchio, N., Tanaka, A., & Bevilacqua, F. (2015). Adaptive gesture recognition with variation estimation for interactive systems. *ACM Transactions on Interactive Intelligent Systems*, *4*(4), 18.

Carrer, L. R. J., Pompéia, S., & Miranda, M. C. (2022). Sensorimotor synchronization with music and metronome in school-aged children. *Psychology of Music*.

Carter, E. J., & Pollick, F. E. (2014). Not quite human: What virtual characters have taught us about person perception. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality* (p. 145). Oxford University Press.

Caruso, G., Coorevits, E., Nijs, L., & Leman, M. (2016). Gestures in Contemporary Music Performance: A Method to Assist the Performer’s Artistic Process. *Contemporary Music Review*, *35*(4–5), 402–422.

Casetti, F. (2018). Mediascapes: A decalogue. *Perspecta*, *51*, 21–43.

Casey, E. S. (2010). Aesthetic Experience. In H. R. Sepp & L. Embree (Eds.), *Contributions To Phenomenology* (Vol. 59, pp. 1–7). Springer.

Castellano, G., Mortillaro, M., Camurri, A., Volpe, G., & Scherer, K. (2008). Automated analysis of body movement in emotionally expressive piano performances. *Music Perception*, *26*(2), 103–119.

Catricalà, V., & Eugeni, R. (2020). Technologically Modified Self–Centred Worlds - Modes of Presence as Effects of Sensein Virtual, Augmented, Mixed and Extended Reality. In F. Biggio, V. Dos Santos, & G. T. Giuliana (Eds.), *Meaning–Making in Extended Reality* (pp. 63–90). Aracne Editrice.

Chang, A., Kragness, H. E., Livingstone, S. R., Bosnyak, D. J., & Trainor, L. J. (2019). Body sway reflects joint emotional expression in music ensemble performance. *Scientific Reports*, *9*(1), 1–11.

Chang, A., Livingstone, S. R., Bosnyak, D. J., & Trainor, L. J. (2017). Body sway reflects leadership in joint music performance. *Proceedings of the National Academy of Sciences*, *114*(21), E4134–E4141.

Chauvigné, L. A. S., Walton, A., Richardson, M. J., & Brown, S. (2019). Multi-person and multisensory synchronization during group dancing. *Human Movement Science*, *63*, 199–208.

Christensen, R. H. B. (2015). Analysis of ordinal data with cumulative link models—estimation with the R-package ordinal. *R-Package Version*, *28*, 145.

Ciciliani, M. (2020). Virtual 3D environments as composition and performance spaces. *Journal of New Music Research*, *49*(1), 104–113.

Cipresso, P. (2015). Modeling behavior dynamics using computational psychometrics within virtual worlds. *Frontiers in Psychology*, *6*.

Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, *9*.

Clark, M. E., & Hirschman, R. (1990). Effects of paced respiration on anxiety reduction in a clinical population. *Biofeedback and Self-Regulation*, *15*(3), 273–284.

Cochrane, T. (2017). Group flow. In *The Routledge Companion to Embodied Music Interaction* (pp. 133–140). Routledge.

Collins, N. (2011). Trading faures: Virtual musicians and machine ethics. *Leonardo Music Journal*, *21*, 35–39.

Cook, N. (2013). *Beyond the score: Music as performance*. Oxford University Press.

Coorevits, E., Maes, P.-J., Six, J., & Leman, M. (2020). The influence of performing gesture type on interpersonal musical timing, and the role of visual contact and tempo. *Acta Psychologica*, *210*, 103166.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Critchley, H. D. (2002). Electrodermal responses: What happens in the brain. *Neuroscientist*, *8*(2), 132–142.

Cross, I., & Morley, I. R. M. (2010). The evolution of music: Theories, definitions and the nature of the evidence. In S. Malloch & C. Trevarthen (Eds.), *Communicative Musicality: Exploring the Basis of Human Companionship* (pp. 61–82). Oxford University Press.

Cruz-Neira, C. (2003). Computational humanities: The new challenge for VR. *IEEE Computer Graphics and Applications*, *23*(3), 10–13.

Cui, G. (2013). Evaluating Online Social Presence: An Overview of Social Presence Assessment. *Journal of Educational Technology Development and Exchange*, *6*(1), 13–30.

D’Amario, S., Daffern, H., & Bailes, F. (2018). Synchronization in singing duo performances: The roles of visual contact and leadership instruction. *Frontiers in Psychology*, *9*.

D’Ausilio, A., Novembre, G., Fadiga, L., & Keller, P. E. (2015). What can music tell us about social interaction? *Trends in Cognitive Sciences*, *19*(3), 111–114.

Daffern, H., Camlin, D. A., Egermann, H., Gully, A. J., Kearney, G., Neale, C., & Rees-Jones, J. (2019). Exploring the potential of virtual reality technology to investigate the health and well being benefits of group singing. *International Journal of Performance Arts and Digital Media*, *15*(1), 1–22.

Damasio, A. R. (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain*. Harcourt Books.

de Borst, A. W., & de Gelder, B. (2015). Is it the real deal? Perception of virtual characters versus humans: An affective cognitive neuroscience perspective. *Frontiers in Psychology*, *6*.

De Bruyn, L., Leman, M., Demey, M., Desmet, F., & Moelants, D. (2008). Measuring and quantifying the impact of social interaction on listeners’ movement to music. In K. K. Jensen (Ed.), *Proceedings of the 2008 Computers in Music Modeling and Retrieval and Network for Cross-Disciplinary Studies of Music and Meaning Conference* (pp. 298–305).

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447.

De Oliveira, E. C., Bertrand, P., Lesur, M. E. R., Palomo, P., Demarzo, M., Cebolla, A., Banos, R., & Tori, R. (2016). Virtual Body Swap: A New Feasible Tool to Be Explored in Health and

Education. In R. Bilof (Ed.), *Proceedings - 18th Symposium on Virtual and Augmented Reality* (pp. 81–89).

De Wolf, T., & Holvoet, T. (2004). Emergence and Self-Organisation: a statement of similarities and differences. In S. A. Brueckner, G. M. Serugendo, A. Karageorgos, & R. Nagpal (Eds.), *Proceedings of the International Workshop on Engineering Self-Organising Applications* (pp. 96–110).

Deacon, T., Bryan-Kinns, N., Healey, P. G. T., & Barthet, M. (2019). Shaping sounds: The role of gesture in collaborative spatial music composition. In S. Dow, M. Lou Maher, A. Kerne, & C. Latulipe (Eds.), *Proceedings of the 2019 Creativity and Cognition* (pp. 121–132).

Deacon, T., Stockman, T., & Barthet, M. (2017). User experience in an interactive music virtual reality system: An exploratory study. In M. Aramaki, R. Kronland-Martinet, & S. Ystad (Eds.), *Lecture Notes in Computer Science* (Vol. 10525, pp. 192–216).

Debord, G. (1995). *Society of the Spectacle* (N.-S. Donald (ed.)). Zone Books.

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.

Dell'Anna, A., Buhmann, J., Six, J., Maes, P. J., & Leman, M. (2020). Timing Markers of Interaction Quality During Semi-Hocket Singing. *Frontiers in Neuroscience*, *14*.

Dell'Anna, A., Leman, M., & Berti, A. (2021). Musical Interaction Reveals Music as Embodied Language. *Frontiers in Neuroscience*, *15*.

Demos, A. P., Chaffin, R., Begosh, K. T., Daniels, J. R., & Marsh, K. L. (2012). Rocking to the beat: Effects of music and partner's movements on spontaneous interpersonal coordination. *Journal of Experimental Psychology: General*, *141*(1), 49–53.

Demos, A. P., Chaffin, R., & Kant, V. (2014). Toward a dynamical theory of body movement in musical performance. *Frontiers in Psychology*, *5*.

Demos, A. P., Chaffin, R., & Logan, T. (2018). Musicians body sway embodies musical structure and expression: A recurrence-based approach. *Musicae Scientiae*, *22*(2), 244–263.

Demos, A. P., Frank, T., & Chaffin, R. (2011). Understanding movement during performance: A recurrence quantization approach. In A. Williamon, D. Edwards, & L. Bartel (Eds.), *International Symposium on Performance Science* (pp. 259–264).

Demos, A. P., Layeghi, H., Wanderley, M. M., & Palmer, C. (2019). Staying Together: A Bidirectional Delay–Coupled Approach to Joint Action. *Cognitive Science*, *43*(8), e12766.

Desmet, F., Nijs, L., Demey, M., Lesaffre, M., Martens, J. P., & Leman, M. (2012). Assessing a Clarinet Player's Performer Gestures in Relation to Locally Intended Musical Targets. *Journal of New Music Research*, *41*(1), 31–48.

Doel, M. A., & Clarke, D. B. (2013). Virtual worlds: Simulation, suppletion, s(ed)uction and simulacra. In M. Crang, P. Crang, & J. May (Eds.), *Virtual geographies* (pp. 268–290). Routledge.

Doffman, M. (2009). Making It Groove! Entrainment, Participation and Discrepancy in the "Conversation" of a Jazz Trio. *Language & History*, *52*(1), 130–147.

Dotov, D., Bosnyak, D., & Trainor, L. J. (2021). Collective music listening: Movement energy is enhanced by groove and visual social cues. *Quarterly Journal of Experimental Psychology*, *74*(6), 1037–1053.

Dotov, D., Delasanta, L., Cameron, D. J., Large, E., & Trainor, L. (2022). Collective dynamics support group drumming, reduce variability, and stabilize tempo drift. *ELife*, *11*.

Dourish, P. (2001). Where the Action Is: The Foundations of Embodied Interaction. *Where the Action Is the Foundations of Embodied Interaction*, *36*(3), 233.

Driessen, P. F., Darcie, T. E., & Pillay, B. (2011). The effects of network delay on tempo in musical performance. *Computer Music Journal*, *35*(1), 76–89.

Dumas, G., Lachat, F., Martinerie, J., Nadel, J., & George, N. (2011). From social behaviour to brain synchronization: Review and perspectives in hyperscanning. *IRBM*, *32*(1), 48–53.

Dura, M. T. (2006). The phenomenology of the music-listening experience. *Arts Education Policy Review*, *107*(3), 25–32.

Duranti, A. (2010). Husserl, intersubjectivity and anthropology. *Anthropological Theory*, *10*(1–2), 16–35.

Eberle, T. S. (2014). Phenomenology as a Research Method. In U. Flick (Ed.), *The SAGE Handbook of Qualitative Data Analysis* (Vol. 10, pp. 184–202). Sage publications.

Eerola, T., Jakubowski, K., Moran, N., Keller, P. E., & Clayton, M. (2018). Shared periodic performer movements coordinate interactions in duo improvisations. *Royal Society Open Science*, *5*(2), 171520.

Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science*, *317*(5841), 1048.

Ehrsson, H. H. (2019). Multisensory processes in body ownership. In *Multisensory Perception: From Laboratory to Clinic* (pp. 179–200). Elsevier.

Ekman, P. (1992). Are There Basic Emotions? *Psychological Review*, *99*(3), 550–553.

Engel, A. K., Friston, K. J., & Kragic, D. (2016). Where's the Action? In *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science* (pp. 1–15). MIT Press.

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, *32*(3), 158–172.

Estève, F., Blanc-Gras, N., Gallego, J., & Benchetrit, G. (1996). The effects of breathing pattern training on ventilatory function in patients with COPD. *Biofeedback and Self-Regulation*, *21*(4), 311–321.

Feijs, L., Langereis, G., & van Boxtel, G. (2010). Designing for heart rate and breathing movements. In L.-L. Chen, T. Djajadiningrat, L. Feijs, S. Kyffin, D. Steffen, & B. Young (Eds.), *Design and semantics of form and movement* (Vol. 57, pp. 57–67). Koninklijke Philips Electronics NV.

Fiebrink, R., Trueman, D., Britt, C., Nagai, M., Kaczmarek, K., Early, M., Daniel, M. R., Hege, A., & Cook, P. (2010). Toward understanding human-computer interaction in composing the instrument. *International Computer Music Conference*, 135–142.

Fiorella, L., Van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology*, *109*(5), 653–665.

Fogel, A. (1993). Two principles of communication: Co-regulation and framing. *New Perspectives in Communicative Developmnet*, 9–22.

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual Reality: A Survival Guide for the Social Scientist. *Journal of Media Psychology*, *21*(3), 95–113.

Françoise, J., Caramiaux, B., & Bevilacqua, F. (2012). A hierarchical approach for the design of gesture-to-sound mappings. In X. Shao, K. Qian, X. Wang, & K. Zhang (Eds.), *Proceedings of the 9th Sound and Music Computing Conference*.

Frankel, R., & Krebs, V. J. (2021). *Human virtuality and digital life: Philosophical and psychoanalytic* *investigations*. Routledge.

Frich, J., Biskjaer, M. M., & Dalsgaard, P. (2018). Twenty years of creativity research in human-computer interaction: Current state and future directions. In I. Koskinen & Y.-K. Lim (Eds.), *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 1235–1258).

Fried, R. (1990). Integrating music in breathing training and relaxation: II. Applications. *Biofeedback and Self-Regulation*, *15*(2), 171–177.

Friedman, D., Pizarro, R., Or-Berkers, K., Neyret, S., Pan, X., & Slater, M. (2014). A method for generating an illusion of backwards time travel using immersive virtual reality - an exploratory study. *Frontiers in Psychology*, *5*.

Froese, T., & Fuchs, T. (2012). The extended body: A case study in the neurophenomenology of social interaction. *Phenomenology and the Cognitive Sciences*, *11*(2), 205–235.

Gallagher, S. (2001). The practice of mind: Theory, simulation or primary interaction? *Journal of Consciousness Studies*, *8*(6–7), 83–108.

Gallagher, S. (2012). Phenomenology. In *Phenomenology* (pp. 7–18). Palgrave-Macmillan.

Gallagher, S., & Allen, M. (2018). Active inference, enactivism and the hermeneutics of social cognition. *Synthese*, *195*(6), 2627–2648.

Gallese, V. (2020). The aesthetic world in the digital era a call to arms for experimental aesthetics. *Reti Saperi Linguaggi*, *7*(1), 55–83.

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences*, *17*(4), 160–165.

Garau, M., Slater, M., Pertaub, D. P., & Razzaque, S. (2005). The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, *14*(1), 104–116.

Garner, T. A., & Grimshaw, M. (2014). Sonic Virtuality: Understanding Audio In A Virtual World. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality* (pp. 364–377). Oxford University Press.

Geeves, A., & Sutton, J. (2015). Embodied Cognition, Perception, and Performance in Music. *Empirical Musicology Review*, *9*(3–4), 247.

Gibson, W. (1984). *Neuromancer* (p. 271). Ace Books.

Giordano, F., Scarlata, E., Baroni, M., Gentile, E., Puntillo, F., Brienza, N., & Gesualdo, L. (2020). Receptive music therapy to reduce stress and improve wellbeing in Italian clinical staff involved in COVID-19 pandemic: A preliminary study. *Arts in Psychotherapy*, *70*, 101688.

Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., & Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology*, *16*(4), 428–432.

Glowinski, D., Baron, N., Shirole, K., Yahia Coll, S., Chaabi, L., Ott, T., Rappaz, M.-A., & Grandjean, D. (2015). Evaluating music performance and context-sensitivity with Immersive Virtual Environments. *EAI Endorsed Transactions on Creative Technologies*, *2*(2), e3.

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, *26*(5), 427–438.

Gonçalves, G., Melo, M., Barbosa, L., Vasconcelos-Raposo, J., & Bessa, M. (2022). Evaluation of the impact of different levels of self-representation and body tracking on the sense of presence and embodiment in immersive VR. *Virtual Reality*, *26*(1), 1–14.

González-Franco, M., Pérez-Marcos, D., Spanlang, B., & Slater, M. (2010). The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment. In B. Lok, G. Klinker, & R. Nakatsu (Eds.), *Proceedings - IEEE Virtual Reality* (pp.

111–114).

Gonzalez-Sanchez, V. E., Zelechowska, A., & Jensenius, A. R. (2018). Correspondences between music and involuntary human micromotion during standstill. *Frontiers in Psychology*, *9*.

Gordon, I., Gilboa, A., Cohen, S., Milstein, N., Haimovich, N., Pinhasi, S., & Siegman, S. (2020). Physiological and Behavioral Synchrony Predict Group Cohesion and Performance. *Scientific Reports*, *10*(1), 1–12.

Goupil, L., Wolf, T., Saint-Germier, P., Aucouturier, J. J., & Canonne, C. (2021). Emergent Shared Intentions Support Coordination During Collective Musical Improvisations. *Cognitive Science*, *45*(1), e12932.

Grassini, S., Laumann, K., & Rasmussen Skogstad, M. (2020). The Use of Virtual Reality Alone Does Not Promote Training Performance (but Sense of Presence Does). *Frontiers in Psychology*, *11*.

Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2007). Listening to music as a re-creative process: Physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Perception*, *24*(3), 297–314.

Grimshaw, M. (2014). *The Oxford handbook of virtuality*. Oxford University Press.

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, *11*(5/6), 561–566.

Hagan, K. (2008). Textural composition and its space. In M. Supper & S. Weinzierl (Eds.), *Proceedings of the 5th Sound and Music Computing Conference*.

Halák, J. (2016). Merleau-Ponty on embodied subjectivity from the perspective of subject-object circularity. *AUC KINANTHROPOLOGICA*, *52*(2), 26–40.

Hale, J., & Hamilton, A. F. D. C. (2016). Testing the relationship between mimicry, trust and rapport in virtual reality conversations. *Scientific Reports*, *6*(1), 1–11.

Hamilton, R. (2019). Mediated Musical Interactions in Virtual Environments. In *Springer Series on Cultural Computing* (pp. 243–257). Springer.

Hamilton, R., Caceres, J. P., Nanou, C., & Platz, C. (2011). Multi-modal musical environments for mixed-reality performance. *Journal on Multimodal User Interfaces*, *4*(3–4), 147–156.

Harris, J., Parnandi, A., Vance, S., Gutierrez-Osuna, R., & Fernandes, O. (2014). Sonic respiration: Controlling respiration rate through auditory biofeedback. In M. Jones & P. Palanque (Eds.), *Conference on Human Factors in Computing Systems - Proceedings* (pp. 2383–2388).

Harris, V. A., Katkin, E. S., Lick, J. R., & Habberfield, T. (1976). Paced Respiration as a Technique for the Modification of Autonomic Response to Stress. *Psychophysiology*, *13*(5), 386–391.

Harrison, S., Tatar, D., & Sengers, P. (2007). The three paradigms of HCI. *Alt. Chi. Session at the SIGCHI Conference on Human Factors in Computing*, 1–18.

Harry, B., & Keller, P. E. (2019). Tutorial and simulations with ADAM: an adaptation and anticipation model of sensorimotor synchronization. *Biological Cybernetics*, *113*(4), 397–421.

Heelan, P. A., & Schulkin, J. (1998). Hermeneutical philosophy and pragmatism: A philosophy of science. *Synthese*, *115*(3), 269–302.

Heggli, O. A., Cabral, J., Konvalinka, I., Vuust, P., & Kringelbach, M. L. (2019). A Kuramoto model of self-other integration across interpersonal synchronization strategies. *PLoS Computational Biology*, *15*(10).

Heggli, O. A., Konvalinka, I., Kringelbach, M. L., & Vuust, P. (2019). Musical interaction is influenced by underlying predictive models and musical expertise. *Scientific Reports*, *9*(1), 11048.

Heidegger, M. (2014). The question concerning technology. In R. C. Scharff & V. Dusek (Eds.), *Philosophy of technology: The technological condition* (pp. 305–317). John Wiley & Sons.

Heilig, M. L. (1960). Stereoscopic-television apparatus for individual use. In *US Patent 2,955,156*. Google Patents. https://patents.google.com/patent/US2955156A/en

Heim, M. (1994). The metaphysics of virtual reality. In *The Metaphysics of Virtual Reality*. Oxford University Press.

Herath, S. S., Kusumi, A., Note, Y., Nakamura, E., Nozawa, A., & Nomura, S. (2018). Enhancing cardiac parasympathetic nervous activity using a posture respiration ambient biofeedback system: a pilot study. *Artificial Life and Robotics*, *23*(1), 23–27.

Herrera, F., Bailenson, J., Weisz, E., Ogle, E., & Zaki, J. (2018). Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PloS One*, *13*(10), e0204494.

Herrera, G., Jordan, R., & Vera, L. (2006). Agency and presence: A common dependence on subjectivity? *Presence: Teleoperators and Virtual Environments*, *15*(5), 539–552.

Hickman, L. (2014). Putting Pragmatism (especially Dewey's) to Work. In R. C. Scharff & V. Dusek (Eds.), *Philosophy of technology: The technological condition* (Vol. 2, pp. 406–420). John Wiley & Sons.

Hilt, P. M., Badino, L., D'Ausilio, A., Volpe, G., Tokay, S., Fadiga, L., & Camurri, A. (2020). Multi-layer adaptation of group coordination in musical ensembles. *Scientific Reports*, *10*(1), 597.

Hilton, C., Plant, N., Fiebrink, R., Perry, P., Diaz, C. G., Gibson, R., Martelli, B., Zbyszynski, M., & Gillies, M. (2021). InteractML: Making machine learning accessible for creative practitioners working with movement interaction in immersive media. In Y. Itoh, K. Takashima, P. Punpongsanon, M. Sra, K. Fujita, S. Yoshida, & T. Piumsomboon (Eds.), *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (pp. 1–10).

Holland, S., Wilkie, K., Mulholland, P., & Seago, A. (2013). Music Interaction: Understanding Music and Human-Computer Interaction. In *Springer Series on Cultural Computing* (pp. 1–28). Springer.

Holmes, E. A., O'Connor, R. C., Perry, V. H., Tracey, I., Wessely, S., Arseneault, L., Ballard, C., Christensen, H., Cohen Silver, R., Everall, I., Ford, T., John, A., Kabir, T., King, K., Madan, I., Michie, S., Przybylski, A. K., Shafran, R., Sweeney, A., … Bullmore, E. (2020). Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *The Lancet Psychiatry*, *7*(6), 547–560.

Hong, J. Y., Lam, B., Ong, Z. T., Ooi, K., Gan, W. S., Kang, J., Feng, J., & Tan, S. T. (2019). Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications. *Building and Environment*, *149*, 1–14.

Honigman, C., Walton, A., & Kapur, A. (2013). The third room: A 3d virtual music paradigm. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 29–34.

Horne, J., Langsdorf, L., & Smith, A. (1998). Changing the Subject: Rorty and Contemporary Rhetorical Theory. In L. Langsdorf & A. Smith (Eds.), *Recovering Pragmatism's Voice: The Classical Tradition, Rorty, and the Philosophy of Communication* (pp. 235–249). SUNY Press.

Huang, H., Solah, M., Li, D., & Yu, L. F. (2019). Audible panorama: Automatic spatial audio generation for panorama imagery. In S. Brewster & G. Fitzpatrick (Eds.), *Conference on Human Factors in Computing Systems - Proceedings* (pp. 1–11).

Hume, K., & Ahtamad, M. (2013). Physiological responses to and subjective estimates of soundscape elements. *Applied Acoustics*, *74*(2), 275–281.

Hunt, N., McGrath, D., & Stergiou, N. (2014). The influence of auditory-motor coupling on fractal dynamics in human gait. *Scientific Reports*, *4*, 5879.

Hunter, P. G., & Schellenberg, E. G. (2010). Music and emotion. In *Music Perception* (pp. 129–164). Springer.

Husserl, E. (1970). *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy* (D. Carr (ed.)). Northwestern University Press.

Hutto, D. D., & McGivern, P. (2015). How embodied is cognition? *The Philosophers' Magazine*, *68*, 77–83.

Ihde, D. (2009). A phenomenology of technics. In C. Hanks (Ed.), *Technology and values: Essential readings* (pp. 134–155). John Wiley & Sons.

Ijsselsteijn, W., & Riva, G. (2003). Being there: The experience of presence in mediated environments. In G. Riva, F. Davide, & W. A. Ijsselsteijn (Eds.), *Being there: Concepts, effects and measurement of user presence in synthetic environments* (pp. 4–14). IOS Press.

Impett, L., Herman, I., Wollner, P. K. A., & Blackwell, A. F. (2015). Musician fantasies of dialectical interaction: Mixed-initiative interaction and the open work. In M. Kurosu (Ed.), *Lecture Notes in Computer Science* (Vol. 9170, pp. 184–195).

Iqbal, T., & Riek, L. D. (2018). Human-Robot Teaming: Approaches from Joint Action and Dynamical Systems. In *Humanoid Robotics: A Reference* (pp. 2293–2312). Springer.

Jacucci, G. (2015). Interaction as performance: Performative strategies in designing interactive experiences. In U. Ekman, J. D. Bolter, L. Diaz, M. Sondergaard, & M. Engberg (Eds.), *Ubiquitous Computing, Complexity and Culture* (pp. 350–363). Routledge.

Jafari, H., de Broek, K., Plaghki, L., Vlaeyen, J. W. S., den Bergh, O., & Van Diest, I. (2016). Respiratory hypoalgesia? Breath-holding, but not respiratory phase modulates nociceptive flexion reflex and pain intensity. *International Journal of Psychophysiology*, *101*, 50–58.

Jäger, L., & Kim, J. H. (2021). Transparency and Opacity: Interface Technology of Mediation in New Media Art. In U. Seifert, J. H. Kim, & A. Moore (Eds.), *Paradoxes of Interactivity* (pp. 44–61). Transcript Verlag.

Jenkins, H. (2004). Game Design as Narrative Architecture. *Computer*, *44*(3), 118–130.

Jerath, R., Edry, J. W., Barnes, V. A., & Jerath, V. (2006). Physiology of long pranayamic breathing: Neural respiratory elements may provide a mechanism that explains how slow deep breathing shifts the autonomic nervous system. *Medical Hypotheses*, *67*(3), 566–571.

Johanson, M. (2015). The Turing Test for Telepresence. *The International Journal of Multimedia & Its Applications*, *7*(4/5), 01–20.

Jun, J., Jung, M., Kim, S. Y., & Kim, K. (2018). Full-body ownership illusion can change our emotion. In R. Mandryk & M. Hancock (Eds.), *Conference on Human Factors in Computing Systems - Proceedings* (Vols. 2018-April, pp. 1–11).

Juslin, P. N., & Sloboda, J. A. (2013). Music and Emotion. In *The Psychology of Music* (pp. 583–645). Elsevier Academic Press.

Kalckert, A., & Ehrsson, H. H. (2014). The moving rubber hand illusion revisited: Comparing movements and visuotactile stimulation to induce illusory ownership. *Consciousness and Cognition*, *26*(1), 117–132.

Katz, B. F. G., & Parseihian, G. (2012). Perceptually based head-related transfer function database optimization. *The Journal of the Acoustical Society of America*, *131*(2), EL99–EL105.

Kawase, S. (2014). Gazing behavior and coordination during piano duo performance. *Attention,* *Perception, and Psychophysics*, *76*(2), 527–540.

Kearney, R. (1995). *Poetics of modernity: Toward a hermeneutic imagination*. Humanity Books.

Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 260–282). Oxford University Press.

Keller, P. E., & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception*, *28*(1), 27–46.

Keller, P. E., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society: Biological Sciences*, *369*(1658), 20130394.

Keller, P. E., Weber, A., & Engel, A. (2011). Practice makes too perfect: Fluctuations in loudness indicate spontaneity in musical improvisation. *Music Perception*, *29*(1), 109–114.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. MIT press.

Kelso, J. A. S. (2009). Coordination Dynamics. In *Encyclopedia of Complexity and Systems Science* (pp. 1537–1564). Springer.

Kendall, G. S. (2010). Spatial perception and cognition in multichannel audio for electroacoustic music. *Organised Sound*, *15*(3), 228–238.

Kern, A. C., & Ellermeier, W. (2020). Audio in VR: Effects of a Soundscape and Movement-Triggered Step Sounds on Presence. *Frontiers in Robotics and AI*, *7*.

Kets, N. Van, Moens, B., Bombeke, K., Durnez, W., Maes, P.-J., Wallendael, G. Van, Marez, L. De, Leman, M., & Lambert, P. (2021). Art and Science Interaction Lab - {A} highly flexible and modular interaction science research facility. *CoRR*, *2101.1*. https://arxiv.org/abs/2101.11691

Kilteni, K., Bergstrom, I., & Slater, M. (2013). Drumming in immersive virtual reality: The body shapes the way we play. *IEEE Transactions on Visualization and Computer Graphics*, *19*(4), 597–605.

Kilteni, K., Groten, R., & Slater, M. (2013). The Sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, *22*(1), 373–387.

Kilteni, K., Maselli, A., Kording, K. P., & Slater, M. (2015). Over my fake body: Body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in Human Neuroscience*, *9*.

Kirschenbaum, M. G. (2016). What is digital humanities and what's it doing in English departments? *Defining Digital Humanities*, 211–220.

Kitson, A., DiPaola, S., & Riecke, B. E. (2019). Lucid loop: A virtual deep learning biofeedback system for lucid dreaming practice. In S. Brewster & G. Fitzpatrick (Eds.), *Conference on Human Factors in Computing Systems - Proceedings* (pp. 1–6).

Kiverstein, J. (2011). Social understanding without mentalizing. *Philosophical Topics*, *39*(1), 41–65.

Koban, L., Ramamoorthy, A., & Konvalinka, I. (2019). Why do we fall into sync with others? Interpersonal synchronization and the brain's optimization principle. *Social Neuroscience*, *14*(1), 1–9.

Kobayashi, M., Ueno, K., & Ise, S. (2015). The effects of spatialized sounds on the sense of presence in auditory virtual environments: A psychological and physiological study. *Presence: Teleoperators and Virtual Environments*, *24*(2), 163–174.

Koelsch, S. (2015). Music-evoked emotions: Principles, brain correlates, and implications for therapy. *Annals of the New York Academy of Sciences*, *1337*(1), 193–201.

Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: How can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience*, *6*.

Konvalinka, I., Sebanz, N., & Knoblich, G. (2023). The role of reciprocity in dynamic interpersonal coordination of physiological rhythms. *Cognition*, *230*, 105307.

Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me: Continuous mutual prediction and adaptation in joint tapping. *Quarterly Journal of Experimental Psychology*, *63*(11), 2220–2230.

Kornelsen, J. (1991). *Virtual Reality? Marshall McLuhan and a Phenomenological Investigation of the Construction of Virtual Worlds*. Simon Fraser University, CA.

Kostrubiec, V., Dumas, G., Zanone, P. G., & Scott Kelso, J. A. (2015). The virtual teacher (VT) paradigm: Learning new patterns of interpersonal coordination using the human dynamic clamp. *PLoS ONE*, *10*(11), e0142029.

Kostrubiec, V., Zanone, P. G., Fuchs, A., & Scott Kelso, J. A. (2012). Beyond the blank slate: Routes to learning new coordination patterns depend on the intrinsic dynamics of the learner-experimental evidence and theoretical model. *Frontiers in Human Neuroscience*, *6*.

Kothgassner, O. D., & Felnhofer, A. (2020). Does virtual reality help to cut the Gordian knot between ecological validity and experimental control? *Annals of the International Communication Association*, *44*(3), 210–218.

Kothgassner, O. D., Griesinger, M., Kettner, K., Wayan, K., Völkl-Kernstock, S., Hlavacs, H., Beutl, L., & Felnhofer, A. (2017). Real-life prosocial behavior decreases after being socially excluded by avatars, not agents. *Computers in Human Behavior*, *70*, 261–269.

Kox, M., Van Eijk, L. T., Zwaag, J., Van Den Wildenberg, J., Sweep, F. C. G. J., Van Der Hoeven, J. G., & Pickkers, P. (2014). Voluntary activation of the sympathetic nervous system and attenuation of the innate immune response in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(20), 7379–7384.

Kramer, G., Walker, B., Bonebright, T., Cook, P., & Flowers, J. H. (2010). Sonification Report: Status of the Field and Research Agenda. In *Faculty Publications, Department of Psychology* (Vol. 444). https://digitalcommons.unl.edu/psychfacpub/444

Krause, B. L. (1993). The Niche Hypothesis: A Virtual Symphony of Animal Sounds, The Origins of Musical Expression and the Health of Habitats. *The Soundscape Newsletter*, *6*(6), 6–10.

Kreps, D. (2014). Virtuality and humanity. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality* (pp. 712–726). Oxford University Press.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27.

Lamb, M., Nalepka, P., Kallen, R. W., Lorenz, T., Harrison, S. J., Minai, A. A., & Richardson, M. J. (2019). A Hierarchical Behavioral Dynamic Approach for Naturally Adaptive Human-Agent Pick-and-Place Interactions. *Complexity*.

Landy, L. (2019). Understanding the Art of Sound Organization. In *Understanding the Art of Sound Organization*. MIT Press.

Lang, M., Shaw, D. J., Reddish, P., Wallot, S., Mitkidis, P., & Xygalatas, D. (2016). Lost in the Rhythm: Effects of Rhythm on Subsequent Interpersonal Coordination. *Cognitive Science*, *40*(7), 1797–1815.

Lanier, J. (1988). A vintage virtual reality interview. *Whole Earth Review*, *64*, 108–119.

Large, E. W. (2008). Resonating to musical rhythm: theory and experiment. In *Psychology of time*.

Emerald West Yorkshire.

Larsson, P., Väljamäe, A., Västfjäll, D., Tajadura-Jiménez, A., & Kleiner, M. (2010). Auditory-Induced Presence in Mixed Reality Environments and Related Technology. In *The engineering of mixed reality systems* (pp. 143–163). Springer.

Launay, J., Dean, R. T., & Bailes, F. (2014). Synchronising movements with the sounds of a virtual partner enhances partner likeability. *Cognitive Processing*, *15*(4), 491–501.

Lee, H. (2020). *A Conceptual Model of Immersive Experience in Extended Reality* (Issue SEP). PsyArXiv.

Lee, K. M. (2004). Presence, explicated. *Communication Theory*, *14*(1), 27–50.

Lefebvre, H. (2012). From the production of space. In J. Collins & A. Nisbet (Eds.), *Theatre and Performance Design* (pp. 81–84). Routledge.

Lehrer, P. M., Vaschillo, E., & Vaschillo, B. (2000). Resonant frequency biofeedback training to increase cardiac variability: Rationale and manual for training. *Applied Psychophysiology Biofeedback*, *25*(3), 177–191.

Leman, M. (2007). *Embodied music cognition and mediation technology*. MIT press.

Leman, M. (2016). *The expressive moment: How interaction (with music) shapes human empowerment*. MIT press.

Leman, M. (2021). Co-regulated timing in music ensembles: A Bayesian listener perspective. *Journal of New Music Research*, *50*(2), 121–132.

Leman, M., & Maes, P.-J. (2015). The Role of Embodiment in the Perception of Music. *Empirical Musicology Review*, *9*(3–4), 236.

Leman, M., & Six, J. (2018). Beyond documentation–The digital philology of interaction heritage. *Journal of New Music Research*, *47*(4), 309–320.

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. *Science*, *317*(5841), 1096–1099.

Lennox, P. (2009). The emotional contents of the ‘space’ in spatial music. *International Conference on Music and Emotion*, *SEP*.

Leow, L. A., Parrott, T., & Grahn, J. A. (2014). Individual differences in beat perception affect gait responses to low- and high-groove music. *Frontiers in Human Neuroscience*, *8*.

Lerch, A., Arthur, C., Pati, A., & Gururani, S. (2020). An Interdisciplinary Review of Music Performance Analysis. *Transactions of the International Society for Music Information Retrieval*, *3*(1), 221–245.

Lesaffre, M., & Leman, M. (2020). Integrative Research in Art and Science: A Framework for Proactive Humanities. *Critical Arts*, *34*(5), 39–54.

Lesaffre, M., Maes, P.-J., & Leman, M. (2017). *The Routledge companion to embodied music interaction*. Taylor & Francis.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence: Teleoperators and Virtual Environments*, *10*(3), 282–297.

Levitin, D. J., & Tirovolas, A. K. (2009). Current advances in the cognitive neuroscience of music. *Annals of the New York Academy of Sciences*, *1156*(1), 211–231.

Lichty, P. (2014). The Translation of Art in Virtual Worlds. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality* (pp. 444–462). Oxford University Press.

Loehr, J. D. (2022). The sense of agency in joint action: An integrative review. *Psychonomic Bulletin* *and Review*, *29*(4), 1089–1117.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, *3*(2), JCMC321.

Loveridge, B. (2020). Networked Music Performance in Virtual Reality: Current Perspectives. *Journal of Network Music and Arts*, *2*(1), 2.

Lüdecke, D. (2014). sjPlot: Data Visualization for Statistics in Social Science. In *R package version 2.8.12*. https://cran.r-project.org/package=sjPlot

Määttänen, P. (2007). Semiotics of space: Peirce and Lefebvre. *Semiotica*, *166*, 453–461.

Maes, P. J. (2016). Sensorimotor grounding of musical embodiment and the role of prediction: A review. *Frontiers in Psychology*, *7*.

Maes, P. J., Buhmann, J., & Leman, M. (2016). 3Mo: A model for music-based biofeedback. *Frontiers in Neuroscience*, *10*.

Maes, P. J., Leman, M., Palmer, C., & Wanderley, M. M. (2014). Action-based effects on music perception. *Frontiers in Psychology*, *4*.

Maes, P. J., Lorenzoni, V., Moens, B., Six, J., Bressan, F., Schepers, I., & Leman, M. (2018). Embodied, Participatory Sense-Making in Digitally-Augmented Music Practices: Theoretical Principles and the Artistic Case “SoundBikes.” *Critical Arts*, *32*(3), 77–94.

Maes, P. J., Lorenzoni, V., & Six, J. (2019). The SoundBike: musical sonification strategies to enhance cyclists’ spontaneous synchronization to external music. *Journal on Multimodal User Interfaces*, *13*(3), 155–166.

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, *72*, 276–285.

Malloch, J., Garcia, J., Wanderley, M. M., Mackay, W. E., Beaudouin-Lafon, M., & Huot, S. (2019). A Design Workbench for Interactive Music Systems. In *Springer Series on Cultural Computing* (pp. 23–40). Springer.

Mardia, K. V, & Jupp, P. E. (2008). Directional Statistics. In *Directional Statistics* (Vol. 494). John Wiley & Sons.

Martin-Gutierrez, J., Del Rio Guerra, M. S., Lopez-Chao, V., Gastelum, R. H. S., & Bojórquez, J. F. V. (2020). Augmented reality to facilitate learning of the acoustic guitar. *Applied Sciences*, *10*(7), 2425.

Martin, A. J., & Jackson, S. A. (2008). Brief approaches to assessing task absorption and enhanced subjective experience: Examining “short” and “core” flow in diverse performance domains. *Motivation and Emotion*, *32*(3), 141–157.

Martin, A. K., Perceval, G., Davies, I., Su, P., Huang, J., & Meinzer, M. (2019). Visual Perspective Taking in Young and Older Adults. *Journal of Experimental Psychology: General*, *148*(11), 2006.

Martin, C. P., Liu, Z., Wang, Y., He, W., & Gardner, H. (2020). Sonic Sculpture: Activating Engagement with Head-Mounted Augmented Reality. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 48–52.

Martin, D., Malpica, S., Gutierrez, D., Masia, B., & Serrano, A. (2022). Multimodality in VR: A Survey. *ACM Computing Surveys*, *54*(10s), 1–36.

Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*(5–6), 237–329.

Massumi, B. (2014). Envisioning the virtual. In M. Grimshaw (Ed.), *The Oxford Handbook of*

Virtuality (pp. 55–70). Oxford University Press.

Matamala-Gomez, M., Maselli, A., Malighetti, C., Realdon, O., Mantovani, F., & Riva, G. (2021). Virtual body ownership illusions for mental health: A narrative review. *Journal of Clinical Medicine*, *10*(1), 1–19.

Mazzanti, D., Zappi, V., Caldwell, D., & Brogni, A. (2014). Augmented Stage for Participatory Performances. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 29–34.

McCall, R., Wetzel, R., Löschner, J., & Braun, A. K. (2011). Using presence to evaluate an augmented reality location aware game. *Personal and Ubiquitous Computing*, *15*(1), 25–35.

McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2014). Stereoscopic 3D displays and human performance: A comprehensive review. *Displays*, *35*(1), 18–26.

McLuhan, M. (1964). *Understanding media: The extensions of man*. McGraw-Hill Education.

Medvedev, O., Shepherd, D., & Hautus, M. J. (2015). The restorative potential of soundscapes: A physiological investigation. *Applied Acoustics*, *96*, 20–26.

Meehan, M., Insko, B., Whitton, M., & Brooks, F. P. (2002). Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics*, *21*(3), 645–652.

Men, L., & Bryan-Kinns, N. (2019). LeMo: Exploring virtual space for collaborative creativity. In S. Dow & M. Lou Maher (Eds.), *Proceedings of the 2019 Creativity and Cognition* (pp. 71–82).

Merleau-Ponty, M. (2013). Phenomenology of perception. In *Phenomenology of Perception*. Routledge.

Metzinger, T. (2014). First-order embodiment, second-order embodiment, third-order embodiment. In L. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 272–286). Routledge.

Metzinger, T. K. (2018). Why is virtual reality interesting for philosophers? *Frontiers Robotics AI*, *5*.

Michałko, A., Campo, A., Nijs, L., Leman, M., & Van Dyck, E. (2022). Toward a meaningful technology for instrumental music education: Teachers' voice. *Frontiers in Education*, *7*.

Milgram, P., & Kishino, F. (1994). Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, *E77-D*(12), 1321–1329.

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, *53*(4), 1551–1562.

Milward, S. J., & Sebanz, N. (2016). Mechanisms and development of self--other distinction in dyads and groups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150076.

Minksy, M. (1980). Telepresence. *Omni*, 45–52.

Mitcham, C. (2014). Three Ways of Being-With Technology. In R. C. Scharff & V. Dusek (Eds.), *Philosophy of technology: The technological condition* (pp. 523–538). John Wiley & Sons.

Mitchell, R. W. (2018). The expressive moment: How interaction (with music) shapes human empowerment , by Marc Leman. *Interaction Studies*, *19*(3), 519–521.

Miura, A., Kudo, K., & Nakazawa, K. (2013). Action-perception coordination dynamics of whole-body rhythmic movement in stance: A comparison study of street dancers and non-dancers. *Neuroscience Letters*, *544*, 157–162.

Miyata, K., Varlet, M., Miura, A., Kudo, K., & Keller, P. E. (2017). Modulation of individual auditory-motor coordination dynamics through interpersonal visual coupling. *Scientific Reports*, *7*(1), 1–11.

Moens, B., & Leman, M. (2015). Alignment strategies for the entrainment of music and movement rhythms. *Annals of the New York Academy of Sciences*, *1337*(1), 86–93.

Moens, B., van Noorden, L., & Leman, M. (2010). D-Jogger: Syncing music with walking. *Proceedings of the 7th Sound and Music Computing Conference, SMC 2010*, 23.

Møller, C., Stupacher, J., Celma-Miralles, A., & Vuust, P. (2021). Beat perception in polyrhythms: Time is structured in binary units. *PLoS ONE*, *16*(8 August), e0252174.

Moore, C., & Barresi, J. (2017). The role of second-person information in the development of social understanding. *Frontiers in Psychology*, *8*.

Moore, E., Schaefer, R. S., Bastin, M. E., Roberts, N., & Overy, K. (2017). Diffusion tensor MRI tractography reveals increased fractional anisotropy (FA) in arcuate fasciculus following music-cued motor training. *Brain and Cognition*, *116*, 40–46.

Moore, F. R. (1988). Dysfunctions of MIDI. *Computer Music Journal*, *12*(1), 19–28.

Moran, N. (2014). Social implications arise in embodied music cognition research which can counter musicological "individualism." *Frontiers in Psychology*, *5*.

Moran, S., Jäger, N., Schnädelbach, H., & Glover, K. (2016). ExoPranayama: a biofeedback-driven actuated environment for supporting yoga breathing practices. *Personal and Ubiquitous Computing*, *20*(2), 261–275.

Moraveji, N., Ben, O., Nguyen, T., Saadat, M., Khalighi, Y., Pea, R., & Heer, J. (2011). Peripheral paced respiration: Influencing user physiology during information work. In J. Pierce (Ed.), *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (pp. 423–427).

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, *19*(2), 98–100.

Mörtl, A., Lorenz, T., & Hirche, S. (2014). Rhythm patterns interaction - Synchronization behavior for human-robot joint action. *PLoS ONE*, *9*(4), e95195.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2), e89642.

Murgia, M., Santoro, I., Tamburini, G., Prpic, V., Sors, F., Galmonte, A., & Agostini, T. (2016). Ecological sounds affect breath duration more than artificial sounds. *Psychological Research*, *80*(1), 76–81.

Murphy, E., & King, E. A. (2016). Testing the accuracy of smartphones and sound level meter applications for measuring environmental noise. *Applied Acoustics*, *106*, 16–22.

Nabi, D. A., & Charlton, J. P. (2014). The psychology of addiction to virtual environments: The allure of the virtual self. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality* (pp. 187–204). Oxford University Press.

Naef, A. C., Jeitziner, M. M., Knobel, S. E. J., Exl, M. T., Müri, R. M., Jakob, S. M., Nef, T., & Gerber, S. M. (2022). Investigating the role of auditory and visual sensory inputs for inducing relaxation during virtual reality stimulation. *Scientific Reports*, *12*(1), 1–11.

Nakagawa, R., Komatsubara, R., Ota, T., & Ohmura, H. (2018). Air maestros: A multi-user audiovisual experience using MR. *Proceedings of the Symposium on Spatial User Interaction*, 168.

Naveda, L., & Leman, M. (2010). The spatiotemporal representation of dance and music gestures using topological gesture analysis (TGA). *Music Perception*, *28*(1), 93–111.

Nelson, B. W., & Allen, N. B. (2019). Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: Intraindividual validation study. *JMIR MHealth and UHealth*, *7*(3), e10828.

Newen, A., Gallagher, S., & De Bruin, L. (2018). 4E Cognition: Historical Roots, Key Concepts, and Central Issues. In *The Oxford Handbook of 4E Cognition* (pp. 1–16). Oxford University Press.

Neyret, S., Navarro, X., Beacco, A., Oliva, R., Bourdin, P., Valenzuela, J., Barberia, I., & Slater, M. (2020). An Embodied Perspective as a Victim of Sexual Harassment in Virtual Reality Reduces Action Conformity in a Later Milgram Obedience Scenario. *Scientific Reports*, *10*(1), 1–18.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, *9*(3), 184–211.

Nijs, L., Coussement, P., Moens, B., Amelinck, D., Lesaffre, M., & Leman, M. (2012). Interacting with the music paint machine: Relating the constructs of flow experience and presence. *Interacting with Computers*, *24*(4), 237–250.

Nijs, L., & Leman, M. (2014). Interactive technologies in the instrumental music classroom: A longitudinal study with the Music Paint Machine. *Computers and Education*, *73*, 40–59.

Nilsson, U. (2011). Music: A nursing intervention. *European Journal of Cardiovascular Nursing*, *10*(2), 73–74.

Noel, J. P., Pfeiffer, C., Blanke, O., & Serino, A. (2015). Peripersonal space as the space of the bodily self. *Cognition*, *144*, 49–57.

Nordahl, R., & Nilsson, N. C. (2014). The Sound of Being There: Presence and Interactive Audio in Immersive Virtual Reality. In *The Oxford Handbook of Interactive Audio* (pp. 213–233). Oxford University Press.

Normand, J. M., Giannopoulos, E., Spanlang, B., & Slater, M. (2011). Multisensory stimulation can induce an illusion of larger belly size in immersive virtual reality. *PLoS ONE*, *6*(1), e16128.

Ochs, C., & Sonderegger, A. (2022). The Interplay Between Presence and Learning. *Frontiers in Virtual Reality*, *3*.

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers Robotics AI*, *5*.

Onderdijk, K. E., Swarbrick, D., Van Kerrebroeck, B., Mantei, M., Vuoskoski, J. K., Maes, P. J., & Leman, M. (2021). Livestream Experiments: The Role of COVID-19, Agency, Presence, and Social Context in Facilitating Social Connectedness. *Frontiers in Psychology*, *12*.

Orman, E. K. (2004). Effect of virtual reality graded exposure on anxiety levels of performing musicians: A case study. *Journal of Music Therapy*, *41*(1), 70–78.

Orman, E. K., Price, H. E., & Russell, C. R. (2017). Feasibility of Using an Augmented Immersive Virtual Reality Learning Environment to Enhance Music Conducting Skills. *Journal of Music Teacher Education*, *27*(1), 24–35.

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, *26*(5), 489–504.

Oxman, N. (2016). Age of Entanglement. *Journal of Design and Science*.

Pai, Y. S., Hajika, R., Gupta, K., Sasikumar, P., & Billinghurst, M. (2020). NeuralDrum: Perceiving Brain Synchronicity in XR Drumming. *SIGGRAPH Asia 2020 Technical Communications*, 1–4.

Palmer, C., Spidle, F., Koopmans, E., & Schubert, P. (2019). Ears, heads, and eyes: When singers synchronise. *Quarterly Journal of Experimental Psychology*, *72*(9), 2272–2287.

Pan, X., & Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417.

Paradiso, J. A., & Landay, J. A. (2009). Guest Editors' Introduction: Cross reality environments. *IEEE Pervasive Computing*, *8*(3), 14–15.

Parker, E. B., Short, J., Williams, E., & Christie, B. (1978). The Social Psychology of Telecommunications. *Contemporary Sociology*, *7*(1), 32.

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*.

Parsons, T. D., Gaggioli, A., & Riva, G. (2017). Virtual reality for research in social neuroscience. *Brain Sciences*, *7*(4), 42.

Parsons, T. D., Riva, G., Parsons, S., Mantovani, F., Newbutt, N., Lin, L., Venturini, E., & Hall, T. (2017). Virtual reality in pediatric psychology. *Pediatrics*, *140*, S86–S91.

Peng, C. K., Buldyrev, S. V, Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, *49*(2), 1685–1689.

Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: Perceptual illusion of body swapping. *PLoS ONE*, *3*(12), e3832.

Pickering, A. (2014). The Mangle of Practice. In R. C. Scharff & V. Dusek (Eds.), *Philosophy of technology: The technological condition* (pp. 260–265). John Wiley & Sons.

Pike, A. (1967). The Phenomenological Analysis and Description of Musical Experience. *Journal of Research in Music Education*, *15*(4), 316–319.

Pitman, D. (2021). Sounds in Space: A portfolio of spatial music works for VR. In *Doctoral dissertation*. https://hdl.handle.net/2440/135218

Prpa, M., Tatar, K., Schiphorst, T., & Pasquier, P. (2018). Respire: A breath away from the experience in virtual environment. In R. Mandryk & M. Hancock (Eds.), *Conference on Human Factors in Computing Systems - Proceedings* (pp. 1–6).

Pysiewicz, A., & Weinzierl, S. (2016). Instruments for spatial sound control in real time music performances. A review. In T. Bovermann, A. de Campo, H. Egermann, S.-I. Hardjowirogo, & S. Weinzierl (Eds.), *Musical Instruments in the 21st Century: Identities, Configurations, Practices* (pp. 273–296). Springer.

Randles, C. (2012). Phenomenology: A review of the literature. *Update: Applications of Research in Music Education*, *30*(2), 11–21.

Rankin, S. K., & Large, E. W. (2009). Fractal tempo fluctuation and pulse prediction. *Music Perception*, *26*(5), 401–413.

Ravignani, A., & Norton, P. (2017). Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations. *Journal of Language Evolution*, *2*(1), 4–19.

Razafsha, M., Behforuzi, H., Azari, H., Zhang, Z., Wang, K. K., Kobeissy, F. H., & Gold, M. S. (2012). Qualitative versus quantitative methods in psychiatric research. *Methods in Molecular Biology*, *829*, 49–62.

Reich, S. (2002). *Writings on music, 1965-2000*. Oxford University Press.

Reidsma, D., Radha, M., & Nijholt, A. (2014). Mediated interactions and musical expression - A survey. In *Digital Da Vinci: Computers in Music* (pp. 79–98). Springer.

Repp, B. H. (2012). Musical synchronization. In *Music, Motor Control and the Brain* (pp. 55–76).

Oxford University Press.

Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: A review of recent research (2006-2012). *Psychonomic Bulletin and Review*, *20*(3), 403–452.

Reyes-Lecuona, A., Bouchara, T., & Picinali, L. (2022). Immersive Sound for XR. In *Roadmapping Extended Reality* (pp. 75–102). Wiley Online Library.

Reza Namazi, H. (2017). Fractal-based analysis of the influence of music on human respiration. *Fractals*, *25*(6), 1750059.

Richardson, M. J., Harrison, S. J., Kallen, R. W., Walton, A., Eiler, B. A., Saltzman, E., & Schmidt, R. C. (2015). Self-organized complementary joint action: Behavioral dynamics of an interpersonal collision-avoidance task. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 665.

Riva, G. (2006). Being-in-the-world-with: Presence meets social and cognitive neuroscience. In *From communication to presence: Cognition, emotions and culture towards the ultimate communicative experience* (pp. 47–80). IOS Press.

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., & Alcañiz, M. (2007). Affective interactions using virtual reality: The link between presence and emotions. *Cyberpsychology and Behavior*, *10*(1), 45–56.

Riva, G., Waterworth, J. A., & Waterworth, E. L. (2004). The layers of presence: A bio-cultural approach to understanding presence in natural and mediated environments. *Cyberpsychology and Behavior*, *7*(4), 402–416.

Riva, G., Waterworth, J. A., Waterworth, E. L., & Mantovani, F. (2011). From intention to action: The role of presence. *New Ideas in Psychology*, *29*(1), 24–37.

Riva, G., Wiederhold, B. K., & Mantovani, F. (2019). Neuroscience of Virtual Reality: From Virtual Exposure to Embodied Medicine. *Cyberpsychology, Behavior, and Social Networking*, *22*(1), 82–96.

Roholt, T. C. (2013). In Praise of Ambiguity: Musical Subtlety and Merleau-Ponty. *Contemporary Aesthetics*, *11*(1), 19.

Rose, D., Ott, L., Guérin, S. M. R., Annett, L. E., Lovatt, P., & Delevoye-Turrell, Y. N. (2021). A general procedure to measure the pacing of body movements timed to music and metronome in younger and older adults. *Scientific Reports*, *11*(1), 1–16.

Rosso, M., Maes, P. J., & Leman, M. (2021). Modality-specific attractor dynamics in dyadic entrainment. *Scientific Reports*, *11*(1), 1–13.

Rottondi, C., Chafe, C., Allocchio, C., & Sarti, A. (2016). An overview on networked music performance technologies. *IEEE Access*, *4*, 8823–8843.

Ruan, J., & Xie, D. (2021). Networked vr: State of the art, solutions, and challenges. *Electronics*, *10*(2), 1–18.

Sale, J. E. M., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and Quantity*, *36*(1), 43–53.

Salselas, I., Penha, R., & Bernardes, G. (2021). Sound design inducing attention in the context of audiovisual immersive environments. *Personal and Ubiquitous Computing*, *25*(4), 737–748.

Sanchez-Vives, M. V, & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, *6*(4), 332–339.

Sänger, J., Müller, V., & Lindenberger, U. (2013). Directionality in hyperbrain networks discriminates between leaders and followers in guitar duets. *Frontiers in Human*

Neuroscience, *7*.

Santini, G. (2019). Composing space in the space: An Augmented and Virtual Reality sound spatialization system. In I. Barbancho, L. J. Tardón, A. Peinado, & A. M. Barbancho (Eds.), *Proceedings of the Sound and Music Computing Conferences* (pp. 229–233).

Scarfe, P., & Glennerster, A. (2019). The Science behind Virtual Reality Displays. *Annual Review of Vision Science*, *5*(18), 529–547.

Schacher, J. C., & Kocher, P. (2006). Ambisonics spatialization tools for max/MSP. In G. Essl & T. H. Park (Eds.), *International Computer Music Conference* (Vol. 500, Issue 1, pp. 274–277).

Schäfer, T., Fachner, J., & Smukalla, M. (2013). Changes in the representation of space and time while listening to music. *Frontiers in Psychology*, *4*.

Scharff, R. C., & Dusek, V. (2014). *Philosophy of technology: The technological condition: An anthology*. John Wiley & Sons.

Schiavio, A., & De Jaegher, H. (2017). Participatory sense-making in joint musical practice. In *The Routledge Companion to Embodied Music Interaction* (pp. 31–39). Routledge.

Schiavio, A., Menin, D., & Matyja, J. (2014). Music in the flesh: Embodied simulation in musical understanding. *Psychomusicology: Music, Mind, and Brain*, *24*(4), 340–343.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414.

Schlagowski, R., Gupta, K., Mertes, S., Billinghurst, M., Metzner, S., & Andre, E. (2022). Jamming in MR: Towards Real-Time Music Collaboration in Mixed Reality. *IEEE Conference on Virtual Reality and 3D User Interfaces*, 854–855.

Schmele, T., & Finney, N. (2011). Exploring 3D Audio as a New Musical Language [Universitat Pompeu Fabra]. In *Master thesis*. http://mtg.upf.edu/system/files/publications/Schmele-Timothy-Master-thesis-2011.pdf

Schmidt, P., Reiss, A., Dürichen, R., & Laerhoven, K. Van. (2019). Wearable-based affect recognition—a review. *Sensors*, *19*(19), 4079.

Schultz, B. G., & Palmer, C. (2019). The roles of musical expertise and sensory feedback in beat keeping and joint action. *Psychological Research*, *83*(3), 419–431.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76.

Sebanz, N., & Knoblich, G. (2009). Prediction in Joint Action: What, When, and Where. *Topics in Cognitive Science*, *1*(2), 353–367.

Sebanz, N., & Knoblich, G. (2021). Progress in joint-action research. *Current Directions in Psychological Science*, *30*(2), 138–143.

Seinfeld, S., Arroyo-Palacios, J., Iruretagoyena, G., Hortensius, R., Zapata, L. E., Borland, D., de Gelder, B., Slater, M., & Sanchez-Vives, M. V. (2018). Offenders become the victim in virtual reality: impact of changing perspective in domestic violence. *Scientific Reports*, *8*(1), 1–11.

Selfridge, R., & Barthet, M. (2019). Augmented live music performance using mixed reality and emotion feedback. In M. Aramaki, O. Derrien, R. Kronland-Martinet, & Ystad (Eds.), *14th Int. Symposium on Computer Music Multidisciplinary Research* (p. 210).

Şenel, G., & Slater, M. (2020). Conversation with Your Future Self About Nicotine Dependence. In P. Bourdot, V. Interrante, R. Kopper, A.-H. Olivier, H. Saito, & G. Zachmann (Eds.), *Virtual Reality and Augmented Reality: 17th EuroVR International Conference* (pp. 216–223).

Serafin, S., Adjorlu, A., Nilsson, N., Thomsen, L., & Nordahl, R. (2017). Considerations on the use of virtual and augmented reality technologies in music education. In M. M. O. da Silva, S. Cutchin, E. Southgate, & J. A. Fails (Eds.), *IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual and Augmented Reality* (pp. 1–4).

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

Serafin, S., Franinović, K., Hermann, T., Lemaitre, G., Rinott, M., & Rocchesso, D. (2011). Sonic interaction design. In T. Hermann, H. Andy, & J. G. Neuhoff (Eds.), *The sonification handbook* (pp. 87–110). Logos Publishing House.

Servotte, J. C., Goosse, M., Campbell, S. H., Dardenne, N., Pilote, B., Simoneau, I. L., Guillaume, M., Bragard, I., & Ghuysen, A. (2020). Virtual Reality Experience: Immersion, Sense of Presence, and Cybersickness. *Clinical Simulation in Nursing*, *38*, 35–43.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573.

Seth, A. K., & Tsakiris, M. (2018). Being a Beast Machine: The Somatic Basis of Selfhood. *Trends in Cognitive Sciences*, *22*(11), 969–981.

Shahal, S., Wurzberg, A., Sibony, I., Duadi, H., Shniderman, E., Weymouth, D., Davidson, N., & Fridman, M. (2020). Synchronization of complex human networks. *Nature Communications*, *11*(1), 1–10.

Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, *7*(3), 269–284.

Sheridan, T. B. (1992). Musings on Telepresence and Virtual Presence. *Presence: Teleoperators and Virtual Environments*, *1*(1), 120–126.

Shin, M., Song, S. W., Kim, S. J., & Biocca, F. (2019). The effects of 3D sound in a 360-degree live concert video on social presence, parasocial interaction, enjoyment, and intent of financial supportive action. *International Journal of Human Computer Studies*, *126*, 81–93.

Shirzadian, N., Redi, J. A., Röggla, T., Panza, A., Nack, F., & Cesar, P. (2017). Immersion and togetherness: How live visualization of audience engagement can enhance music events. In A. D. Cheok, M. Inami, & T. Romao (Eds.), *Advances in Computer Entertainment Technology: 14th International Conference* (pp. 488–507).

Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(1), 70–75.

Siler, T. (2011). The artscience program for realizing human potential. *Leonardo*, *44*(5), 417–424.

Six, J. (2018). *Engineering systematic musicology: Methods and services for computational and empirical music research*. Ghent University.

Skarbez, R., Brooks, F. P., & Whitton, M. C. (2017). A survey of presence and related concepts. *ACM Computing Surveys*, *50*(6), 1–39.

Slater, M., Lotto, B., Arnold, M. M., & Sanchez-Vives, M. V. (2009). How we experience immersive virtual environments: The concept of presence and its measurement. *Anuario de Psicologia*, *40*(2), 193–210.

Slater, M., Perez-Marcos, D., Ehrsson, H. H., & Sanchez-Vives, M. V. (2009). Inducing illusory ownership of a virtual body. *Frontiers in Neuroscience*, *3*.

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality.

Frontiers Robotics AI, *3*.

Slater, M., Spanlang, B., Sanchez-Vives, M. V, & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS ONE*, *5*(5), e10564.

Slater, M., & Steed, A. (2000). A virtual presence counter. *Presence: Teleoperators and Virtual Environments*, *9*(5), 413–434.

Slepian, M. L., & Ambady, N. (2012). Fluid movement and creativity. *Journal of Experimental Psychology: General*, *141*(4), 625–629.

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category-ELAN and ISO DCR. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, & D. Tapias (Eds.), *6th international Conference on Language Resources and Evaluation*.

Small, C. (1999). Musicking — the meanings of performing and listening. A lecture. *Music Education Research*, *1*(1), 9–22.

Smith, D. W. (2008). Phenomenology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Smith, F. J. (1995). Toward a Phenomenology of Music : A Musician ' s Composition Journal Stable. *Philosophy of Music Education Review*, *3*(1), 21–33.

Smith, J. C. (2001). Advances in ABC Relaxation: Applications and inventories. In *Advances in ABC Relaxation: Applications and inventories*. Springer.

Spampinato, F. M., & Carticalà, V. (2021). Contemporary Art and Virtual Reality: New Conditions of Viewership. *Cinergie--Il Cinema E Le Altre Arti*, *19*, 121–133.

Speicher, M., Hall, B. D., & Nebeling, M. (2019). What is mixed reality? In S. Brewster & G. Fitzpatrick (Eds.), *Conference on Human Factors in Computing Systems - Proceedings* (pp. 1–15).

Spence, C., Lee, J., & Van der Stoep, N. (2020). Responding to sounds from unseen locations: crossmodal attentional orienting in response to sounds presented from the rear. *European Journal of Neuroscience*, *51*(5), 1137–1150.

Spence, J., Frohlich, D. M., & Andrews, S. (2013). Performative Experience Design. In W. E. Mackay (Ed.), *Conference on Human Factors in Computing Systems - Proceedings* (pp. 2049–2058).

Steed, A. (2014). Recreating Visual Reality in Virtuality. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality* (pp. 420–443). Oxford University Press.

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, *18*(8), 1380–1393.

Stiegler, B. (2019). *The age of disruption: Technology and madness in computational capitalism*. Polity Press.

Stupacher, J. (2019). The experience of flow during sensorimotor synchronization to musical rhythms. *Musicae Scientiae*, *23*(3), 348–361.

Stupacher, J., Maes, P. J., Witte, M., & Wood, G. (2017). Music strengthens prosocial effects of interpersonal synchronization – If you move in time with the beat. *Journal of Experimental Social Psychology*, *72*, 39–44.

Sutherland, I. E. (1968). A head-mounted three dimensional display. *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference*, *33*, 757–764.

Suzuki, K., Wakisaka, S., & Fujii, N. (2012). Substitutional reality system: a novel experimental platform for experiencing alternative reality. *Scientific Reports*, *2*, 459.

Svare, H. (2006). The Embodied Mind. In *Studies in German Idealism* (Vol. 6, pp. 11–60). Springer.

Tajadura-Jiménez, A., Pantelidou, G., Rebacz, P., Västfjäll, D., & Tsakiris, M. (2011). I-space: The effects of emotional valence and source of music on interpersonal distance. *PLoS ONE*, *6*(10), e26083.

Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence* (pp. 366–381). Springer.

Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantzf, L., & De Gelder, B. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(42), 17661–17666.

Tanaka, A. (2019). Embodied Musical Interaction. In *New Directions in Music and Human-Computer Interaction* (pp. 135–154). Springer.

Tarr, B., Launay, J., & Dunbar, R. I. M. (2014). Music and social bonding: "Self-other" merging and neurohormonal mechanisms. *Frontiers in Psychology*, *5*.

Tarr, B., Slater, M., & Cohen, E. (2018). Synchrony and social connection in immersive Virtual Reality. *Scientific Reports*, *8*(1), 1–8.

Taylor, J. (1997). The emerging geographies of virtual worlds. *Geographical Review*, *87*(2), 172–192.

Taylor, S., Landry, C. A., Paluszek, M. M., Fergus, T. A., McKay, D., & Asmundson, G. J. G. (2020). Development and initial validation of the COVID Stress Scales. *Journal of Anxiety Disorders*, *72*, 102232.

Teo, W. P., Muthalib, M., Yamin, S., Hendy, A. M., Bramstedt, K., Kotsopoulos, E., Perrey, S., & Ayaz, H. (2016). Does a combination of virtual reality, neuromodulation and neuroimaging provide a comprehensive platform for neurorehabilitation? - A narrative review of the literature. *Frontiers in Human Neuroscience*, *10*.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, *24*(1), 1–34.

Thielen, E., Letellier, J., Sieck, J., & Thoma, A. (2018). Bringing a Virtual String Quartet to Life. In H. Weinschiers-Theopilus, I. van Zyl, N. Goagoses, D. Singh Jat, E. G. Belay, R. Orji, & A. Peters (Eds.), *ACM International Conference Proceeding Series* (pp. 208–211).

Tognoli, E., Zhang, M., Fuchs, A., Beetle, C., & Kelso, J. A. S. (2020). Coordination Dynamics: A Foundation for Understanding Social Behavior. *Frontiers in Human Neuroscience*, *14*.

Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. *Music Perception*, *28*(1), 59–70.

Torales, J., O'Higgins, M., Castaldelli-Maia, J. M., & Ventriglio, A. (2020). The outbreak of COVID-19 coronavirus and its impact on global mental health. *International Journal of Social Psychiatry*, *66*(4), 317–320.

Tress, B., Tress, G., & Fry, G. (2007). Defining concepts and process of knowledge production in integrative research. In *From landscape research to landscape planning* (Vol. 12, pp. 13–26). Springer.

Trommer, M. (2020). Points Further North: An acoustemological cartography of non-place. *Journal of New Music Research*, *49*(1), 73–87.

Trujano, F., Khan, M., & Maes, P. (2018). Arpiano efficient music learning using augmented reality. In T.-T. Wu, Y.-M. Huang, R. Shadiev, L. Lin, & A. I. Starčič (Eds.), *Innovative Technologies and Learning: First International Conference* (pp. 3–17).

Tsioutas, K., & Xylomenos, G. (2021). On the Impact of Audio Characteristics to the Quality of Musicians' Experience in Network Music Performance. *Journal of the Audio Engineering Society*, *69*(12), 914–923.

Turchet, L., Fischione, C., Essl, G., Keller, D., & Barthet, M. (2018). Internet of Musical Things: Vision and Challenges. *IEEE Access*, *6*, 61994–62017.

Turchet, L., Hamilton, R., & Çamcı, A. (2021). Music in Extended Realities. *IEEE Access*, *9*, 15810–15832.

Turchet, L., & Rottondi, C. (2022). On the relation between the fields of Networked Music Performances, Ubiquitous Music, and Internet of Musical Things. *Personal and Ubiquitous Computing*, 1–10.

Van Den Haak, M. J., De Jong, M. D. T., & Schellens, P. J. (2003). Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour and Information Technology*, *22*(5), 339–351.

van der Schyff, D., Schiavio, A., Walton, A., Velardo, V., & Chemero, A. (2018). Musical creativity and the embodied mind: Exploring the possibilities of 4E cognition and dynamical systems theory. *Music and Science*, *1*, 1–18.

Van Diest, I., Verstappen, K., Aubert, A. E., Widjaja, D., Vansteenwegen, D., & Vlemincx, E. (2014). Inhalation/Exhalation Ratio Modulates the Effect of Slow Breathing on Heart Rate Variability and Relaxation. *Applied Psychophysiology Biofeedback*, *39*(3–4), 171–180.

Van Dixhoorn, J. (1998). Cardiorespiratory effects of breathing and relaxation instruction in myocardial infarction patients. *Biological Psychology*, *49*(1–2), 123–135.

Van Doorn, P., Folgering, H., & Colla, P. (1982). Control of the end-tidal P(CO2) in the hyperventilation syndrome: Effects of biofeedback and breathing instructions compared. *Clinical Respiratory Physiology*, *18*(6), 829–836.

van Dyck, E., Six, J., Soyer, E., Denys, M., Bardijn, I., & Leman, M. (2017). Adopting a music-to-heart rate alignment strategy to measure the impact of music and its tempo on human heart rate. *Musicae Scientiae*, *21*(4), 390–404.

Van Kerrebroeck, B., Caruso, G., & Maes, P. J. (2021). A Methodological Framework for Assessing Social Presence in Music Interactions in Virtual Reality. *Frontiers in Psychology*, *12*.

van Noorden, L., & Moelants, D. (1999). Resonance in the perception of musical pulse. *International Journal of Phytoremediation*, *21*(1), 43–66.

Vandenberg, F., Berghman, M., & Schaap, J. (2021). The 'lonely raver': music livestreams during COVID-19 as a hotline to collective consciousness? *European Societies*, *23*(S1), S141–S152.

Varni, G., Dubus, G., Oksanen, S., Volpe, G., Fabiani, M., Bresin, R., Kleimola, J., Välimäki, V., & Camurri, A. (2012). Interactive sonification of synchronisation of motoric behaviour in social active listening to music with mobile devices. *Journal on Multimodal User Interfaces*, *5*(3–4), 157–173.

Västfjäll, D. (2003). The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. *Cyberpsychology and Behavior*, *6*(2), 181–188.

Verbeek, P.-P. (2014). Postphenomenology of Technology. In R. C. Scharff & V. Dusek (Eds.), *Philosophy of technology: The technological condition* (pp. 561–572). John Wiley & Sons.

Verberne, F. M. F., Ham, J., Ponnada, A., & Midden, C. J. H. (2013). Trusting digital chameleons: The effect of mimicry by a virtual social agent on user trust. In S. Berkovsky & J. Freyne (Eds.), *Persuasive Technology: 8th International Conference* (pp. 234–245).

Vuoskoski, J. K., Thompson, M. R., Spence, C., & Clarke, E. F. (2016). Interaction of sight and sound in the perception and experience of musical performance. *Music Perception*, *33*(4), 457–471.

Waddell, G., Perkins, R., & Williamon, A. (2019). The evaluation simulator: A new approach to training music performance assessment. *Frontiers in Psychology*, *10*.

Walton, A. E., Richardson, M. J., Langland-Hassan, P., & Chemero, A. (2015). Improvisation and the self-organization of multiple musical bodies. *Frontiers in Psychology*, *6*.

Walton, A. E., Washburn, A., Langland-Hassan, P., Chemero, A., Kloos, H., & Richardson, M. J. (2018). Creating Time: Social Collaboration in Music Improvisation. *Topics in Cognitive Science*, *10*(1), 95–119.

Walton, A., Washburn, A., Richardson, M. J., & Chemero, A. (2016). Empathy and groove in musical movement. In S. Penny & K. Donahey (Eds.), *Proceedings of A Body of Knowledge - Embodied Cognition and the Arts conference*.

Warren, W. H. (2006). The dynamics of perception and action. *Psychological Review*, *113*(2), 358–389.

Waterworth, J. A., & Waterworth, E. L. (2014). Distributed Embodiment: Real Presence in Virtual Bodies. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality* (pp. 589–601). Oxford University Press.

Wernli, D., Darbellay, F., & Universities., L. of E. R. (2016). *Interdisciplinarity and the 21 st century university* (Issue November). League of European Research Universities.

Williamon, A., Aufegger, L., & Eiholzer, H. (2014). Simulating and stimulating performance: Introducing distributed simulation to enhance musical learning and performance. *Frontiers in Psychology*, *5*.

Williamon, A., & Davidson, J. W. (2002). Exploring co-performer communication. *Musicae Scientiae*, *6*(1), 53–72.

Witmer, B. G., Jerome, C. J., & Singer, M. J. (2005). The factor structure of the Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, *14*(3), 298–312.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240.

Wojciechowski, R., & Cellary, W. (2013). Evaluation of learners' attitude toward learning in ARIES augmented reality environments. *Computers and Education*, *68*, 570–585.

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics*, *79*(7), 2064–2072.

Wozniewski, M., Settel, Z., & Cooperstock, J. R. (2018). Sonic Interaction via Spatial Arrangement in Mixed-Reality Environments. In *Sonic Interaction Design* (p. 329). MIT Press.

Yang, L., Huang, T., Hong-An, W., & Guo-Zhong, D. a. (2019). Gesture interaction in virtual reality. *Virtual Reality and Intelligent Hardware*, *1*(1), 84–112.

Zabel, G. (2014). Through the Looking Glass: Philosophical Reflections on the Art of Virtual Worlds. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality* (pp. 407–419). Oxford University Press.

Zahavi, D. (2001). Beyond empathy: Phenomenological approaches to intersubjectivity. *Journal of Consciousness Studies*, *8*(6–7), 151–167.

Zelli, B. (2009). Spatialization as a musical concept. In G. Klinker, H. Saito, & T. Höllerer (Eds.), *IEEE 2009 International Symposium on Mixed and Augmented Reality* (pp. 35–38).

Zhao, Q., Meng, M., Kumar, R., Wu, Y., Huang, J., Lian, N., Deng, Y., & Lin, S. (2020). The impact of COPD and smoking history on the severity of COVID-19: A systemic review and meta-analysis. *Journal of Medical Virology*, *92*(10), 1915–1921.

Zibrek, K., Martin, S., & McDonnell, R. (2019). Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Transactions on Applied Perception*, *16*(3), 1–19.

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). An analysis and critique of research through design: Towards a formalization of a research approach. In O. W. Bertelsen & P. Krogh (Eds.), *Proceedings of the 8th ACM Conference on Designing Interactive Systems* (pp. 310–319).