



A Methodological Framework for Assessing Social Presence in Music Interactions in Virtual Reality

Bavo Van Kerrebroeck*, Giusy Caruso and Pieter-Jan Maes

Department of Art, Music, and Theatre Sciences, IPEM, Ghent University, Ghent, Belgium

Virtual reality (VR) brings radical new possibilities to the empirical study of social music cognition and interaction. In the present article, we consider the role of VR as a research tool, based on its potential to create a sense of “social presence”: the illusory feeling of being, and socially interacting, inside a virtual environment. This makes VR promising for bridging ecological validity (“research in the wild”) and experimental control (“research in the lab”) in empirical music research. A critical assumption however is the actual ability of VR to simulate real-life social interactions, either *via* human-embodied avatars or computer-controlled agents. The mediation of social musical interactions *via* VR is particularly challenging due to their embodied, complex, and emotionally delicate nature. In this article, we introduce a methodological framework to operationalize social presence by a combination of factors across interrelated layers, relating to the performance output, embodied co-regulation, and subjective experiences. This framework provides the basis for the proposal of a pragmatic approach to determine the level of social presence in virtual musical interactions, by comparing the outcomes across the multiple layers with the outcomes of corresponding real-life musical interactions. We applied and tested this pragmatic approach *via* a case-study of piano duet performances of the piece Piano Phase composed by Steve Reich. This case-study indicated that a piano duet performed in VR, in which the real-time interaction between pianists is mediated by embodied avatars, might lead to a strong feeling of social presence, as reflected in the measures of performance output, embodied co-regulation, and subjective experience. In contrast, although a piano duet in VR between an actual pianist and a computer-controlled agent led to a relatively successful performance output, it was inadequate in terms of both embodied co-regulation and subjective experience.

OPEN ACCESS

Edited by:

Steve Richard DiPaola,
Simon Fraser University, Canada

Reviewed by:

Francisco Rebelo,
University of Lisbon, Portugal
Eriko Aiba,
The University of Electro-
Communications, Japan

*Correspondence:

Bavo Van Kerrebroeck
bavo.vankerebroeck@ugent.be

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 03 February 2021

Accepted: 28 April 2021

Published: 11 June 2021

Citation:

Van Kerrebroeck B, Caruso G and
Maes P-J (2021) A Methodological
Framework for Assessing Social
Presence in Music Interactions in
Virtual Reality.
Front. Psychol. 12:663725.
doi: 10.3389/fpsyg.2021.663725

Keywords: presence, virtual reality, music, embodiment, social interaction, interpersonal coordination, ecologically valid research

INTRODUCTION

Virtual reality (VR) encompasses a plethora of technologies to create new environments, or simulate existing ones, *via* computer-generated multisensory displays (Sutherland, 1968; Taylor, 1997; Scarfe and Glennerster, 2019). 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 provide hence a technological mediation between performed actions and multisensory perceptions, extending the natural sensorimotor capacities of humans into the digital world (Kornelsen, 1991; Biocca and Delaney, 1995; Ijsselstein and Riva, 2003). 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 and 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 (Minsky, 1980; Sheridan, 1992; Slater and 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 to create the illusory feeling for a user of owning, controlling, and being inside another body than its physical one (self-presence) (Kilteni et al., 2012; De Oliveira et al., 2016; Braun et al., 2018; Matamala-Gomez et al., 2020). 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 (Short et al., 1976; Garau et al., 2005; Parsons et al., 2017; Oh et al., 2018).

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. In the first part of the article, we will discuss in more detail how a VR-based approach has its roots in earlier, human-centered 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 skillful co-regulation of bodily articulations (Leman, 2008; Leman and 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 signaling communicative cues, related to music-structural aspects and emotional expression (Williamson and Davidson, 2002; Goebel and Palmer, 2009; Keller and Appel, 2010; Coorevits et al., 2020). 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 to animate 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-modeling and simulation of human behavior (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 behavioral 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. 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.

VR: A RESEARCH TOOL FOR STUDYING SOCIAL INTERACTIONS

Around the 21st century, VR started to develop as a valuable methodological tool in human-centered research, including the social and cognitive (neuro)sciences (Biocca, 1992; Biocca and Delaney, 1995; Blascovich et al., 2002; Fox et al., 2009; Parsons, 2015; Slater and Sanchez-Vives, 2016; Parsons et al., 2017; Pan and de Hamilton, 2018), philosophy (Metzinger, 2018), the humanities (Cruz-Neira, 2003), product design (Berg and 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-centered 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. 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 (Parsons, 2015; Kothgassner and Felnhofer, 2020). 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 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 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 and 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-centered research, this ability of VR is seldomly employed as a form of mere escapism from the physical world. In contrast, 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 are 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 (Parsons et al., 2017; Metzinger, 2018). 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 and Ehrsson, 2008). It has the potential to influence different representational layers of the human

self-model (Metzinger, 2018) leading to phenomena such as virtual embodiment, (virtual) body swapping (Petkova and Ehrsson, 2008; De Oliveira et al., 2016) and increasingly frequent and complex “social hallucinations” (Metzinger, 2018).

Given these traits and their potential, the use of VR in music research has increased over the recent decade (Çamci and 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 (Orman, 2004; Bissonnette et al., 2016), music education (Orman et al., 2017; Serafin et al., 2017), music performance (Williamon et al., 2014; Glowinski et al., 2015), and the relation between sound and presence (Västfjäll, 2003; Kern and Ellermeier, 2020; Kobayashi et al., 2020). A good example of simulating a real-life setting is given by Glowinski et al. (2015), who investigated the influence of social context on performance. 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 toward new virtual instruments (Honigman et al., 2013; Berthaut et al., 2014; Serafin et al., 2016; Hamilton, 2019) to new interactions and the development of interaction design principles (Deacon et al., 2016; Atherton and Wang, 2020). 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 behavioral changes (Kilteni 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 to the quote by 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 are so immersed in it.” 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, and gender), environmental properties, the actual musical behavior, 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.

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, behavioral, 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, behavior, 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 behavior and responses in simulated VR contexts resemble behavior and responses in corresponding real-life musical contexts (Minsky, 1980; Slater et al., 2009; Johanson, 2015; Scarfe and Glennerster, 2019).

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 (Gallagher, 2001; De Jaegher and Di Paolo, 2007; Kiverstein, 2011; Froese and Fuchs, 2012; Gallotti and Frith, 2013; Schilbach et al., 2013; Fogel, 2017). Proponents of the interaction theory consider social cognition essentially as an embodied and participatory practice, emerging in real-time co-regulated interaction and 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 and 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 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 judgments of the performance output by the performers themselves.

Layer 2: Embodied Co-regulation

A successful musical output relies on a skillful, 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 continuously 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 (Sebanz and Knoblich, 2009; Gallagher and Allen, 2018; Koban et al., 2019). 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 time-series 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.

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 (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,

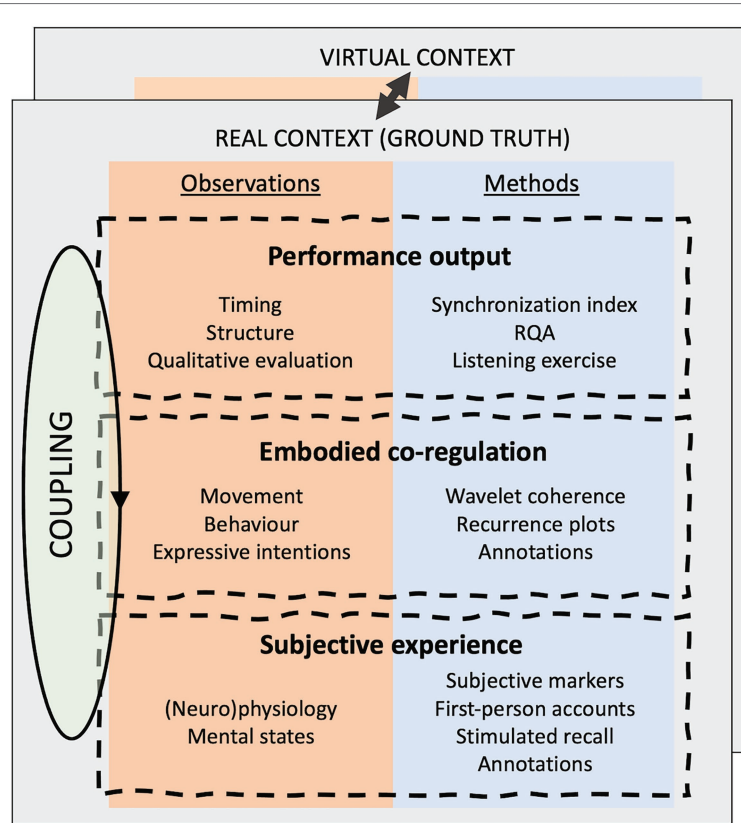


FIGURE 1 | Overview of the methodological framework to operationalize social presence in virtual reality (VR) music contexts. The core of the framework consists of a comparative analysis of a simulated virtual context, with the corresponding real-life music context (which functions as “ground truth”) across three interrelated layers; performance output, embodied co-regulation, and subjective experience. (RQA, recurrence quantification analysis)

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 (Witmer and Singer, 1998; Lessiter et al., 2001; Martin and Jackson, 2008). 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.

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, 2008). 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 distinguishes interactions using a specific operationalization of qualitative and quantitative methods in each layer. These operationalizations can vary from simple setups with for example, audio-, video-recordings, and annotations in the performance, co-regulation, or subjective layer to the more complex setups that will be presented in the case-study below. 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. Direct assessment is preferred over post-experimental reports to avoid the influence of self-referential cognitive processes or interfering in the interaction. Application of the

framework then allows to identify interactions with for example close coordination and intense subjective experiences but nevertheless 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.

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. The piece was chosen as 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.

Research Question

The case-study was meant to empirically evaluate and test our pragmatic approach for 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 (ground truth) performance of Piano Phase under normal, “real-life” conditions (see Design section). In all conditions, we captured an elaborated set of quantitative data related to the experience, behavior, 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.

Participants

The case-study involved three expert pianists: one test subject and two research confederates. The experimental protocol was reviewed and approved by the ethical commission of the University of Ghent. All pianists had over 10 years of professional music experience. The test subject (female, 38 years, in the following termed “Test pianist”) 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 “Confederate pianist 1”) functioned as research confederate in the first two conditions and had concert experiences in performing Piano Phase. Finally, the third pianist (female, 44 years, in the following termed “Confederate pianist 2”) was another research confederate in the third condition and co-author of this study. She had no experience in performing the piece but did have experience with VR.

Design

The experiment consisted of three conditions as presented in the schematic overview in **Figure 2**. Conditions are placed along a virtuality continuum represented by the arrow in the figure (Milgram et al., 1995) and correspond from left to right to an unmodelled, partially modeled, and fully modeled world. In each condition, the Test pianist performed Piano Phase together with a research confederate while wearing a Head-Mounted-Display (HMD) (Confederate pianist 1 in the first

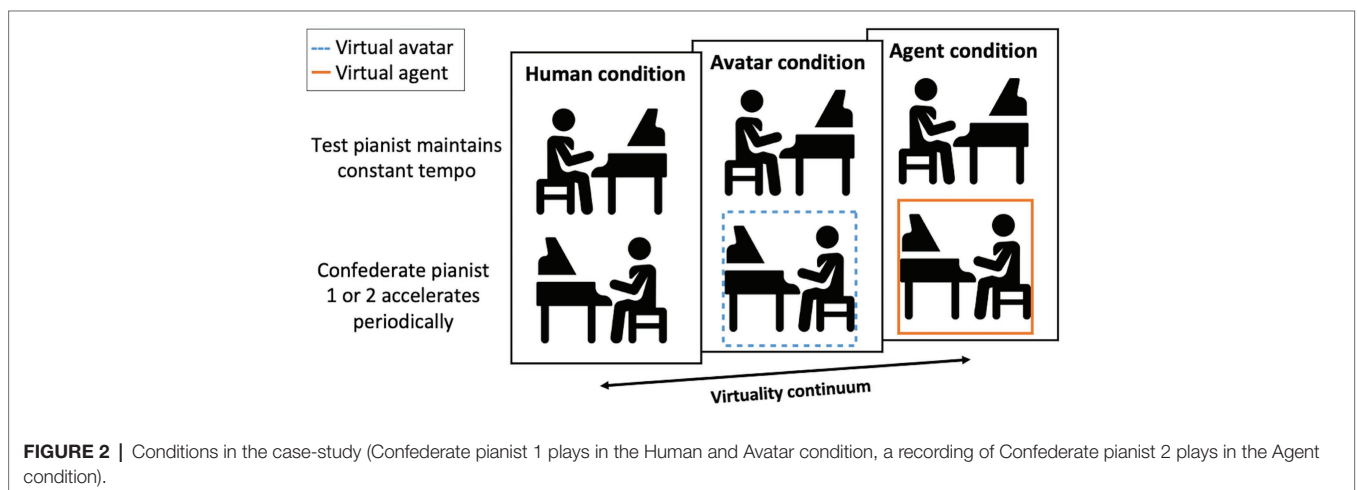


FIGURE 2 | Conditions in the case-study (Confederate pianist 1 plays in the Human and Avatar condition, a recording of Confederate pianist 2 plays in the Agent condition).

two conditions and Confederate pianist 2 in the third condition). The fundamental distinction between the three conditions was the level of behavioral realism of the confederate partner as perceived by the Test pianist:

1. Human condition (ground truth): the Test pianist and Confederate pianist 1 performed Piano Phase under normal, “real-life” concert conditions. The Test pianist visually perceived Confederate pianist 1 in a natural, physical manner. To match the two other performance conditions, we asked the Test pianist to wear a HMD with the pass-through camera activated. This was done 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 the Test pianist and Confederate pianist.
2. Avatar condition: the Test pianist and Confederate pianist 1 performed together in real-time, but the Test pianist visually perceived Confederate pianist 1 as a human-embodied virtual avatar. The Test pianist was disconnected from visual information coming from the physical environment, and all visual information related to the virtual room, piano keyboard, hands, and the co-performing Confederate pianist 1 was provided to the Test pianist *via* the HMD. Full body movements and musical instrument digital interface (MIDI) piano performance of the virtual avatar were streamed in real-time from the performance of Confederate pianist 1.
3. Agent condition: the Test pianist performed together with a computer-controlled virtual agent. The Test pianist was disconnected from all direct visual information as similar to the Avatar condition. Full body movements and MIDI piano performance of the virtual agent were rooted in pre-recorded time-series data of an actual pianist (Confederate pianist 2), 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 time-series data and the accompanying audiovisual VR animation to the real-time performance of the Test pianist. This allowed to accurately control the phase of the musical part of the virtual agent with respect to the Test pianist 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 the Avatar condition.

For more information on the display methods, see the Materials and Apparatus section below. We will use Human condition (HC), Avatar condition (AvC), and Agent condition (AgC) abbreviations in the Analysis and Results sections to refer to numerical results from the Human, Avatar, and Agent condition.

Materials and Apparatus

Piano Keyboards

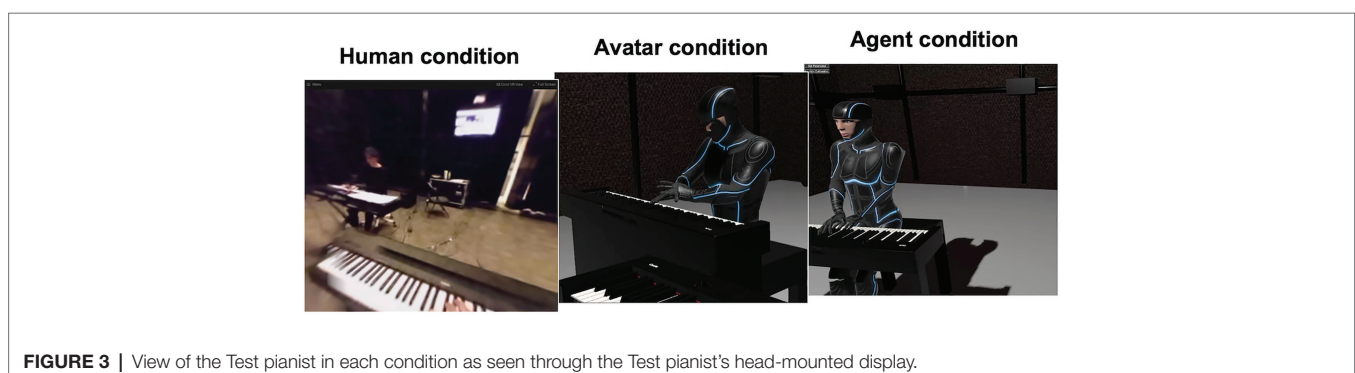
The piano keyboards used for the performances were digital interface MIDI controllers. The Test pianist and Confederate pianist 1 played on a Yamaha P60 and Confederate pianist 2 played on a Roland RD700SX. MIDI information was processed in Ableton Live 10 to generate piano sounds using a Native Instruments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

Performance Setting and Virtual Simulation Displays

The full experiment took place in the Art and Science Interaction Lab of IPeM, a 10 m-by-10 m-by-7 m (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 toward each other (see **Figure 3**). The avatar and agent were based on full body movement captures using the Qualisys system described below. Motion capture data were processed in Unity. An important consideration in the study was to also simulate the hands of the Test pianist as earlier research indicated that this may substantially increase the feeling of (self-)presence (Banakou and Slater, 2014). We used the Leap Motion system for that purpose which allowed to track and display fine finger movements.

Data Measurement Setup

Multiple technologies were used to capture and measure bodily behavior and performance aspects of the pianists into quantitative time-series data. Concerning the performance, we recorded MIDI



data from the piano keyboards, including MIDI note numbers, note-on/off times, and note velocities. Delays from piano keypresses to audio equaled 16 ± 5 ms. We captured full body movements of all pianists (3D position, 120 Hz) using a multi-camera Qualisys optical motion capture system (OQUS 7+ cameras). Real-time streaming of full-body movement data of Confederate pianist 1 to visualization in the HMD had a latency of 54 ± 11 ms. In addition, video was recorded using a four-camera Qualisys Miquis system. Further, we captured how the Test pianist 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 the Test pianist 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 4**.

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 start in unison after which intermittent gradual tempo changes cause increasing phase shifts between the melodic patterns played by each pianist. These shifts, in their dynamic variety of interlocking, then lead to the emergence of a variety of harmonies over the course of the performance until pianists are back in unison. First bars of the piece are shown in **Figure 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 30–60 s) he is one sixteenth

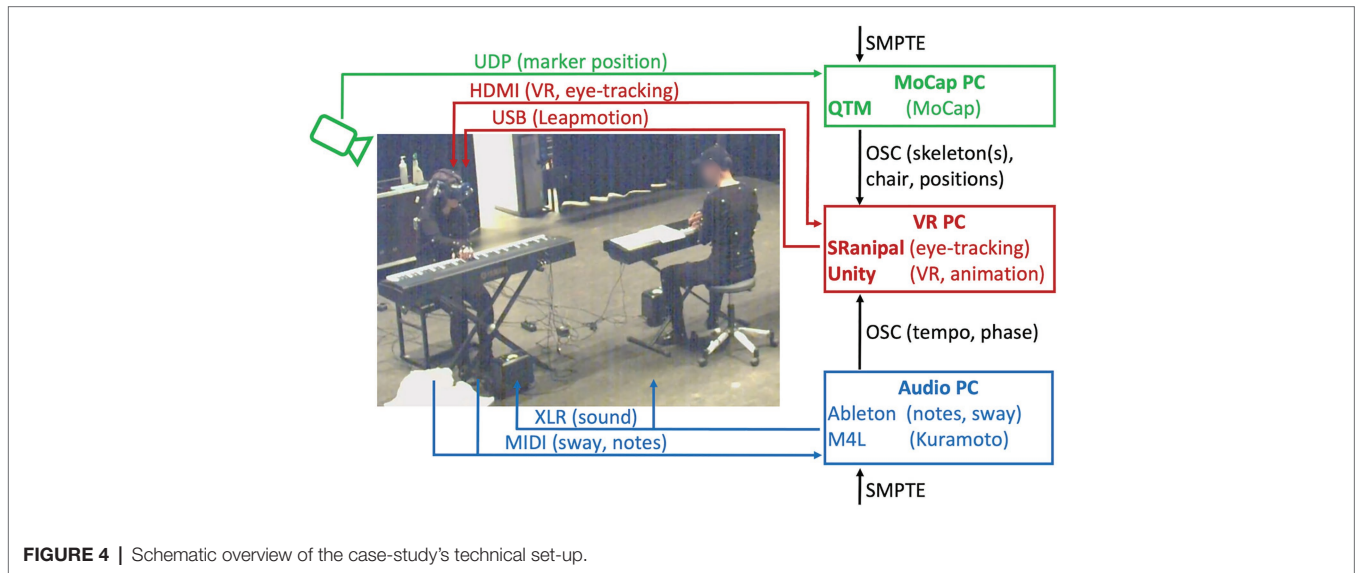


FIGURE 4 | Schematic overview of the case-study's technical set-up.

FIGURE 5 | Annotated first 11 bars of Piano Phase (1967) by Steve Reich (blue = melodic pattern, green = phase difference, and red = tempo instruction; reprinted with kind permission by Universal Edition AG, Vienna).

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 eighth (4), a dotted eighth (5), a quarter (6), etc., ahead of the first until he finally passes through all 12 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, the Test pianist was always assigned the top part of the score, while Confederate pianists 1 and 2 were assigned the bottom part in the Avatar and Agent condition, 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).

A month before the experiment, the Test pianist was asked to prepare for a performance of the musical composition. The Test pianist received an audio recording of her part with isochronous notes, uniform velocities, and linear accelerations to help with practicing. Upon arrival, the Test pianist 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 pianist practiced the task with Confederate pianist 1 for about 15 min without wearing the HMD. When both pianists indicated they were ready for the performance, the Test pianist 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 pianist was not told that the agent in the Agent condition was computer-controlled. The experiment concluded with a semi-structured interview about the experience. Five days after the experiment, the Test pianist was asked to listen and evaluate randomized audio recordings of each condition as well as fill in a final questionnaire.

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 the Results section below.

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 toward 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 and Jupp, 2009). 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 to identify transitions and behavior of a system by analyzing 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 to assess 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 0.3 s for a joint RQA. The minimal diagonal length for calculating DET was set to 0.3 s. Joint RQA parameters were obtained by looking at extrema of mutual information and false-nearest neighbor 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 pianist received six, 15-s, 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.

Layer 2: Embodied Co-regulation

In this layer, we assessed movement, behavior, 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 among 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 and Wittenburg, 2008) to see whether

communicative cues at transitional moments remained consistent across the different performances.

As a measure of coupling and attention toward the other, we recorded the Test pianist'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 0.35 s which were defined using mutual information and false-nearest neighbor metrics. Radii were set to produce around 10% recurrences for each condition [(Radius, RR%) equal (0.065, 10.66)_{HC}, (0.185, 10.63)_{AvC}, and (0.104, 10.55)_{AgC}].

Layer 3: Subjective Experience

In this layer, we looked at physiological signals and self-reported scores as windows into the Test pianist's experience during and after the interaction. The immersive tendencies questionnaire (Witmer and Singer, 1998) was taken before the experiment. Self-reported scores were obtained using the flow short scale (Martin and 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 the 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).

Results

This section presents the results from analyzing the case-study data. We evaluated each layer of the methodological framework in each condition with results shown in **Figures 6–8**. 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.

Layer 1: Performance Output

The score had a tempo indication of 72BPM and participants performed it slightly faster [(mean, STD)_{tempo} equaled (74.73, 4.44)_{HC}, (75.90, 4.62)_{AvC}, and (74.10, 5.67)_{AgC} BPM]. IOI variability was comparable across conditions with a higher variability for the Agent condition [(mean, STD)_{IOIvar} equaled (7.47, 3.63)_{HC}, (7.26, 3.41)_{AvC}, and (12.55, 2.73)_{AgC} ms]. As the agent was programmed to be attracted toward 72BPM, the slightly higher tempo of the performance made its tempo corrections larger when accelerating and synchronizing. Across conditions however, the Agent condition's tempo was closest to the instructed 72BPM.

Bars in the composition represented 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 gray areas in **Figures 6–8**. We have set the synchronization index threshold at 0.99, which allowed the discrimination of 25 measures in the Human condition. This choice was motivated by the fact that the Human condition was taken as the ground truth and the composition prescribed 25 measures of (de)synchronization. The Agent condition had a more variable bar length distribution with extremes of a long period of synchronization at 240 s and a turbulent moment of successive (de)synchronization around 280 s. Interestingly, a longer period of synchronization is found at the same relative phase in the Human condition as well. This relative phase and resulting harmony might have represented 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 **Figures 6B–8B**. The Agent 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 pianist jumping to synchronization with the agent toward the end.

Joint RQA showed alternating periods of (de)synchronization (**Figures 6C–8C**) and comparable average values for RR, DET, and TT across conditions. The Agent condition did contain more variability, especially during the earlier identified turbulent moments in the performance. The Human condition had a slightly higher average DET value indicating more stable synchronization. Fluctuations in DET values from (de)synchronizing were slightly larger in the Avatar condition compared to the Human condition.

Test pianist's scores on expressiveness, technical content, and interaction quality of each audio excerpt are indicated in **Figures 6A–8A** (mean scores: HC = 7.33; AvC = 7.17; AgC = 5.5). The Agent condition 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. Three excerpts from the Agent condition did get the remark "I do not understand the intention of the pianist."

In conclusion, performances in each condition were executed relatively well with stable tempos, fluctuating synchronization and RQA measures as instructed by the score. Increased DET fluctuations in the Avatar condition were indicative of a good performance given they are indicative of clearer alternating moments of (de)synchronization. Subjective scores by the Test pianist were good for the Human and Avatar conditions and just above average for the Agent condition. Analysis of the performance layer thus showed a good execution of the composition in the Human and Avatar conditions and showed more trouble performing successfully in the Agent condition. Underlying reasons might have resided in the embodied dynamics of coordination and communication that will be discussed in the next layer.

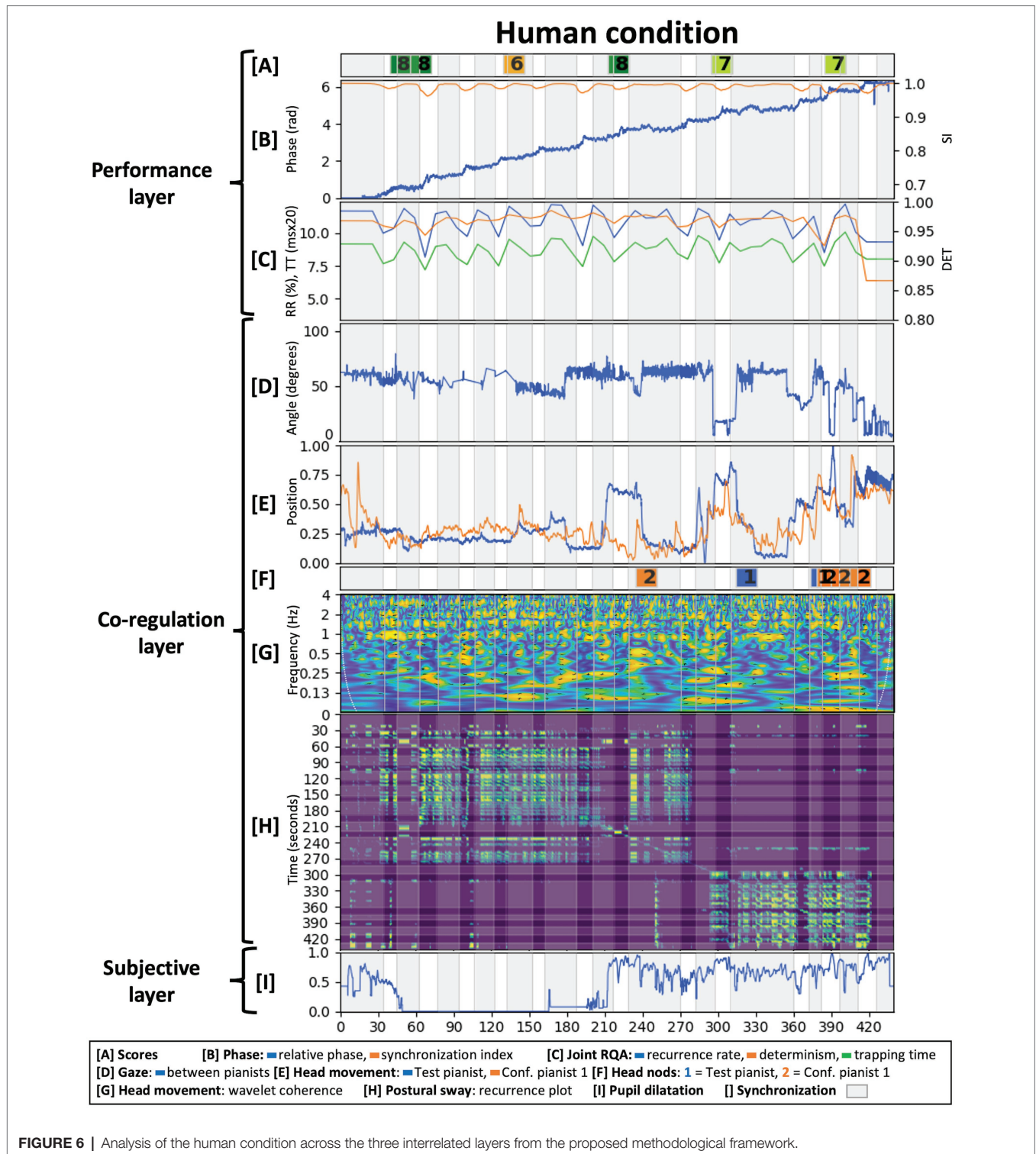


FIGURE 6 | Analysis of the human condition across the three interrelated layers from the proposed methodological framework.

Layer 2: Embodied Co-regulation

The normalized, principal component of the head movement timeseries of both pianists is shown in **Figures 6E–8E**. Wavelet coherence on these timeseries is shown in **Figures 6G–8G**. These plots show maxima at multiples of half the average tempo of the performance (72BPM or 1.2 Hz) with a less

outspoken pattern for the Agent condition. As the score had two beats per measure, it shows how pianists synchronized their head movements coherent with the musical structure. Head movement in the (0.9, 1.5) Hz band had a flat distribution of relative phase angles between pianists across conditions [(mean, resultant vector length) equaled (73°, 0.105)_{HC},

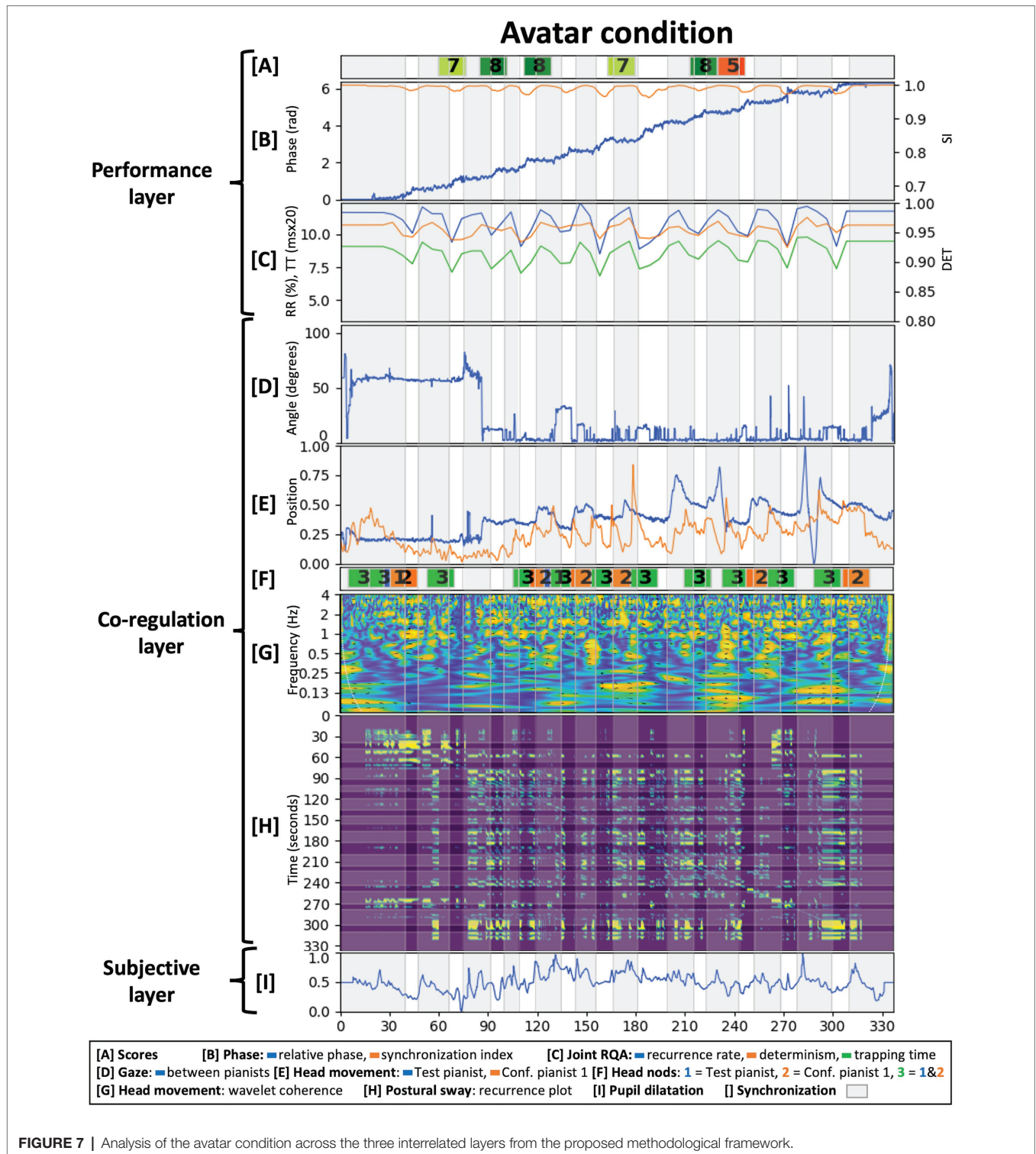


FIGURE 7 | Analysis of the avatar condition across the three interrelated layers from the proposed methodological framework.

$(-140^\circ, 0.127)_{AVC}$, and $(82^\circ, 0.070)_{AgC}$. Phase shifts in the music might have transferred to head movements and might suggest pianists were mainly keeping time for themselves.

Annotations of head nodding between pianists are shown in **Figures 6F–8F**. The Human condition contained cues from Confederate pianist 1 toward the end, the Avatar condition

contained several synchronized head nods between pianists and the Agent condition showed the absence of communication cues from the Test pianist toward the virtual agent. Synchronized head nodding in the Avatar condition also went together with a closely coupled gaze and regular RQA measures. This close coordination was felt by The Test pianist as she commented

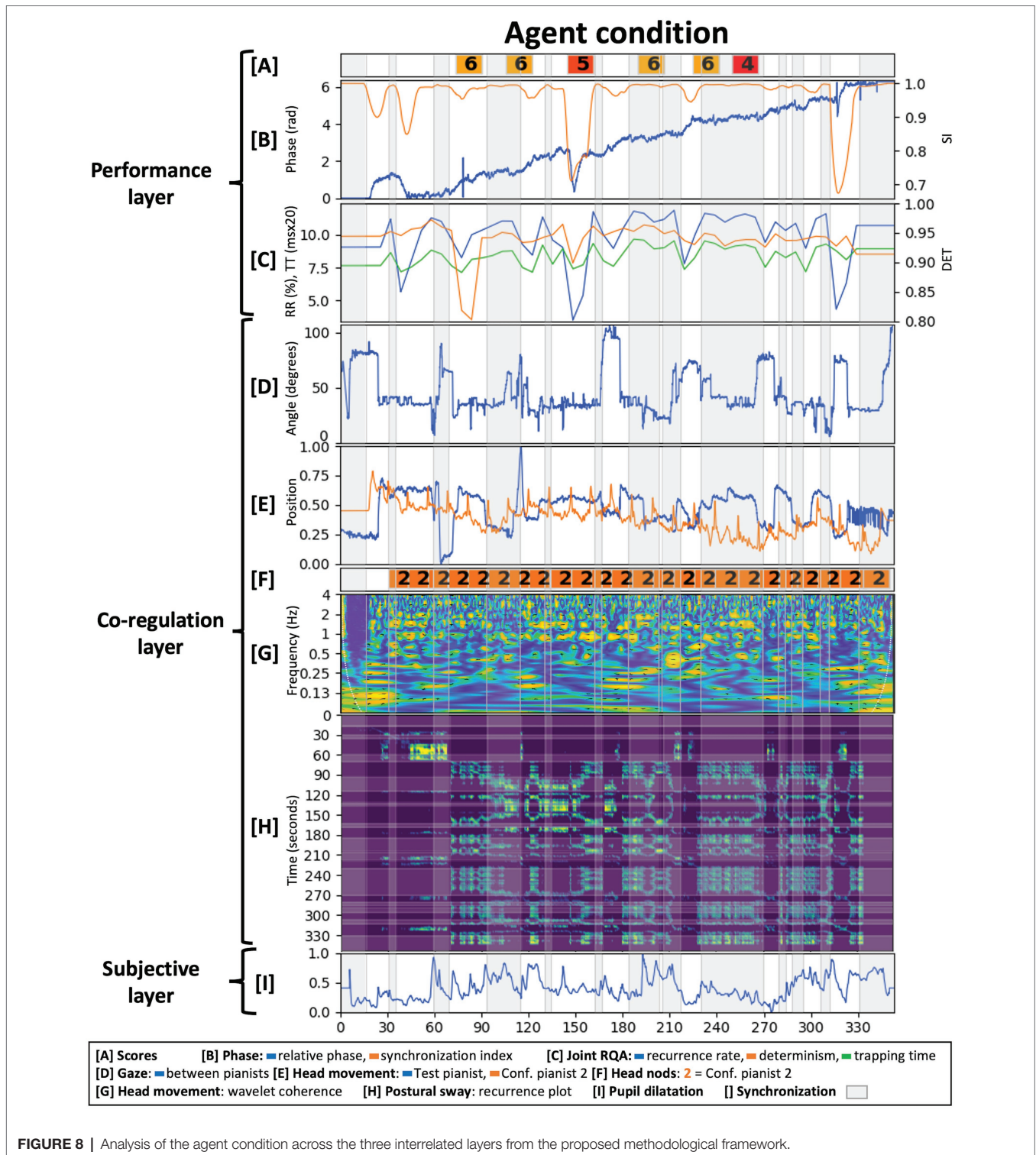


FIGURE 8 | Analysis of the agent condition across the three interrelated layers from the proposed methodological framework.

after the experiment that she missed the “posture mirroring” of the other pianist in the Agent condition as present in the Avatar condition.

Next, one can see differences between conditions for the gaze angle throughout time, between pianists in **Figures 6D–8D**. The Human condition had the Test pianist mainly looking

forward as the other pianist sat at an angle of 60°. **Figure 6D** shows a significant decrease of the gaze angle at 300 s in the Human condition just before a longer musical synchronization period. This transition is followed by a head nod of Confederate pianist 1, showing how attention shifted toward the Test pianist. The Avatar condition also had an important transition early

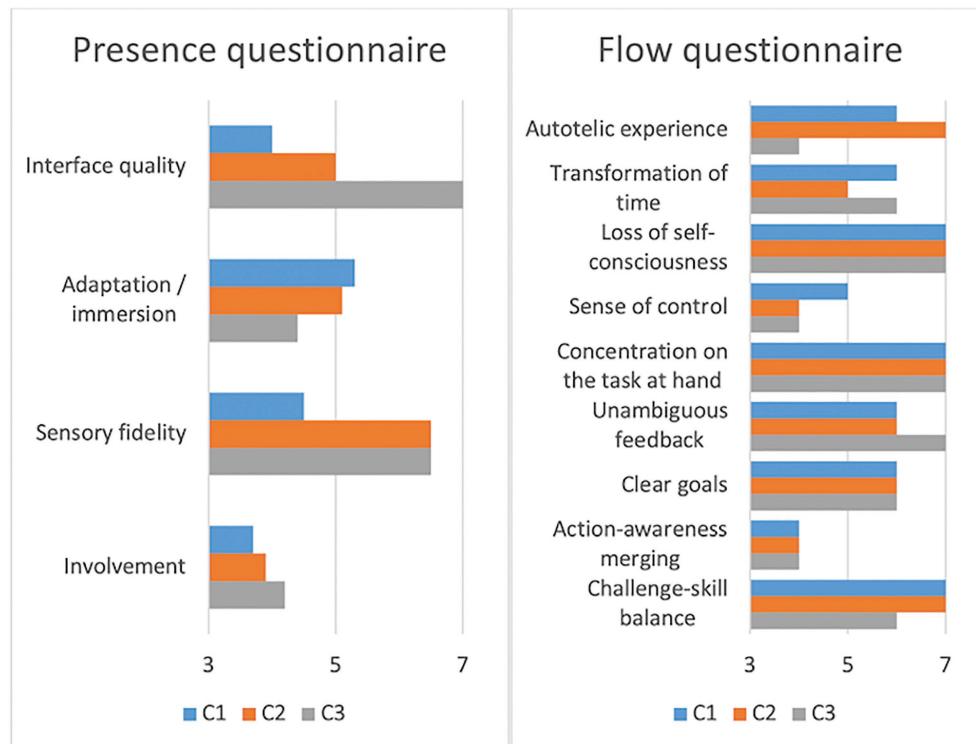


FIGURE 9 | Presence and flow scores of the Test pianist in all conditions (scores ranging from 1 = Strongly disagree to 7 = Strongly agree for the flow questionnaire and from 1 = Not at all to 7 = Completely for the presence questionnaire).

in the performance after which the Test pianist kept gaze directed toward the Confederate pianist 1. This attention shift resulted in a closer coordination as illustrated by synchronized head nods. In the Agent condition, pianists demonstrated less co-regulation with a gaze angle between pianists that never reached 0 degrees. The Test pianist had five moments of looking forward (around 70 degrees at 5, 65, 220, 265, and 315 s), a moment of looking away from the virtual agent (100 degrees at 170 s) and looking slightly inclined toward the virtual agent for the majority of the performance (around 40 degrees).

Figures 6H–8H shows recurrence plots of the postural sway of the Test pianist in each condition. At first sight, one can see a phase transition in the Human condition at 300 s. 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 pianist 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 the Agent condition 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.

Layer 3: Subjective Experience

For the qualitative aspects, we focused on the subjective experience of the Test pianist after the interaction. The analysis

goals were to evaluate how the Test pianist experienced each performance globally and in specific moments.

Global scores on the interaction from the custom questions showed satisfactory enjoyment for the Human and Avatar conditions [(HC, AvC, AgC) = (7,7,4)], the most natural interaction in the Human condition [(HC, AvC, AgC) = (4,1,1)] and interestingly, higher scores for experienced closeness to the musical partner in the Avatar condition [(HC, AvC, AgC) = (4,6,1)]. The Test pianist commented about the Avatar condition 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.”* The Agent condition had lowest scores on enjoyment, closeness, and naturalness. Comments of the Test pianist 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 pianist's profile (involvement = 6.38,

focus = 6.57, games = 5.00). Presence and flow scores are shown in **Figure 9**. Flow scores were relatively close across conditions ($\text{mean}_{\text{HC}} = 6.00$, $\text{mean}_{\text{AvC}} = 5.89$, $\text{mean}_{\text{AgC}} = 5.67$) with slightly less challenge-skill matching and autotelic experience in the Agent condition. Presence scores were comparable across conditions as well, with the Agent condition 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” (HC = 6, AvC = 7, AgC = 7), low scores in “spatial presence” for the Avatar condition (HC = 6, AvC = 1, AgC = 7), and low scores in “similarity to a real place” for the Agent condition (HC = 3, AvC = 5, AgC = 1).

Pupil dilations of the Test pianist are shown in **Figures 6I–8I**. It remained too challenging to extract meaningful insights or perform comparisons across conditions due to a technical issue in the Human condition (missing data for 2.5 min) and different light conditions across conditions (pass-through camera in the Human condition while a visualization of the virtual environment in the Avatar and Agent conditions). 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.

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 and Di Paolo, 2007). Once established, second person information in the interaction will have characteristics, such as self-directedness, contingency, reciprocity, affective engagement, and shared intentional relations (Moore and Barresi, 2017). This might have taken place in the Avatar condition 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; Walton et al., 2018b). 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 (Walton et al., 2018a).

A shared context implies the participants are involved in participatory sense-making (De Jaegher and 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 (Moore and Barresi, 2017). The relevance of virtual scenarios, avatars, and agents for this approach is exemplified

by the application of our methodological framework on the Avatar and Agent conditions in our case-study. Self-reports from the Test pianist directly referred to “the presence of a real person” about the avatar in the Avatar condition and the realization about the agent in the Agent condition “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 et al., 2017). The other layers in our framework further demonstrate the diffuse intentional relations in the Agent condition by a turbulent coordination and less movement synchronization resulting from a lack of communication between the pianists. As the Test pianist commented about a perceived “non-intentional agent,” it might have been unclear for the Test pianist when the performance was in a (de)synchronizing measure of the score.

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

Flow questionnaire scores were comparable across conditions. The presence questionnaire showed a low sense of spatial presence in the Avatar condition 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 the Agent condition 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 pianist to become more (negatively) emotionally involved and become more aware of playing with the piano interface. Musical performance in the Agent condition was not satisfactory for the pianist despite relatively satisfactory relative phase progression and joint RQA metrics. On the other hand, scores indicated a successful execution and enjoyment of the performance in the Human and Avatar conditions.

We argue that the combined successful musical performance, close coupling and shared intentional relations across different layers between the interacting individuals could be necessary and sufficient conditions for the feeling of social presence. Our methodological framework allowed 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 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.

GENERAL DISCUSSION AND FUTURE WORK

The design and analysis of the case-study based on our proposed methodological framework have allowed us to describe performative, behavioral, 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-centered interactions that involve emotional engagement with joint attention and joint goal-direction actions (Moore and Barresi, 2017). As such, their analysis could support the move toward a second person approach to social understanding and help close the gap between first person experiential and third 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 behavioral 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 behavior

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 (Mörzl et al., 2014; Shahal et al., 2020; Tognoli et al., 2020). While the main differentiator between conditions has been the behavioral realism of the virtual humans, future studies could vary their appearance realism as well and investigate possible influences on interaction dynamics and social presence (Bailenson et al., 2006; Roth et al., 2016).

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 analyze 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 (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 amnesic re-embodiment (Suzuki et al., 2012; Metzinger, 2018) 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 and Science Interaction lab using a modified version of the virtual environment described in this paper.¹

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 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 behavioral and experiential layers for interactions between the real pianist and virtual avatar and a good performance without interpersonal coordination for the interaction with the virtual agent. The analyses demonstrated the potential of our proposed framework in assessing social presence as well as in highlighting opportunities and challenges in developing better virtual interactions with and models of virtual humans.

¹A video is available at: <https://youtu.be/GIVaMPCotzM>.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in an online repository. The repository can be found at: <https://github.com/ArtScienceLab/Piano-Phase>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ghent University Faculty of Arts and Philosophy Ethics Committee. The participants provided their written informed consent to participate in this study.

REFERENCES

- Alcañiz, M., Guixeres, J., and Bigne, E. (2019). Virtual reality in marketing: a framework, review and research agenda. *Front. Psychol.* 10:1530. doi: 10.3389/fpsyg.2019.01530
- Atherton, J., and Wang, G. (2020). Doing vs. being: a philosophy of design for artful VR. *J. New Music Res.* 49, 35–59. doi: 10.1080/09298215.2019.1705862
- Bailenson, J. N., Yee, N., Merget, D., and Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators Virtual Environ.* 15, 359–372. doi: 10.1162/pres.15.4.359
- Banakou, D., and Slater, M. (2014). Body ownership causes illusory self-attribution of speaking and influences subsequent real speaking. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17678–17683. doi: 10.1073/pnas.1414936111
- Baumgartner, T., Valko, L., Esslen, M., and Jäncke, L. (2006). Neural correlate of spatial presence in an arousing and noninteractive virtual reality: an EEG and psychophysiology study. *CyberPsychol. Behav.* 9, 30–45. doi: 10.1089/cpb.2006.9.30
- Benedikt, M. (1991). *Cyberspace: First Steps*. Cambridge, MA: MIT Press.
- Berg, L. P., and Vance, J. M. (2017). Industry use of virtual reality in product design and manufacturing: a survey. *Virtual Reality* 21, 1–17. doi: 10.1007/s10055-016-0293-9
- Berthaut, F., Zappi, V., and Mazzanti, D. (2014). “Scenography of immersive virtual musical instruments.” in *2014 IEEE VR Workshop: Sonic Interaction in Virtual Environments (SIVE)*. 19–24.
- Biocca, F. (1992). Communication within virtual reality: creating a space for research. *J. Commun.* 42, 5–22. doi: 10.1111/j.1460-2466.1992.tb00810.x
- Biocca, F., and Delaney, B. (1995). “Immersive virtual reality technology.” in *Communication in the Age of Virtual Reality*. eds. F. Biocca and M. R. Levy (Hillsdale, NJ: Lawrence Erlbaum Associates Publishers), 57–124.
- Bissonnette, J., Dubé, F., Provencher, M. D., and Moreno Sala, M. T. (2016). Evolution of music performance anxiety and quality of performance during virtual reality exposure training. *Virtual Reality* 20, 71–81. doi: 10.1007/s10055-016-0283-y
- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., and Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* 13, 103–124. doi: 10.1207/s15327965PLI1302_01
- Braun, N., Debener, S., Spychala, N., Bongartz, E., Sörös, P., Müller, H. H. O., et al. (2018). The senses of agency and ownership: a review. *Front. Psychol.* 9:535. doi: 10.3389/fpsyg.2018.00535
- Çamci, A., and Hamilton, R. (2020). Audio-first VR: new perspectives on musical experiences in virtual environments. *J. New Music Res.* 49, 1–7. doi: 10.1080/09298215.2019.1707234
- Camurri, A., De Poli, G., Leman, M., and Volpe, G. (2001). “A multi-layered conceptual framework for expressive gesture applications.” in *Proc. Intl MOSART Workshop*, Barcelona.
- Caruso, G., Coorevits, E., Nijls, L., and Leman, M. (2016). Gestures in contemporary music performance: a method to assist the performer’s artistic process. *Contemp. Music. Rev.* 35, 402–422. doi: 10.1080/07494467.2016.1257292

AUTHOR CONTRIBUTIONS

BK co-designed the study, developed the experimental set-up, conducted experiments, performed data analysis, and co-drafted the manuscript. GC co-designed the study and provided the performance data for creation of the avatar. P-JM co-designed the study and co-drafted, revised, and approved the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by a starting grant of the special research fund (BOF19/STA/035) to P-JM.

- Castellano, G., Mortillaro, M., Camurri, A., Volpe, G., and Scherer, K. (2008). Automated analysis of body movement in emotionally expressive piano performances. *Music. Percept.* 26, 103–119. doi: 10.1525/mp.2008.26.2.103
- Cipresso, P. (2015). Modeling behavior dynamics using computational psychometrics within virtual worlds. *Front. Psychol.* 6:1725. doi: 10.3389/fpsyg.2015.01725
- Coorevits, E., Maes, P.-J., Six, J., and Leman, M. (2020). The influence of performing gesture type on interpersonal musical timing, and the role of visual contact and tempo. *Acta Psychol.* 210:103166. doi: 10.1016/j.actpsy.2020.103166
- Cruz-Neira, C. (2003). Computational humanities: the new challenge for VR. *IEEE Comput. Graph. Appl.* 23, 10–13. doi: 10.1109/MCG.2003.1198257
- Cui, G. (2013). Evaluating online social presence: an overview of social presence assessment. *J. Educ. Technol. Develop. Exch.* 6, 13–30. doi: 10.18785/jetde.0601.02
- D’Ausilio, A., Novembre, G., Fadiga, L., and Keller, P. E. (2015). What can music tell us about social interaction? *Trends Cogn. Sci.* 19, 111–114. doi: 10.1016/j.tics.2015.01.005
- Deacon, T., Stockman, T., and Barthelet, M. (2016). “User experience in an interactive music virtual reality system: an exploratory study.” in *International Symposium on Computer Music Multidisciplinary Research*. 192–216.
- De Jaegher, H., and Di Paolo, E. (2007). Participatory sense-making. *Phenomenol. Cogn. Sci.* 6, 485–507. doi: 10.1007/s11097-007-9076-9
- Demos, A. P., Chaffin, R., and Logan, T. (2018). Musicians body sway embodies musical structure and expression: a recurrence-based approach. *Music. Sci.* 22, 244–263. doi: 10.1177/1029864916685928
- De Oliveira, E. C., Bertrand, P., Lesur, M. E. R., Palomo, P., Demarzo, M., Cebolla, A., et al. (2016). “Virtual body swap: a new feasible tool to be explored in health and education.” in *Proceedings - 18th Symposium on Virtual and Augmented Reality, SVR 2016*. Gramado, 81–89.
- Doffman, M. (2009). Making it groove! Entrainment, participation and discrepancy in the “conversation” of a jazz trio. *Lang. Hist.* 52, 130–147. doi: 10.1179/175975309x452012
- Dumas, G., Lachat, F., Martinerie, J., Nadel, J., and George, N. (2011). From social behaviour to brain synchronization: review and perspectives in hyperscanning. *IRBM* 32, 48–53. doi: 10.1016/j.irbm.2011.01.002
- Eerola, T., Jakubowski, K., Moran, N., Keller, P. E., and Clayton, M. (2018). Shared periodic performer movements coordinate interactions in duo improvisations. *R. Soc. Open Sci.* 5:171520. doi: 10.1098/rsos.171520
- Ekman, P. (1992). *Are There Basic Emotions?* Washington, DC: American Psychological Association.
- Fogel, A. (2017). “Two principles of communication: co-regulation and framing.” in *New Perspectives in Early Communicative Development*. eds. J. Nadel and L. Camaioni (Routledge), 9–22.
- Fox, J., Arena, D., and Bailenson, J. N. (2009). Virtual reality: a survival guide for the social scientist. *J. Media Psychol.* 21, 95–113. doi: 10.1027/1864-1105.21.3.95
- Friedman, D., Pizarro, R., Or-Berkers, K., Neyret, S., Pan, X., and Slater, M. (2014). A method for generating an illusion of backwards time travel using immersive virtual reality—an exploratory study. *Front. Psychol.* 5:943. doi: 10.3389/fpsyg.2014.00943

- Froese, T., and Fuchs, T. (2012). The extended body: a case study in the neurophenomenology of social interaction. *Phenomenol. Cogn. Sci.* 11, 205–235. doi: 10.1007/s11097-012-9254-2
- Gallagher, S. (2001). The practice of mind. Theory, simulation or primary interaction? *J. Conscious. Stud.* 8, 83–108.
- Gallagher, S., and Allen, M. (2018). Active inference, enactivism and the hermeneutics of social cognition. *Synthese* 195, 2627–2648. doi: 10.1007/s11229-016-1269-8
- Gallotti, M., and Frith, C. D. (2013). Social cognition in the we-mode. *Trends Cogn. Sci.* 17, 160–165. doi: 10.1016/j.tics.2013.02.002
- Garau, M., Slater, M., Pertaub, D.-P., and Razaque, S. (2005). The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators Virtual Environ.* 14, 104–116. doi: 10.1162/1054746053890242
- Gibson, W. (1984). *Neuromancer*. Berkeley, CA: Ace Books.
- Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., and Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Curr. Biol.* 16, 428–432. doi: 10.1016/j.cub.2006.01.019
- Glowinski, D., Baron, N., Shirole, K., Coll, S. Y., Chaabi, L., Ott, T., et al. (2015). Evaluating music performance and context-sensitivity with immersive virtual environments. *EAI Endors. Trans. Creat. Technol.* 2:e3. doi: 10.4108/ct.2.2.e3
- Goebel, W., and Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music. Percept.* 26, 427–438. doi: 10.1525/mp.2009.26.5.427
- Hamilton, R. (2019). “Mediated musical interactions in virtual environments,” in *New Directions in Music and Human-Computer Interaction*. eds. S. Holland, T. Mudd, K. Wilkie-McKenna, A. McPherson and M. M. Wanderley (Springer), 243–257.
- Heim, M. (1994). *The Metaphysics of Virtual Reality*. New York: Oxford University Press.
- Herrera, G., Jordan, R., and Vera, L. (2006). Agency and presence: a common dependence on subjectivity? *Presence Teleop. Virt.* 15, 539–552. doi: 10.1162/pres.15.5.539
- Hilt, P. M., Badino, L., D’Ausilio, A., Volpe, G., Tokay, S., Fadiga, L., et al. (2020). Author correction: multi-layer adaptation of group coordination in musical ensembles. *Sci. Rep.* 10:597. doi: 10.1038/s41598-019-55965-3
- Honigman, C., Walton, A., and Kapur, A. (2013). “The third room: a 3d virtual music paradigm.” in *Proceedings of the International Conference on New Interfaces for Musical Expression*. 29–34.
- Ijsselstein, W., and Riva, G. (2003). “Being there: the experience of presence in mediated environments,” in *Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments*. eds. G. Riva, F. Davide and W. A. Ijsselstein (Amsterdam, Netherlands: IOS Press), 4–14.
- Johanson, M. (2015). The turing test for telepresence. *Int. J. Multimed. Appl.* 7, 1–20. doi: 10.5121/ijma.2015.7501
- Keller, P. E., and Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music. Percept.* 28, 27–46. doi: 10.1525/mp.2010.28.1.27
- Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT press.
- Kern, A. C., and Ellermeier, W. (2020). Audio in VR: effects of a soundscape and movement-triggered step sounds on presence. *Front. Robot. AI* 7:20. doi: 10.3389/frobot.2020.00020
- Kiltner, K., Bergstrom, I., and Slater, M. (2013). Drumming in immersive virtual reality: the body shapes the way we play. *IEEE Trans. Vis. Comput. Graph.* 19, 597–605. doi: 10.1109/TVCG.2013.29
- Kiltner, K., Groten, R., and Slater, M. (2012). The sense of embodiment in virtual reality. *Presence Teleop. Virt.* 21, 373–387. doi: 10.1162/pres_a_00124
- Kiverstein, J. (2011). Social understanding without mentalizing. *Philos. Top.* 39, 41–65. doi: 10.5840/philtopics201139113
- Koban, L., Ramamoorthy, A., and Konvalinka, I. (2019). Why do we fall into sync with others? Interpersonal synchronization and the brain’s optimization principle. *Soc. Neurosci.* 14, 1–9. doi: 10.1080/17470919.2017.1400463
- Kobayashi, M., Ueno, K., and Ise, S. (2020). The effects of spatialized sounds on the sense of presence in auditory virtual environments: a psychological and physiological study. *Presence Teleop. Virt.* 24, 163–174. doi: 10.1162/pres_a_00226
- Kornelsen, J. J. A. (1991). *Virtual Reality?: Marshall McLuhan and a Phenomenological Investigation of the Construction of Virtual Worlds*. dissertation/master’s thesis. Canada: Simon Fraser University
- Kothgassner, O. D., and Felnhöfer, A. (2020). Does virtual reality help to cut the Gordian knot between ecological validity and experimental control? *Ann. Int. Commun. Assoc.* 44, 210–218. doi: 10.1080/23808985.2020.1792790
- Kothgassner, O. D., Griesinger, M., Kettner, K., Wayan, K., Völkl-Kernstock, S., Hlavacs, H., et al. (2017). Real-life prosocial behavior decreases after being socially excluded by avatars, not agents. *Comput. Hum. Behav.* 70, 261–269. doi: 10.1016/j.chb.2016.12.059
- Laeng, B., Sirois, S., and Gredebäck, G. (2012). Pupillometry: a window to the preconscious? *Perspect. Psychol. Sci.* 7, 18–27. doi: 10.1177/1745691611427305
- Lanier, J. (1988). A Vintage Virtual Reality Interview. Available at: <http://www.jaronlanier.com/vrint.html> (Accessed January 11, 2021).
- Leman, M. (2008). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press.
- Leman, M., and Maes, P.-J. (2015). The role of embodiment in the perception of music. *Empir. Musicol. Rev.* 9, 236–246. doi: 10.18061/emr.v9i3-4.4498
- Lerch, A., Arthur, C., Pati, A., and Gururani, S. (2020). An interdisciplinary review of music performance analysis. *Trans. Int. Soc. Music. Inf. Retr.* 3, 221–245. doi: 10.5334/tismir.53
- Lessiter, J., Freeman, J., Keogh, E., and Davidoff, J. (2001). A cross-media presence questionnaire: the ITC-sense of presence inventory. *Presence: Teleoperators Virtual Environ.* 10, 282–297. doi: 10.1162/105474601300343612
- Lombard, M., and Ditton, T. (1997). At the heart of it all: the concept of presence. *J. Comput.-Mediat. Commun.* 3:JCMC321.
- Mardia, K. V., and Jupp, P. E. (2009). *Directional Statistics*. Hoboken, NJ: John Wiley & Sons.
- Martin, A. J., and Jackson, S. A. (2008). Brief approaches to assessing task absorption and enhanced subjective experience: examining ‘short’ and ‘core’ flow in diverse performance domains. *Motiv. Emot.* 32, 141–157. doi: 10.1007/s11031-008-9094-0
- Marwan, N., Carmenromano, M., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Phys. Rep.* 438, 237–329. doi: 10.1016/j.physrep.2006.11.001
- Matamala-Gomez, M., Maselli, A., Malighetti, C., Realdon, O., Mantovani, F., and Riva, G. (2020). Body illusions for mental health: a systematic review. *PsyArXiv [Preprint]*.
- McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. New York, NY: McGraw-Hill Education.
- Meehan, M., Insko, B., Whitton, M., and Brooks, F. B. (2002). Physiological measures of presence in stressful virtual environments. *ACM Trans. Graph.* 21, 645–652. doi: 10.1145/566654.566630
- Metzinger, T. K. (2018). Why is virtual reality interesting for philosophers? *Front. Robot. AI* 5:101. doi: 10.3389/frobot.2018.00101
- Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). “Augmented reality: a class of displays on the reality-virtuality continuum,” in *Telematics and Telepresence Technologies*. ed. H. Das (Bellingham, WA: SPIE), 282–292.
- Minsky, M. (1980). Telepresence. *Omni* 2, 45–52.
- Moore, C., and Barresi, J. (2017). The role of second-person information in the development of social understanding. *Front. Psychol.* 8:1667. doi: 10.3389/fpsyg.2017.01667
- Mörzl, A., Lorenz, T., and Hirche, S. (2014). Rhythm patterns interaction-synchronization behavior for human-robot joint action. *PLoS One* 9:e95195. doi: 10.1371/journal.pone.0095195
- Nijs, L., Coussement, P., Moens, B., Amelinck, D., Lesaffre, M., and Leman, M. (2012). Interacting with the music paint machine: relating the constructs of flow experience and presence. *Interact. Comput.* 24, 237–250. doi: 10.1016/j.intcom.2012.05.002
- Oh, C. S., Bailenson, J. N., and Welch, G. F. (2018). A systematic review of social presence: definition, antecedents, and implications. *Front. Robot. AI* 5:114. doi: 10.3389/frobot.2018.00114
- Orman, E. K. (2004). Effect of virtual reality graded exposure on anxiety levels of performing musicians: a case study. *J. Music. Ther.* 41, 70–78. doi: 10.1093/jmt/41.1.70
- Orman, E. K., Price, H. E., and Russell, C. R. (2017). Feasibility of using an augmented immersive virtual reality learning environment to enhance music conducting skills. *J. Music. Teach. Educ.* 27, 24–35. doi: 10.1177/1057083717697962
- Pan, X., and de Hamilton, A. F. C. (2018). Why and how to use virtual reality to study human social interaction: the challenges of exploring a new research landscape. *Br. J. Psychol.* 109, 395–417. doi: 10.1111/bjop.12290
- Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front. Hum. Neurosci.* 9:660. doi: 10.3389/fnhum.2015.00660

- Parsons, T. D., Gaggioli, A., and Riva, G. (2017). Virtual reality for research in social neuroscience. *Brain Sci.* 7:42. doi: 10.3390/brainsci7040042
- Petkova, V. I., and Ehrsson, H. H. (2008). If I were you: perceptual illusion of body swapping. *PLoS One* 3:e3832. doi: 10.1371/journal.pone.0003832
- Reich, S. (2002). *Writings on Music, 1965–2000*. New York: Oxford University Press.
- 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*. eds. G. Riva, M. T. Anguera, B. K. Wiederhold and F. Mantovani (Amsterdam, Netherlands, IOS Press), 47–80.
- Riva, G., Waterworth, J. A., and Waterworth, E. L. (2004). The layers of presence: a bio-cultural approach to understanding presence in natural and mediated environments. *CyberPsychol. Behav.* 7, 402–416. doi: 10.1089/cpb.2004.7.402
- Riva, G., Wiederhold, B. K., and Mantovani, F. (2019). Neuroscience of virtual reality: from virtual exposure to embodied medicine. *Cyberpsychol. Behav. Soc. Netw.* 22, 82–96. doi: 10.1089/cyber.2017.29099.gri
- Roth, D., Lugrin, J. -L., Galakhov, D., Hofmann, A., Bente, G., Latoschik, M. E., et al. (2016). “Avatar realism and social interaction quality in virtual reality.” in *IEEE Virtual Reality (VR)*. 277–278.
- Scarfe, P., and Glennerster, A. (2019). The science behind virtual reality displays. *Annu. Rev. Vis. Sci.* 5, 1–20. doi: 10.1146/annurev-vision-091718-014942
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660
- Schmidt, P., Reiss, A., Dürichen, R., and Van Laerhoven, K. (2019). Wearable-based affect recognition—a review. *Sensors* 19:4079. doi: 10.3390/s19194079
- Sebanz, N., and Knoblich, G. (2009). Prediction in joint action: what, when, and where. *Top. Cogn. Sci.* 1, 353–367. doi: 10.1111/j.1756-8765.2009.01024.x
- Serafin, S., Adjorlu, A., Nilsson, N., Thomsen, L., and Nordahl, R. (2017). “Considerations on the use of virtual and augmented reality technologies in music education.” in *2017 IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR)*. 1–4.
- Serafin, S., Erkut, C., Kojs, J., Nilsson, N. C., and Nordahl, R. (2016). Virtual reality musical instruments: state of the art, design principles, and future directions. *Comput. Music. J.* 40, 22–40. doi: 10.1162/COMJ_a_00372
- Shahal, S., Wurzburg, A., Sibony, I., Duadi, H., Shniderman, E., Weymouth, D., et al. (2020). Synchronization of complex human networks. *Nat. Commun.* 11, 1–10. doi: 10.1038/s41467-020-17540-7
- Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence Teleop. Virt.* 1, 120–126. doi: 10.1162/pres.1992.1.1.120
- Shirzadian, N., Redi, J. A., Röggl, T., Panza, A., Nack, F., and Cesar, P. (2017). “Immersion and togetherness: how live visualization of audience engagement can enhance music events.” in *International Conference on Advances in Computer Entertainment*. 488–507.
- Short, J., Williams, E., and Christie, B. (1976). *The Social Psychology of Telecommunications*. Hoboken, NJ: John Wiley & Sons.
- Slater, M., Lotto, B., Arnold, M. M., and Maria, S. -V. (2009). How we experience immersive virtual environments: the concept of presence and its measurement. *Anuario de Psicología* 40, 193–210.
- Slater, M., and Sanchez-Vives, M. (2016). Enhancing our lives with immersive virtual reality. *Front. Robot. AI* 3:74. doi: 10.3389/frobt.2016.00074
- Sloetjes, H., and Wittenburg, P. (2008). “Annotation by category-ELAN and ISO DCR.” in *6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Sutherland, I. E. (1968). “A head-mounted three dimensional display.” in *AFIPS’ 68 (Fall, part I): Proceedings of the Fall Joint Computer Conference*. 757–764.
- Suzuki, K., Wakisaka, S., and Fujii, N. (2012). Substitutional reality system: a novel experimental platform for experiencing alternative reality. *Sci. Rep.* 2:459. doi: 10.1038/srep00459
- Tarr, B., Slater, M., and Cohen, E. (2018). Synchrony and social connection in immersive virtual reality. *Sci. Rep.* 8, 1–8. doi: 10.1038/s41598-018-21765-4
- Taylor, J. (1997). The emerging geographies of virtual worlds. *Geogr. Rev.* 87, 172–192. doi: 10.2307/216004
- Teo, W. P., Muthalib, M., Yamin, S., Hendy, A. M., Bramstedt, K., Kotsopoulos, E., et al. (2016). Does a combination of virtual reality, neuromodulation and neuroimaging provide a comprehensive platform for neurorehabilitation? A narrative review of the literature. *Front. Hum. Neurosci.* 10:284. doi: 10.3389/fnhum.2016.00284
- Tognoli, E., Zhang, M., Fuchs, A., Beetle, C., and Kelso, J. A. S. (2020). Coordination dynamics: a foundation for understanding social behavior. *Front. Hum. Neurosci.* 14:317. doi: 10.3389/fnhum.2020.00317
- Västfjäll, D. (2003). The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. *CyberPsychol. Behav.* 6, 181–188. doi: 10.1089/109493103321640374
- Walton, A. E., Washburn, A., Langland-Hassan, P., Chemero, A., Kloos, H., and Richardson, M. J. (2018a). Creating time: social collaboration in music improvisation. *Top. Cogn. Sci.* 10, 95–119. doi: 10.1111/tops.12306
- Walton, A. E., Washburn, A., Richardson, M. J., and Chemero, A. (2018b). “Empathy and groove in musical movement” in *Proceedings of A Body of Knowledge - Embodied Cognition and the Arts conference 2016*; December 8–10, 2016.
- Warren, W. H. (2006). The dynamics of perception and action. *Psychol. Rev.* 113:358. doi: 10.1037/0033-295X.113.2.358
- Williamson, A., Aufegger, L., and Eiholzer, H. (2014). Simulating and stimulating performance: introducing distributed simulation to enhance musical learning and performance. *Front. Psychol.* 5:25. doi: 10.3389/fpsyg.2014.00025
- Williamson, A., and Davidson, J. W. (2002). Exploring co-performer communication. *Music. Sci.* 6, 53–72. doi: 10.1177/102986490200600103
- Witmer, B. G., Jerome, C. J., and Singer, M. J. (2005). The factor structure of the presence questionnaire. *Presence: Teleoperators Virtual Environ.* 14, 298–312. doi: 10.1162/105474605323384654
- Witmer, B. G., and Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence* 7, 225–240. doi: 10.1162/105474698565686
- Yang, L. I., Jin, H., Feng, T., Hong-An, W., and Guo-Zhong, D. (2019). Gesture interaction in virtual reality. *Virtual Real. Intell. Hardw.* 1, 84–112. doi: 10.3724/sp.j.2096-5796.2018.0006
- Zahavi, D. (2001). Beyond empathy. Phenomenological approaches to intersubjectivity. *J. Conscious. Stud.* 8, 151–167.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Van Kerrebroeck, Caruso and Maes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.