

## Research Article

# A Presence- and Performance-Driven Framework to Investigate Interactive Networked Music Learning Scenarios

**Stefano Delle Monache** <sup>1</sup>, **Luca Comanducci** <sup>2</sup>, **Michele Buccoli**,<sup>2</sup> **Massimiliano Zanoni**,<sup>2</sup> **Augusto Sarti**,<sup>2</sup> **Enrico Pietrocola**,<sup>3</sup> **Filippo Berbenni**,<sup>3</sup> and **Giovanni Cospito**<sup>3</sup>

<sup>1</sup>Department of Architecture and Arts, IUAV University of Venice, Venice, Italy

<sup>2</sup>Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy

<sup>3</sup>Department New Technologies and Musical Languages, “G. Verdi” Music Conservatory, Milano, Italy

Correspondence should be addressed to Stefano Delle Monache; [stefano.dellemonache@iuav.it](mailto:stefano.dellemonache@iuav.it)

Received 7 February 2019; Revised 1 July 2019; Accepted 18 July 2019; Published 26 August 2019

Guest Editor: Michele Geronazzo

Copyright © 2019 Stefano Delle Monache et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cooperative music making in networked environments has been subject of extensive research, scientific and artistic. Networked music performance (NMP) is attracting renewed interest thanks to the growing availability of effective technology and tools for computer-based communications, especially in the area of distance and blended learning applications. We propose a conceptual framework for NMP research and design in the context of classical chamber music practice and learning: presence-related constructs and objective quality metrics are used to problematize and systematize the many factors affecting the experience of studying and practicing music in a networked environment. To this end, a preliminary NMP experiment on the effect of latency on chamber music duos experience and quality of the performance is introduced. The degree of involvement, perceived coherence, and immersion of the NMP environment are here combined with measures on the networked performance, including tempo trends and misalignments from the shared score. Early results on the impact of temporal factors on NMP musical interaction are outlined, and their methodological implications for the design of pedagogical applications are discussed.

## 1. Introduction

As distributed and ubiquitous interactive applications are increasingly populating our daily environments, we find ourselves constantly engaged in making sense of geographically displaced social practices and behaviors for the purpose of communicating, sharing, and cooperating. The network, in fact, is progressively evolving from a medium of communication, to a shared space virtually inhabited by bodily presences. Cooperative music making is a form of social practice characterized by peculiar temporal and spatial relationships, and in networked music performance (NMP), such relationships are unavoidably altered by the interposition of the network [1].

Computer systems for networked musical interaction have been categorized according to their temporal

(synchronous vs. asynchronous) and spatial (co-located vs. remote) dimensions of the performance [2], whereas a wide part of the scientific literature has focused on the technological and perceptual issues in real-time performance between musicians located in remote rooms, and requiring the highest degree of synchronicity, typically over teleconference-based communication media [3–5]. From a different angle, for years, research in telematic arts and music has been questioning the NMP communication model, stressing instead the role of the body in the network, and proposing an interpretive model of the emerging space, wherein questions of presence, plausible representations, sense of agency, and flow become crucial [6–8]. According to this view, telematic systems are conceived as proper instruments, whose apparent limitations are exploited instead in terms of creative affordances, for music purposes and interactions [9].

At the intersection between such disciplines, we witness a renewed interest in NMP technologies for music education [10]. NMP tools, in fact, are progressively being introduced in the market, and some solutions are emerging as viable standards in blended and distance learning, that is, a knowledge delivery combined with computer-mediated instruction, synchronous and face-to-face, or asynchronous [11]. The NMP approach to higher music education and pedagogy is the main locus of interest of this paper. Understanding how the use of mediating technologies affects the learning activity and experience becomes critical not only to inform the design and the development of learning environments and setups but also to adopt the appropriate pedagogical strategies [12–14].

The EU-funded project InterMUSIC (Interactive Environment for Music Learning and Practicing, 2017–2020) aims at bringing the solutions developed in NMP research within the context of distance education. The objective of the project is to distill a systematized knowledge in the form of best practices and guidelines for the design of remote environments for music interaction and higher education. Within the project activities, InterMUSIC is authoring three online pilot courses: music theory and composition, chamber music practice, and vocal training. The tools for the courses are developed according to two main paradigms. The former is based on MOOC approaches, namely, massive open online courses, wherein the learning experience is essentially individual and based on asynchronous communication with the master and limited collaboration with peers [15]. In the latter, students have access to NMP environments for attending master-student lessons or rehearse together. This paradigm clearly exhibits stricter requirements in terms of synchronicity and enables one-to-one (or few-to-few) synchronous communication. In this article, we focus on the NMP-based paradigm and we discuss the requirements related to the specific scenario of the chamber music practice. In particular, we approach the issue of synchronicity in NMP as a necessary condition for effective musical interaction, by focusing on how rhythm, melody, and expression affect the interaction and communication between performers. The existing literature on NMP pinpointed several factors affecting the objective quality of the music performance, namely, the unavoidable network latencies [16] and the effect of the resulting temporal separation of action and feedback [17], the timbral properties of the instruments [18], and the rhythmic complexity of the performed pieces [19]. These constitute the point of departure towards an interaction-centered and situated research approach to the case of learning and practicing chamber music in the NMP environment [20].

In the InterMUSIC project scenario, the goal is not to assess the NMP's objective quality, but rather to understand and frame those elements that can enable a comfortable remote rehearsal in teacher-student and student-student communication. Expert musicians are used to coping with the most adverse and diverse performing conditions, by relying on strong sensory-motor associations and by adapting their action planning in response to altered auditory feedback as much as possible [21, 22]. From this

viewpoint, much of the InterMUSIC research on NMP pedagogy is addressed to understand the design of learning environments in which the disruptive effects of the temporal and spatial alterations, due to the inherent physical distance and remoteness of performers, can be minimized or compensated through specific training and exercises.

In this article, we conceptualize the many forms of the networked music performance in a framework situated in the chamber music education scenario. Yet, the framework is abstracted enough to design and conduct our experimental activity and develop the pedagogical scenarios, and therefore it is the choice of appropriate technological solutions. In other words, we generalize the NMP actors and roles, at the same time narrowing the field of inquiry to the chamber music case, that is, we expect, for instance, findings not necessarily valid or prescriptive to other types of music practices (e.g., jazz improvisation). To this end, the emerging conceptual framework equally considers both the user experience, via presence constructs, and performance quality perspectives. In other words, we make use of metrics which describe the musicians' response to the NMP system and objective quality metrics of their performance as hypotheses generator for understanding and framing the design problem.

To operationalize our approach, we outline a pilot study, aimed at understanding how temporal factors (i.e., network latency) affect the interaction and communication of chamber music duos involved in remote collaboration for music making. The premise is that effective music making and communication rely on the availability of auditory and visual cues (i.e., sonic gestures) [23], which are inevitably constrained/limited in NMP and in telepresence environments in general. In other words, we explore how temporal alterations in audio-visual feedback affect the co-representation and coordination of ensemble musicians and seek for design strategies to compensate and facilitate a plausible music experience in the mediated environment. In the practice, we ask chamber music duos to perform a short exercise, under diverse conditions of network delay. The exercise is specifically conceived around musical structures which are functional to pinpointing a set of problems relative to time management, communication mechanisms, and mutual understanding between remote performers. The co-representation and coordination of duos are investigated by means of a presence questionnaire [24] and combined with objective metrics of the quality of the performance [19]. In this respect, we are exploring how and whether simulated network delay affects the subjective experience of being simultaneously present in the shared reality environment [25]. Sensory breadth and depth, degree of control, and anticipation of events, together with the overall interactivity of the environment, represent crucial elements in both presence and performance, being the first a prerequisite for the second [26].

The goal of this pilot study was to set the experimental playground, by exploring the NMP music learning scenario from a perspective (i.e., the presence experience), which is often recalled in the NMP literature, but that has never been systematically investigated [7]. In this respect, the 5 sessions collected do not make the underway study mature enough to

provide conclusions supported by statistical evidence. On the contrary, the information that we are gathering is instrumental to conceptualizing the relevant temporal (and spatial) factors affecting the NMP experience, from both the presence and performance perspectives. In other words, the exploratory framework, discussed in this paper, sets the reference for the experimentation with a more appropriate, extensive number of duos. Finally, the follow-up study will be addressed instead to the investigation of spatial representations, auditory and visual.

The paper is organized as follows: we first introduce the conceptual framework for NMP research in the classical chamber music scenario, in the context of presence studies and NMP music interaction; we then describe the experimental study, the design of the related questionnaire, and the selection of the quality metrics. Finally, we reflect on the methodological and design implications, by providing a narrative of the early results that we are collecting.

## 2. A Framework for InterMUSIC Research

In this section, we frame the relevant literature for the design of the chamber music practice and learning scenarios and experiments. We essentially look at that rich corpus of research focused on the temporal and spatial aspects, network constraints, and sound reproduction strategies, which have been considered the main factors affecting the sensory and control dimensions involved in the experience of playing together in a remote fashion [27].

In InterMUSIC, we approach the problem of designing an effective communication and interaction environment for networked music practice and teaching, by looking also at studies on presence [26, 28, 29] in the embodied cognition framework [24, 30]. The basic assumption is that the experience of music emerges in interaction, as complex network of predictive models of observable patterns and intentional states, which are acquired through knowledge and skills [31].

The literature overview is distilled in a conceptual framework of the NMP chamber music scenario, which is functional to devising and designing our experiments on the NMP performance as a whole and in its constitutive elements, i.e., actors, environment, and medium.

*2.1. Temporal and Spatial Factors in NMP.* The continuity of the time dimension in both remote music performance and tuition (i.e., teaching) is a necessary perceptual condition [32], yet the two activities are different in nature and call for a different time flexibility in system response. While in music tuition, conflicts and breakdowns are managed with the time flexibility of turn-taking in conversation [33], remote playing demands more extreme responsivity. The immediacy of responses between remote performers is essentially altered by the presence of end-to-end delays, dependent on the signal processing, throughout the whole chain, and network delay [4, 5].

The resulting latency level represents a factor that dramatically affects the NMP experience. It has been shown that

latency values in the range of 20–60 ms degrade the quality of the performance, by causing progressive slowdowns in the performance [19]. The effect of unnatural latencies has been explained in terms of incongruent temporal separation between action and auditory information, with respect to the expected “natural” time delays occurring in co-located performances [16, 17, 34]. Rhythmic clapping tasks have been used to assess the impact of latency (in the range of 3–78 ms) on the performance of two musicians playing in the network. It was found that best results are achieved when end-to-end delays approximate a temporal separation corresponding to a setting where the two subjects are in the same room. It was also observed that unnaturally low latency values (i.e.,  $t \leq 10$  ms) cause an acceleration in the musicians’ tempo, which can be explained in terms of sensory-motor synchronization by anticipation [35].

Indeed, these findings stress the importance of considering NMP as culturally situated practice in which bodily schemas are inherently part of the musical meaning and embodied means through which making sense of the networked space [7]. Both music playing and tuition rely on a significant repertoire of nonverbal forms of communication, and in the chamber music context, the types of body language and mutual cues embody the typical semicircular seating arrangement and the musical interpretation of the shared score. Typically, chamber musicians do not need to stare at each other, yet they support mutual understanding and negotiate the performance through glances and peripheral vision.

It has been argued that live video streams (i.e., full frontal views) of the musicians are essentially superfluous for the sake of the co-performance and that they respond to communication needs towards the listeners/observers, rather than between the performers [8]. While this observation is certainly embraceable, we hypothesize that properly designed spatial elements in remote interactive environments may facilitate the compensation of time-dependent misalignments in the performance and communication. In this respect, the design of the stage and the performance scene implies not only the physical displacement of various equipment in the rooms, but also the choice of appropriate solutions in both visual and sound feedback display [10, 33]. Various off-the-shelf strategies, such as life-size and near-life-size visual display to preserve a coherent distance perception in virtual proximity [36], projections and video loop techniques to support synchronous interaction [37], and the use of spatialized audio to increase the sense of presence [28, 38], have been proposed. The use of Ambisonic techniques has been explored to provide expressive and congruent environments in network music applications [39]. Studies in artificial reverberation rendering, in order to improve the realism, showed that virtual anechoic conditions lead to a higher imprecision than the reverberant conditions, while real-reverberation conditions lead to a slightly lower tempo, compared to the analog virtual reverberation conditions [40]. In general, the degree of spatial coherence needed is much dependent on the specific context and scope, whether the music interaction at hand accommodates loose coupling between performer and instrument, or demands close temporal coupling, with little

tolerance for interruption [41]. A minimum of spatial acuity and especially cross-modal congruence must be preserved, where the quality of the auditory content (i.e., timbral realism) represents a relevant cue through which the musicians negotiate their performance (e.g., agogics and instrumental blending) [28].

More recent research studies investigated the concept of Internet acoustics, which is the ambience resulting from the acoustical loop between internet endpoints, and the implementation of room-like Internet reverberators for collaborative music performances [42]. It should be noted, however, that the strict latency requirements imposed by the chamber music scenario make the adoption of any spatial audio rendering system very challenging. A possible solution could be to follow the route proposed in [43] and treat the rendering system as a distributed one, by taking advantage of a client-server architecture in order to meet the needed latency requirements and satisfy the need for spatial realism. Other works consider the avoidance of spatial realism and instead propose to take advantage of the spatial dissonance as a tool for music composition [44], which, however, is not suitable for the kind of music practice taken into account in this work. These audio-visual spatial aspects, however, are not considered in the present study and will be subject of separate investigation.

Finally, while we refer the reader to the extensive overview on NMP technologies discussed in [3, 5], here we briefly describe the three main alternatives of NMP software, as considered from the InterMUSIC perspective. JackTrip is the application developed by the SoundWIRE research group at CCRMA, to support bidirectional music performances [45]. It is based on uncompressed audio transmission through high-speed links such as Internet2. In the current version, it does not support video transmission.

The LOLA project was developed by the Conservatory of Music “G. Tartini” in Trieste, in collaboration with the Italian national computer network for universities and research (GARR) [46]. LOLA is based on low-latency audio/video acquisition hardware, optimized to transmit audio/video contents through a dedicated network connection. The main drawback is that the application is not open source and not serviceable for generic network connections.

On the other side, UltraGrid is an open-source project, whose application allows audio/video low latency transmission [47]. UltraGrid’s performance is not as effective as the LOLA environment, yet it represents a rather flexible solution for generic hardware use and networks connection and especially allows the development and integration of new functionalities. In the InterMUSIC project, we are currently using LOLA to conduct our experiments, although we are also considering the integration of UltraGrid in the next stages of the project.

*2.2. A Presence Perspective on NMP.* Presence is a multifaceted concept that has been interest of research since the early experiences in telepresence and virtual environments and more in general in computer-mediated environments applications [48]. The “feeling of being there” has been

emphasized as relevant factor in immersive mediated environments and assumed as construct that bridges the immersion quality of the mediating technology with the effectiveness of the mediated experience [49]. Ultimately, several definitions and models of presence have been proposed over the last two decades, with the aim of developing generalizable measures of mediated environment effectiveness and performance. A conclusive demonstration has not been provided yet [26], being the main problematic aspect the (lack of a) logical connection between presence, as an observable phenomenon, and task performance as a function of the user interface [50].

Presence frameworks provide, however, a consistent corpus of protocols, to evaluate the subjective experience in mediated environments, the immersive technology, and the quality of the interaction [26, 49]: validated self-reports in the form of postexperiment questionnaires focus on several presence components, including the individual characteristics of the user (e.g., attentional resources and attitudes), the coherence of the scenario (e.g., the realness and overall consistency of the mediated experience), and the immersion of the system (i.e., the set of sensorimotor and valid actions supported by the tracking system) [29, 51]. In addition, several presence-measuring techniques, behavioral and physiological (e.g., heart rate and skin conductance), have been proposed in order to capture the existence of the presence phenomenon in well-defined scenarios, in a rather objective fashion [25, 29]. Coherence and immersion can be controlled by design, to a certain extent, and developed to best reflect the user characteristics.

In the scope of InterMUSIC research, we make an instrumental use of the structural model of presence proposed by Schubert et al. [24], as resulted from the three-order factor analysis of 246 answers to a 75-item survey of questions taken from several established questionnaires on presence. Presence is defined as the cognitive feeling of being in a place [24]. This definition encompasses a concept of presence as embodied experience arising from the perceived self-location, the sense of agency, and the perceived action possibilities with respect to the specific scenario. The sense of presence is tied to action and active engagement in the (mediated) environment and results from the interpretation of the mental model derived. Schubert’s model of presence is composed of three main groups of components, which describe the subjective experiences of presence (spatial presence, attentional involvement, and realness), the evaluation of the immersive technology (quality of immersion, and dramatic involvement), and the evaluation of the interaction (interface awareness, ease of exploration, predictability, and interaction) ([24], p. 271). While the presence factors refer to the psychological experience, the other two groups pertain, respectively, to the presentation of the stimuli and the properties of the interaction. Although the psychological state and the immersion and interaction factors are separate constructs, Schubert claims that the relationship between the immersion and the sense of presence is not one-to-one; instead, it is the bodily and cognitive processes that mediate the impact of immersion on the experience of the psychological state of being in a place.

We agree with this view which fits the specific InterMUSIC scenario. This model of presence strongly resonates with the inherent sensorimotor nature of music communication and performance, where intentionality, corporeal articulations, interpretation, and perception of physical/sound energy represent the pillars of musical signification practices [52]: we see the NMP environment as a mediation technology layer which displaces the coupling of first-person descriptions (i.e., the mental representation of an intended musical gesture) and second-person descriptions (i.e., the corporeal articulation of the intended act), as an effect of the unavoidable alterations in the third-person description (i.e., the conditioning of physical/sound energy as returned by the networked auditory-visual capturing and display). In other words, we conceptualize the NMP environment as an instrument, on which chamber music instrumentalists project their direct involvement and mimetic skills.

On the other side, we also integrate the meta-analysis of presence models proposed in the literature, done by Skarbez et al. [29]. Presence is defined by the authors as the perceived realness of the mediated experience. The embodied, sensorimotor perspective proposed by Schubert (i.e., the cognitive feeling of being in place) is refocused on the balance of the three illusory experiences of (1) being in a place, in terms of action possibilities afforded by the system's immersion (the place illusion [51]); (2) the fidelity and plausible behaviors which make the apparent scenario to be coherent to the users/musicians (the plausibility illusion); and (3) the awareness of the copresence of another sentient being (e.g., the remote performer) and the extent to which a minimum degree of interaction gives rise to social behaviors (the social presence illusion).

The last point is worth a brief discussion, since playing music in ensemble is an essentially social and sharing activity, which entails complex sociocultural, coordination, and cooperation efforts and concerns [53, 54]. According to Skarbez, the social presence illusion refers to the "illusory (false) feeling of being together with and engaging with a real sentient being" ([29], p. 96:5): the term communicative immersion is proposed to trace those characteristics of the system that afford the transmission of communicative signals and ultimately the priming of the feeling of a warm, sensitive, and sociable medium (i.e., the *communicative salience*). Given the peculiarity of the chamber music NMP scenario, one may argue that musical connectedness and flow [55] represent the optimum illusory states of social engagement that skilled instrumentalists may expect to experience. In this respect, auditory immersion, especially in terms of spatial qualities and fidelity of sound, becomes crucial to maintain the sensorimotor loop and engage in plausible musical behaviors (e.g., instrumental blending) [28].

Taken together, we keep Schubert's presence components ([24], p. 279) and we group them according to the rationale of Skarbez's conceptualization ([29], p. 96:24). In addition, we remove the two factors of exploration and dramatic involvement, since they are not particularly meaningful in the frame of the chamber music scenario. Our main interest is not to actually define or model presence,

which is not the focus of InterMUSIC research; instead, we make use of presence studies as an opportunity to examine the chamber musicians' behaviors in the network, with the aim of providing a further perspective on NMPs.

The presence experience is composed then by three major groups of constructs, that is, components:

- (i) The spatial-constructive and attention facets of the experience of being there are, respectively, operationalized into *spatial presence* and *involvement* components
- (ii) The coherence of the scenario, that is, the reasonableness of the events primed to the user, is reflected into the *perceived realness* and *predictability* components
- (iii) The quality of the system's technology is distilled into the two components of the *interface awareness*, that is, distraction factors and the *quality of immersion*

Spatial presence (SP) refers to the emerging relation between the mediated environment as a space and the user's own body. The sense of "being in place" is tight to the role of the active body in constructing a spatial-functional model of the surrounding environment. The interpretation of the mental model emerges as cognitive representation of action possibilities in the actual environmental conditions, meshed with patterns of actions retrieved from memory. Spatial presence, that is, the place illusion (PI), implies a suspension of disbelief, about *how* the world is perceived, which can be ascribed to the immersion of the system, whereas the plausibility illusion (PsI) represents the orthogonal construct, about *what* is perceived [51]. In the realm of NMP, and particularly in the chamber music scenario, PI and PsI rather mirror the fidelity and consistency of the mediated environment, that is, the degree of real world emulation, in terms of coherent behaviors afforded. Indeed, the constraints of the basic, screen-based NMP setup may conceivably hinder the emergence of a spatial mental model of the networked environment.

Involvement (INV) or flow is a recurring concept in the presence literature and retains the attention side of the presence experience. The involvement component refers to the user's active engagement and focused attention and reflects "a psychological state experienced as a consequence of focusing one's energy and attention on a coherent set of stimuli or meaningfully related activities and events" [56]. Spatial presence and involvement are distinct and yet complementary psychological states, according to which individuals develop meaningful representations of a situation, by constructing the spatial mental model while actively suppressing conflicting sensory inputs. In the context of NMP chamber music practice, the involvement is reflected in the relative concentration on the real and the remote environments, in terms of focused allocation of attentional resources, concentration, and action awareness. In this respect, the degree of involvement is tightly connected to the coherence and consistency of the experience with the expectations from the real world.

*Perceived realness* (REA) encompasses reality judgments with respect to the meaningfulness and coherence of the scenario, as a function of the system's ability to provide stimuli which are internally consistent. From the perspective of NMP systems' development and design, it implies to consider appropriate and reasonable strategies to prime the musicians to expect certain types of behaviors and hence to develop adaptation and familiarity. Perceived realness and predictability provide a clue of the overall consistency and meaningfulness of the remote environment and represent a subjective evaluation of the interaction.

*Predictability* (PRED) refers to the possibility to anticipate what will happen next, in terms of activation of motor representations as a consequence of perceiving while playing. These include not only self-generated actions, but also the actions (and effects—which and when) produced by others [21]. The predictability in the chamber music performance is supported by the musical interpretation of the shared score. Similarly to the realness component, the predictability of the NMP system means that the environment should be consistent and undeviating in its behavior, in order to facilitate the musicians' adaptation. Sensory breadth, that is, the amount of sensory channels simultaneously presented, is expected to affect predictability, and hence the coherence of the scenario. Sensory breadth can be modulated in depth, that is, resolution. Higher resolution in the sensory modality essential to the task is expected to lead to more presence.

*Interface awareness and quality* (IA) take in account distraction factors, that is, the obtrusion of control and display devices in terms of interference in the task performance. This factor gives a different viewpoint on the selective attention ability and focused concentration of the user and in general provides a clue of the mastery of the interface in the specific activity at hand. In the NMP scenario, the interface quality refers to the spatial staging, and its acceptance may increase through practice and familiarity. Eventually, a high quality interface may have an impact on the performers' involvement.

*Quality of immersion* (QI) is a component related to the presentation of the stimuli and can be defined as the set of valid actions supported by the mediating environment [29]. It has been shown that high levels of immersion do not necessarily lead to a higher level of presence experience and that spatial immersive features can be more effective than higher quality sensory contents in leading the users to construct a spatial mental model of the environment (i.e., the place illusion) [49]. The immersion emerges as a quality of the system's technology, reified in the vividness, that is, the sensory breadth and resolution and the interactivity of the system, including the update rate and association of controls and displays, and the range of the interactive attributes available for active search and manipulation (e.g., the binaural sound rendering as a function of the performer's head motion).

It is evident that all these components are intertwined, and they mutually affect each other. The design and development of remote interaction environments may favour one or a combination of these constructs, depending on the

characteristic of the application. NMP environments are constrained by a music making task which is very specific and demanding, especially in the context of chamber music practice. Put in design terms, this represents an advantage, potentially providing clearer user requirements. Much of expected outcomes of the current research are aimed at collecting and organizing the instrumentalists' expectations. For example, if the goal of the music practice at hand is to actually rehearse a performance in the network, the induction of a state of involvement or flow might be preferable. Training applications may rather focus on provoking a plausible effect of social presence. Time continuity and connectedness become crucial. On the other hand, when the goal is to transfer the training to the real world, the coherence and fidelity of the scenario are likely most important.

The objective of this research track of the InterMUSIC project is to understand and conceptualize first the most relevant factors affecting the experience of studying and rehearsing music in a networked environment [57]. The experimental research is aimed at framing the design problem of an online learning environment for master students of chamber music. The overall goal is to define design guidelines for the staging of remote rooms in higher music education institutions, including the NMP system's technology.

*2.3. The Framework Conceptualized.* As shown in Figure 1, a *performance* occurs when two or more *subjects* musically interact together through a *medium*, in an *environment*. The performance is the entity at highest hierarchical level and may assume two main configurations, namely, a *performed music composition* or a *taught lesson*. The music performance can be of two main types, a *rehearsal* or a *concert*, and both involve the interaction of at least two peers, that is, the subjects are both musicians. In the lesson configuration, the performance involves subjects with different roles, a teacher and at least one student. When subjects interact in the same room, we have a *local performance*. However, in the InterMUSIC scenario, we consider either the case of geographically displaced subjects (*networked performance*), or the case of a *mixed performance*, when two or more subjects are accommodated in the remote rooms. In this respect, the spatial property of the performance is a function of the *medium*.

When subjects interact by means of a *physical* medium, that is, simple air propagation, the performance is local. In the case of networked performances, the medium is a *networked medium* and includes the Internet connection and the NMP software/hardware equipment to connect the subjects. Mixed performances involve both physical and networked media.

The environment is the space hosting the subjects, with its own specific timbral and spatial properties, i.e., the acoustics, the staging, and the subject(s) displacement in the room. In the case of networked and mixed performance, environments with different characteristics are potentially involved. Given a subject, we define the environment where

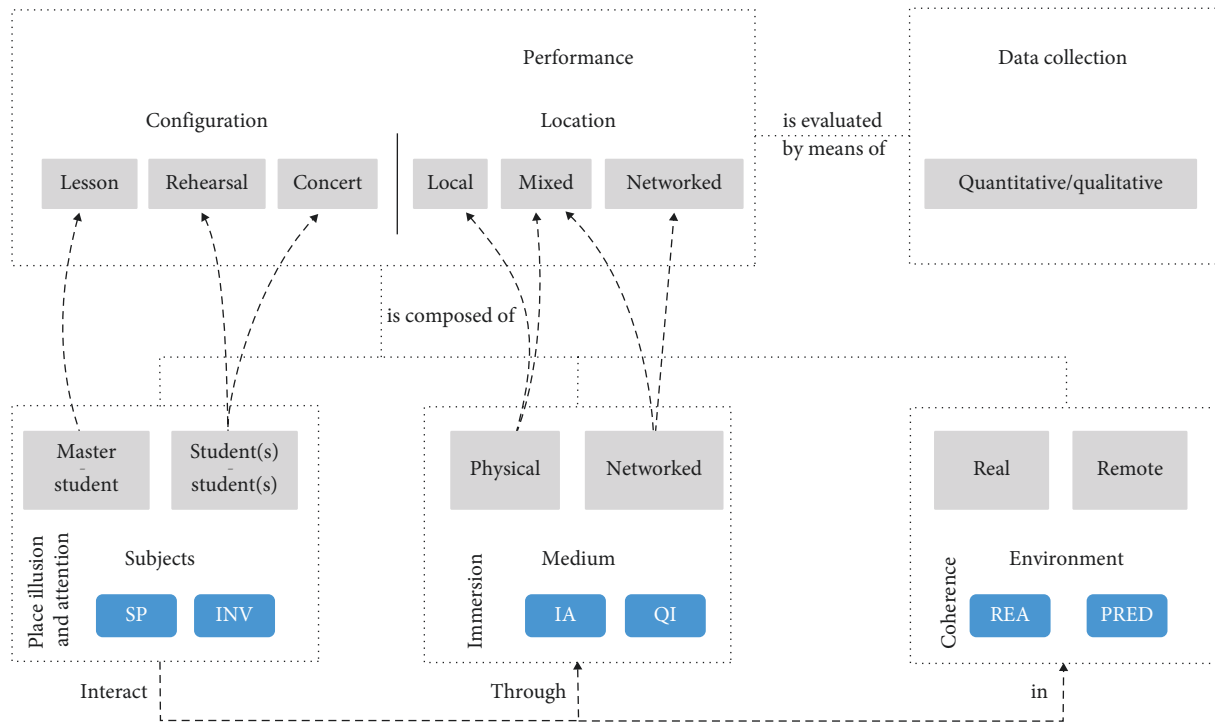


FIGURE 1: The conceptual framework for NMP research in the chamber music practice and learning scenario.

she is playing as the *real environment* and the representation of the environment relative to the geographically displaced co-performer as the *virtual or remote environment*. The performance unfolds in the situatedness of the networked space emerging from the real and the virtual environments. As an example, Figure 2 shows the typical basic NMP staging wherein the real environment of each musician is returned to the other co-performer in the form of a remote audio-video representation. The perceived coherence of the NMP scenario results from the overall congruity of the two environments.

Data collection is essential to the analysis of the performance as a whole and in its constitutive facets. Diverse approaches may come into play to capture the nuances of the music making experience in the networked environment, from the acquisition of multimodal signals to the collection of video-ethnographic observations and self-reports. The first distinction concerns the performance configuration, whether the networked space is meant to host a concert or a lesson. These two types represent the edges of continuum, wherein music interaction (i.e., playing) and gestural and verbal communication clearly have a different relevance and purposes. A performance can be described by date and time, location(s), type, metadata, MIDI or MusicXML symbolic representation, and the score (including its musicological analysis).

The concert type relies entirely on forms of communication, coordination, and leadership, based on musical interpretations previously negotiated and equally shared according to the score indications [58]. These kinds of performance represent a proper task, which is essentially addressed to the external world and experienced from a

second-person perspective, that is, a listener/observer. The concert is the configuration commonly investigated in NMP research and technological development [4, 5, 59]. Multimodal signal acquisition includes the audio recordings from both the remote environments, to compute several measures on the quality of the performance [19], gaze tracing and eye tracking to annotate the visual search of the musicians [60], and the biometric response of the performers for distress estimation [61].

In the lesson configuration, the teacher-student musical interaction rather follows the turn-taking mechanisms of conversations. Typically, musical coordination in the local lesson is maintained through verbalizations, visuospatial gestures, bodily postures, and especially joint annotations of the shared score [33]. However, when the medium is networked, the intrinsic disruptions may affect the management of verbal and musical interruptions and overlaps, despite the higher time flexibility of networked conversational turns [17]. Networked and environmental audio-video recordings of the teaching session can be annotated and parsed in salient conversational turns in order to conduct ethnographic and conversational analyses [62]. The rehearsal represents an intermediate situation, between the concert and the lesson. Synchronous and instantaneous music interaction intertwines with conversational turn-taking mechanisms, for the purpose of practicing or preparing a concert [63].

Self-reports and questionnaires are used to collect information on the quality of the performance, as perceived by the subjects, and in general on their musical experience in the networked space. We make use of a presence questionnaire, whose items reflects constructs that can be

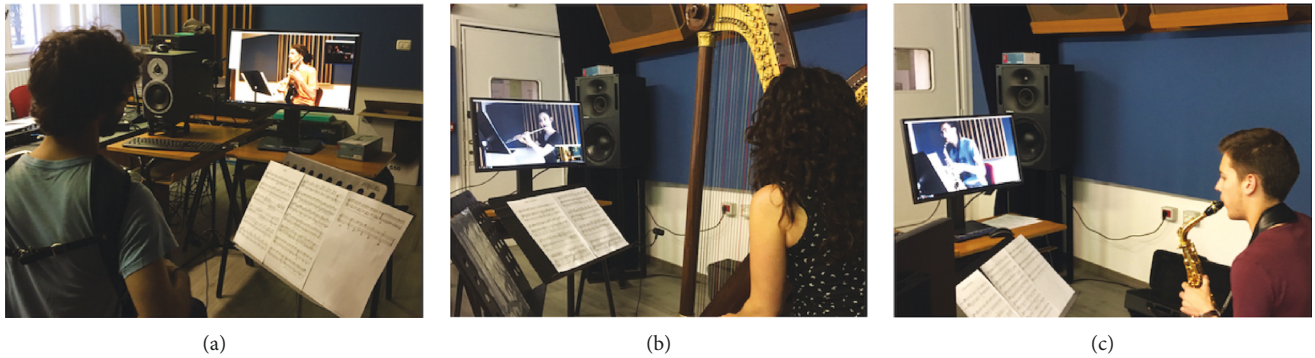


FIGURE 2: Instrumentalists in Room 1, with the frontal view (on screen) of the co-performer displaced in Room 2.

selectively ascribed to the musician herself (spatial presence (SP) and involvement (INV)), the medium (interface awareness (IA) and quality of immersion (QI)), and the environment (realness (REA) and predictability (PRED)).

Subjects are identified by name, age, experience, and musical background. In general, the user types considered in the InterMUSIC scenario are musicians with at least 5 years of academic music practice. The instrument is a relevant feature of the subject, and content-based analyses are used to integrate the user's profile [19]. Previous experiences with virtual environments and motivation are also relevant individual information to account an attitude to adaptability and selective attention.

The environment data include the spatial staging and acoustic properties of both real and remote places. Depending on the types of performance (i.e., concert, rehearsal, and lesson), the environment may exhibit different requirements [10, 62]. In this respect, properties of this entity are the audio/video capture and rendering devices configurations and their degree of experienced coherence with respect to the specific type of performance and scenario. By configuration, we mean the effects of design choices in the electroacoustic chain and signal processing, for example, the types of microphones according to their cost benefit [64], the moulding of a shared space based on Internet acoustics [42, 65], or acoustic scene analysis [66] and spatial rendering techniques over arrays of speakers [67] or headphones [68]. Design choices in visuo-spatial representations, i.e., video capturing and rendering, are mostly constrained by the requirements of the NMP system architecture employed [46].

Data relative to the medium essentially concern the system architecture, hardware and software, and network characteristics, that is, bandwidth and latency. From the subjects' viewpoint, the physical and the networked media affect in a different way the temporal separation between the remote performers. In the concert and rehearsal types, dynamic alterations of the temporal separation, caused by the network latency, result in disruptions of the mutual coordination of the ensemble. The medium features in the actual NMP environment qualify the immersion of the system's technology.

In the following Section, we introduce the pilot experiment [69], which represents an ongoing tool for reflection and conceptualization of the NMP chamber music practice and learning scenario.

### 3. Experimental Study

The objective of the study is to conceptualize the disruptive effects of network delay and interruption on time management, communication mechanisms, and mutual understanding between remote performers and frame them in the subjective experience of playing together in the networked space. In other words, a mixture of quantitative evaluation, based on objective performance quality metrics, and qualitative assessment and observations are aimed at putting in the foreground the users' needs and system requirements. The outcomes are expected to provide relevant design implications in the staging of classrooms dedicated to remote music practice and teaching. The current study focuses on the temporal dimension of the networked experience, while keeping constant the spatial dimension, i.e., strategies in auditory and visual spatial representation, which will be subject of forthcoming inquiry.

*3.1. Method.* The pilot experiment took place at the Conservatory of Music "G. Verdi" of Milano, in two dedicated rooms, equipped with direct network connection and all the necessary facilities. We asked couples of musicians recruited from the conservatory to perform a short exercise, under diverse conditions of network delay. A qualitative assessment through questionnaires on the sense of presence and the perceived quality of the performance [24] was combined with quality metrics of the objective performance [19]. The scores of the stimuli and the presence questionnaire are publicly available.

Ten volunteers (five duos, five males, five females, age ranging from 14 to 29, average age = 21.9 years, SD = 4.7) were recruited from the class of chamber music practice. They were all musicians with at least five years of academic musical practice. Each duo had already a familiarity of minimum two weeks of rehearsal. Table 1 reports the instruments and the performers' location, in rooms 1 or 2 (see Figure 3).

*3.1.1. Presence Questionnaire.* Self-reports in the form of postexperience questionnaires are common measurement means used in presence studies. Several questionnaire models have been proposed, which respond to the different models of presence, variously addressing specific constructs



TABLE 1: Instruments and their location in Room 1 or Room 2.

Room	Couple A	Couple B	Couple C	Couple D	Couple E
1	Mandolin	Accordion	Percussion	Harp	Alto sax
2	Mandolin	Guitar	Percussion	Flute	Alto sax

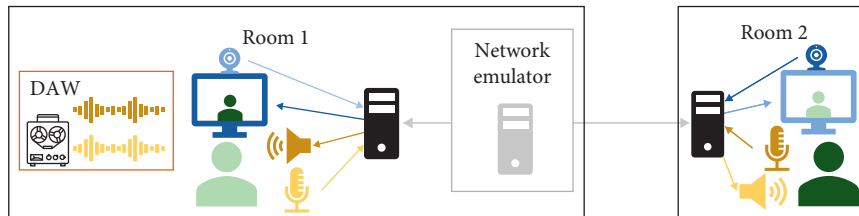


FIGURE 3: Setup of the pilot experiment, where two musicians perform together through a network emulator.

and situations in virtual reality (see [29] for a comprehensive review). One advantage of using published presence questionnaires is that they are the result of extensive validation, which makes them sensitive and reliable, yet the main drawback is that they are intrusive, not concurrent with the experience under inquiry, and prone to subjective (mis)interpretations of potentially difficult constructs. Furthermore, a clarification must be given: presence represents a complex phenomenon whose observable and measurable existence is still under debate, that is, the availability of questionnaires and related data does not demonstrate *per se* an effective measure of presence. In other words, attention must be paid in relying on questionnaires only to draw conclusions on the actual manifestation of the presence experience when interacting in a virtual environment [50]. In the InterMUSIC scenario, we use the information extracted from content-based analyses of the performances to derive observable patterns in music interaction and communication, and we approach the presence perspective to generate hypotheses on the relevant components that may affect the experience of two or more instrumentalists engaged in attending a master class, or rehearsing.

The questionnaire used in the study was constructed by merging three reference questionnaires on presence and selecting the most appropriate items with respect to the specific activity of performing music in the networked space (e.g., questions pertaining to navigation, objects motion, and manipulation in the virtual reality are not applicable to the case of NMP). In particular, we referred to the Witmer-Singer Presence Questionnaire [56, 70], and its revision by the UQO Cyberpsychology Lab [71], and especially the IPQ Igroup Presence Questionnaire by Schubert et al. [24].

Items were partially rephrased and adapted to the musical context and language. The resulting close-ended, 7-point Likert scale questionnaire was edited in Italian. The questionnaire is split in two main parts: a general post-experiment 27-item questionnaire organized in five sections and a postrepetition questionnaire with five questions extracted from the general one. We report the postexperiment questionnaire in Table 2 (with the median, mean, and standard deviation of the answers). The questionnaire is

devised around the three main constructs of presence, with an additional group of items focused on the performance:

- (i) Involvement and place illusion (Section 1): questions encompass the focused concentration and attention allocation to the music performance (Q1.2 and Q1.4–7) and the illusory feeling of connectedness with the remote environment (Q1.1 and Q1.3). It is hypothesized that in the NMP scenario, the attentional resources are more salient than the feeling of place illusion. For instance, while it can be argued that the sense of “being there” (i.e., the SP item, Q1.1) is of less utility in the screen-based NMP interaction, we retained this item exactly to question the suspension of disbelief about the networked space marked by the frontal views of the performers.
- (ii) Coherence (Section 2): this section reflects the perceived coherence of the scenario and includes the subscales of predictability (Q2.1–2, Q2.4–5, and Q2.7–8) and realness (Q2.3 and Q2.6).
- (iii) Immersion (Sections 3 and 4): the immersion quality of the system’s technology is retained in Sections 3 and 4, which, respectively, includes questions relative to distraction factors (interface awareness, Q3.1–5) and the vividness and interactivity of the NMP environment (Q4.1–3). The last set of questions is focused on the effect of the visual and auditory representations as a whole (i.e., screen-based frontal view and monoaural reproduction of the close take of the instrument) on the availability of various information, such as eye contact, foot tapping, breath attack, and instrumental blending, which in turn affect the musicians’ involvement.
- (iv) Quality of the performance (Section 5): this section is dedicated to the subjective assessment of the quality of one’s own performance (Q5.1–4).

Five items with a peculiar emphasis on the experience of delay have been extracted from the general questionnaire and included in the postrepetition questionnaire (items are emphasized in *italics* in Table 2).

TABLE 2: Statistics of the answers to the postexperiment questionnaire.

Question	Median	Mean	Std
<i>(1) Involvement and connectedness</i>			
(1.1) In the remote environment, I had a sense of "being there."	3.5	4.17	1.77
(1.2) The sense of playing in the remote environment was compelling.	5.0	4.50	1.50
(1.3) I had a sense of playing in the remote environment, rather than performing something from outside.	4.0	3.80	1.17
(1.4) How aware were you of the real world surroundings around you during the performance?	4.5	4.00	1.91
(1.5) How completely were you able to actively survey the musical environment using vision?	4.0	3.83	1.57
(1.6) How completely were you able to actively survey the musical environment using audition?	5.0	4.67	0.94
(1.7) The delay affected the sense of involvement.	4.5	4.00	1.63
<i>(2) Coherence</i>			
(2.1) The musical interaction in the remote environment seemed natural.	5.0	5.00	0.82
(2.2) The environment was responsive to actions that I performed.	4.0	4.00	1.83
(2.3) How much did your musical experience in the remote environment seem consistent with your real world experiences?	4.0	4.40	1.36
(2.4) I was able to anticipate the musical outcome in response to my performance in the remote environment.	5.0	5.00	1.00
(2.5) The environment was responsive to actions performed by my partner.	4.5	4.25	1.48
(2.6) How realistic did the remote environment seem to you?	4.5	4.33	0.75
(2.7) I was able to anticipate the musical outcome in response to the performance by my partner in the remote environment	4.0	3.83	1.21
(2.8) It was difficult to cope with the distance performance.	4.0	3.83	1.57
<i>(3) Interface awareness and quality</i>			
(3.1) How well could you concentrate on the music performance rather than on the mechanisms required to perform?	5.5	5.00	1.53
(3.2) How aware were you of the display and control devices/mechanism?	4.0	4.00	1.10
(3.3) How much did the visual display quality interfere or distract from performing?	3.0	3.17	1.21
(3.4) How much did the auditory display quality interfere or distract from performing?	6.0	5.17	1.46
(3.5) How much delay did you experience between your actions and expected outcomes?	4.5	4.33	1.60
<i>(4) Quality of the immersion</i>			
(4.1) The visual representation made me feel involved in the remote environment.	3.0	3.33	0.94
(4.2) The auditory representation made me feel involved in the remote environment.	4.0	4.00	1.73
(4.3) I felt involved in the remote environment experience.	5.0	4.60	0.49
<i>(5) Quality of the music performance</i>			
(5.1) How quickly did you adjust to the experience of playing in the remote environment?	4.5	4.17	1.67
(5.2) It was easy to cope with the delay to adjust the quality of the performance.	3.5	3.33	1.25
(5.3) How proficient in remote music playing did you feel at the end of the experience?	5.0	4.67	1.89
(5.4) The delay affected the quality of my performance.	5.0	4.33	1.80

Postrepetition items are highlighted in italics.

**3.1.2. Objective Quality Metrics.** A content-based analysis was performed on the audio recordings of the duos' performances, in order to derive an objective description of the quality. In the literature, researchers have proposed different metrics, related to the rhythmic trend of the performances [5]. Figure 4 provides an example of the annotation procedure that leads to the computation of these metrics. Given the stimulus represented by a score (Figure 4(a)), the first step is to annotate, in the recordings, the instants when an onset occurs on the beat, as  $t_1, t_2, \dots, t_N$  (Figure 4(b)). In order to address the issues of beats occurring without an onset, e.g., because of a four-quarter note, we also annotated the amount of beats occurring between the  $n$ -th instant and the following ( $t_n$  and  $t_{n+1}$ ) as  $v_n$ . In Figure 4(b), for example, the onset occurring at  $t_{n+3}$  is followed by a two-quarter pause, hence the onset at  $t_{n+4}$  occurs after three beats and  $v_{n+3} = 3$ . We convert the set of annotations  $(t_1, v_1), \dots, (t_N, v_N)$  to the tempo samples (in BPM)  $\bar{\delta}(n) = (60 \cdot v_n) / (t_{n+1} - t_n)$ .

Following the previous example, we show the final BPM annotation in Figure 4(c); we apply a 5-second moving average filter to  $\bar{\delta}(n)$  to compensate the variance due to the musical agogics and the resulting imprecision of the manual annotation.

From the set of  $t_n$ ,  $v_n$ , and  $\bar{\delta}(n)$ , we can compute several metrics [5] related to the tempo slope or the asymmetry between the two performers.

The *tempo* slope  $\kappa$  provides a compact descriptor of the tempo trend as the slope of its linear approximation. In particular  $\kappa = 0$  when the tempo remains steady for the whole performance, and it assumes positive or negative values in case of acceleration or deceleration, respectively.

The *asymmetry*  $\alpha$  provides a metrics of misalignment between the two performers, and it is strictly related to the beat and the score. Let us define  $t_n^{(A)}$  and  $t_n^{(B)}$  as the beat instants, corresponding to the execution of the two performers  $A$  and  $B$  playing during the same performance. Given the score of the performance, we can define a set of pairs  $(n^{(A)}, n^{(B)})$  related to the beats common to both performances. This process is shown in Figure 5, where two example measures are shown, with the indications of the quarter onsets corresponding to  $n^{(A)}$  and  $n^{(B)}$ . Not all quarter onsets are comparable; specifically, in the case shown in Figure 5, it is possible to compute the misalignment only between the annotated time instants  $t_n^{(A)}$  and  $t_n^{(B)}$  corresponding to the notes indexes  $\{(n^{(A)}, n^{(B)})\} = \{(39.2^{(A)}, 39.2^{(B)}), (40.2^{(A)}, 40.2^{(B)})\}$ . Now, we can define the intersubject time difference (ISD) between  $A$  and  $B$  as the time distance between the two instants:

$$t_n^{(AB)} = t_n^{(A)} - t_n^{(B)}. \quad (1)$$

If  $t_n^{(AB)} < 0$ , it means that in that particular beat, the performer  $A$  has anticipated performer  $B$  and vice versa, while  $t_n^{(AB)} = 0$  indicates the performers played in the same instant. In order to obtain the asymmetry, we average the ISDs through parts of the performance, or through the whole performance, to have a more global descriptors of the interaction between the two performers. Let us define for our convenience the set of common beats  $\mathcal{N} = \{(n^{(A)}, n^{(B)})$

$\},$  of size  $|\mathcal{N}|$ ; then, we can write the asymmetry as

$$\alpha^{(AB)} = \frac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} t_n^{(AB)}. \quad (2)$$

It must be noted that the value of the ISDs, and hence of the asymmetry, depends on the point of the recording. In our case, the recording was performed in room corresponding to the subject  $A$  (i.e., Room 1, in Figure 3). We can infer the ISDs of subject  $B$  using the information on the two-way latency  $\lambda$ , as

$$\begin{aligned} t_n^{(BA)} &= t_n^{(B)} - t_n^{(A)} = \left( t_n^{(B)} - \frac{\lambda}{2} \right) - \left( t_n^{(A)} + \frac{\lambda}{2} \right) \\ &= t_n^{(B)} - t_n^{(A)} - \lambda = -t_n^{(AB)} - \lambda. \end{aligned} \quad (3)$$

Note that this may lead to some contradictory behaviors. Suppose, for example, that  $t_n^{(AB)} = -25$  ms (where we neglect the  $n$  for the sake of clarity), which means that performer  $A$  is anticipating performer  $B$  by 25 ms. However, if the two-way latency is  $\lambda = 50$  ms, this means that in the room where  $B$  is performing, we measure  $\alpha^{(BA)} = -\alpha^{(AB)} - \lambda = -25$  ms, hence performer  $B$  is also anticipating performer  $A$ .

The analysis of asymmetry, therefore, must be conducted analyzing both sides of the medium in order to draw meaningful considerations. The asymmetry  $\alpha^{(BA)}$  corresponding to the point of view of the room where performer  $B$  is playing, can be computed as

$$\alpha^{(BA)} = \frac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} t_n^{(BA)} = -\lambda - \frac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} t_n^{(AB)}, \quad (4)$$

where we use equation (3) for the second part of the formula.

The tempo slope  $\kappa$  and the asymmetry  $\alpha$  enable us to consider two different aspects of the interaction between the musicians during the performance with an objective formulation.

**3.1.3. Apparatus.** The research activity took place in two dedicated rooms with direct connection at the Conservatory of Music of Milano. The rooms are two acoustically treated studios and are located in two different floors of the building with no direct sound interference. The experimental setup is shown in Figure 3.

The performance in each room was captured by means of one Audio Technica ATM350 cardioid condenser clip-on microphone applied to the instrument (monoaural acquisition) and a low latency Ximea MQ13CG-E2 USB 3.1 Gen 1 camera (with a Tamron TA-M118FM08 lens) placed in front of the performer and rendered by means of one Dynaudio BM5 mk3 7 "studio monitor loudspeaker (monoaural rendering) and a 27" 144 hz Asus ROG video monitor in the same frontal position of the camera. This basic staging reflects the current usage in remote music practice and tuition [46, 62]. Figure 2 shows the staging in Room 1. We consider this spatial arrangement as baseline for further investigations, assuming that it may represents the worst case scenario.

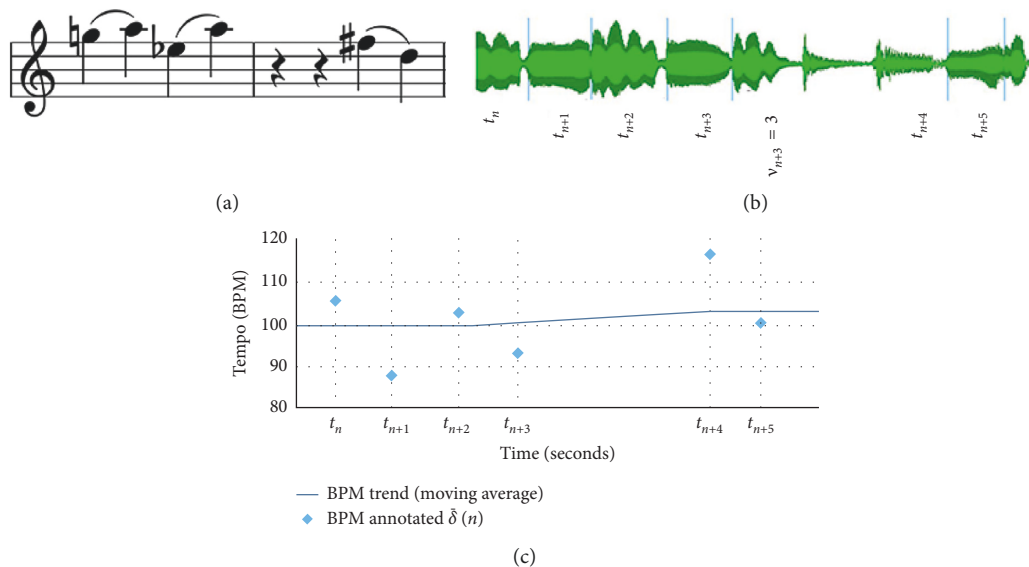


FIGURE 4: Representation of the annotation procedure for computing objective content-based metrics. (a) The stimulus. (b) The signal annotated with  $t_n$  and  $v_n$ . (c) The computed BPM trend  $\bar{\delta}(n)$ .

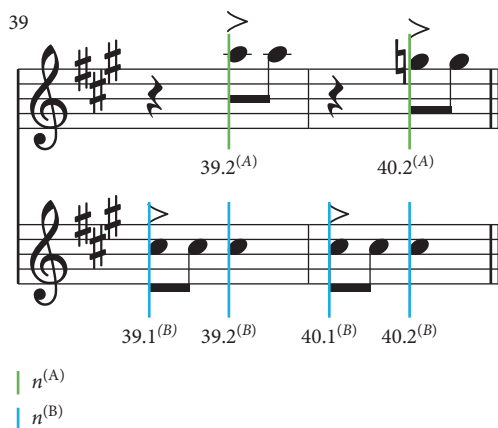


FIGURE 5: Example stimulus, with notes corresponding to quarter onsets suitable for the computation of the ISD.

The hardware equipment was connected to two computers, namely, two high-end Intel/Nvidia powered workstations with i7 esa/octa core processors, using PCIe audio cards, running Windows 10 OS, according to both LOLA and UltraGrid hardware and software requirements. The computers communicated together through a Gigabit Cat6 ethernet connection to a common server. The server, equipped with two Gigabit ethernet interfaces, acted as a Network Emulator to add a fixed delay to both audio and video streams. The effect of jitter, i.e., stochastic variation of the delay, will be considered in future experiments [19]. The server was placed in Room 1 to be easily accessible during the tests and to ease troubleshooting in case of network issues.

The audio output and audio input of the performer in Room 1 were redirected to a Digital Audio Workstation to record the performance (from the perspective of Room 1). These recordings were used to compute content-based metrics of quality, as described in the previous subsection.

Even when the latency was set to zero on the network emulator, a certain amount of it was still present, namely, the processing time, caused by the processing chain composed of analog-digital conversion, acquisition, preparation of packets, network stack, etc. In our experiment, we estimated the processing time by short-circuiting the output to the input in the equipment in Room 2, and we generated an impulse from Room 1 and recorded the delayed output, when the network delay was set to 0. We estimated the two-way processing as 28 ms, which can be seen (acoustically) as two musicians playing four meters apart.

**3.1.4. Stimuli.** The stimuli proposed to the musicians consisted of a score composed by one of the authors and were designed to take into account diverse basic structures of musical interaction in classical chamber music, with respect to time management and communication strategies. We have considered the chamber music duo as the basic instrumental group to approach different kinds of musical interaction. The rationale of the stimuli (i.e., the scores proposed to the duos) concerns simple, yet constraining aspects of synchronicity in musical time, as established in western music tradition, that is, the tight link between the musical dimensions of rhythm, melody, and expression. The objective was to direct the performers towards a complete musical interaction, leaving out any form of purely technical or quantitative test.

In this respect, we looked at Bartók's *Mikrokosmos* piano pieces [72], which represent a valuable methodological compendium of exercises in meaningful *rhythm-melody-expression* relationships: the didactic and technical purpose is immediately connected to the musical sense. In Table 3, we pinpoint eight types of musical structures which combine diverse expressive relationships of rhythm, melody, and expression (this expression includes musical markings

TABLE 3: Types of musical structures in rhythm-melody-expression relationship and examples of their combination present in the exercises in the score (leftmost column).

Ex.	Rhythm	Melody	Expression (dynamics, articulation, agogics)
3a	Homorhythm	Octave or unison	Static
8a	Homorhythm	Opposite direction	Static
5	Heterorhythm	Opposite direction	Alternation
7	Heterorhythm	Homodirection	Climax
6b	Phasing	Octave or unison	Static
4b	Slicing	Slicing	Dynamic
2b	Imitation	Imitation	Imitation
3b	Ostinato	Pedal tones	Static

regarding dynamics, articulation, and agogics). The leftmost column reports a few exemplary exercises as referenced in the score, to ease the reader's understanding. From the perspective of rhythm, a musical structure can be *homorhythmic* or *eterorhythmic*, whether the articulation for each part is, respectively, the same and coincident or different. A *phasing* articulation occurs when the parts have the same rhythm, but their alignment is characterized by a short time delay. A *slicing* articulation refers to a musical phrase in which the rhythm as a whole is alternatively split between the parts. *Imitation* and *ostinato* refer to common repetition strategies in musical practice. The melodic structures essentially reflect the pitch directionality as articulated in the parts. Finally, the expression articulations between the parts can be *static* when there are no variations of expression markings, *alternated*, or arranged in a *climax*.

Figure 6 shows an example of homorhythmic, unison melody, with static expression relationship, respectively, extracted from the scores for flute and harp (left) and percussions (right). The score is internally composed of 14 exercises which represents diverse combinations of musical articulations and structures and is reported in Figure 7: These can be grouped in two main types: 9 exercises mostly emphasizing a synchronicity in rhythm articulation (light gray) and 5 exercises mostly centered on synchronicity in melodic and expression articulation (dark gray). The musical stimuli have a duration of 3 minutes and a reference tempo of 112 BPM.

**3.1.5. Procedure.** Each duo had to perform the exercise, under six different conditions of emulated network delay, reported in Table 4. The minimum amount of latency, due to the two-way processing, was estimated in 28 ms, hence representing the first condition in Table 4. The sequence of six conditions was randomized for each duo. Before each session, each duo was briefed in Room 1, and the task was introduced, within the scope of InterMUSIC, without disclosing any information about the six network delay conditions. The score of the exercise was explained and handed out. Participants were informed about the duration of the exercise and the approximately overall duration of the experimental session (90 min) and introduced to the questionnaire on presence. They were asked to fill in the 5-item

questionnaire after each single repetition and the general 27-item questionnaire at the end of the whole session. Further comments were collected at the end of the test.

After the brief, the musicians settled in their respective room, and a 15-minute rehearsal was devoted to adjust their positioning, framing, and volume levels in order to provide a comfortable environment. In addition, they could rehearse and get acquainted with the score.

## 4. Results and Discussion

Being the experimental campaign for the pilot study limited to a reduced number of collected sessions ( $S = 5$ , for a total of 10 participating musicians), we provide the reader with a narrative of the information that we are able to extract from the analysis of the subjective and objective evaluations. In fact, we consider only three sessions out of five, that is, the couples C (percussions), D (harp/flute), and E (alto sax), since two sessions were either not fully completed or deeply biased (couples A, mandolins, and B, accordion/guitar). That being said, the analysis of the results shows the usefulness of the proposed framework as means to conceptualize and systematize the diverse aspects that affect the quality of a networked music learning scenario. The reduced sample size does not allow to stress any conclusive result; nonetheless, this pilot generated valuable methodological implications and hypotheses concerning the relevance of latency in NMP and the overall complexity at play.

**4.1. Subjective Evaluation.** The subjective evaluation is aimed at understanding the sense of presence in NMP performance and at providing a qualitative, yet reliable measure of NMP interaction and system, via presence constructs. In the current study, we consider the network latency as the main variable affecting the subjective experience of playing together in the networked space, that is, the focused concentration, the coherence, and immersion of the overall experience in the real and remote environments, in addition to the perceived quality of the performance.

The visual inspection of Table 2 returns a picture of an overall experience which is mostly perceived as puzzling, and yet intriguing. As it was expected, the musicians felt mostly neutral to the feeling of a place illusion (Q1.1 and Q1.3), and yet the sense of playing in the remote environment was experienced as sufficiently compelling (Q1.2). Despite the low effectiveness of the overall environment in generating a meaningful sharing experience, the musicians were able to concentrate on their performance (Q3.1). We hypothesize that if any sense of being together was felt, this was due to the inherent social characteristics of the music making task. Indeed, the overall quality of the display is perceived as poor (Q3.4), when not useless (Q3.3), which in turn affects the focused concentration on playing with the remote co-performer (Q4.1, Q4.2, and Q5.2).

Despite the distress, the musicians seemed to adjust to the experienced difficulties to a certain extent (Q5.1 and Q5.2) and make sense of the NMP interaction as a coherent whole (Q2.1). Again, we interpret this result rather as a goal-

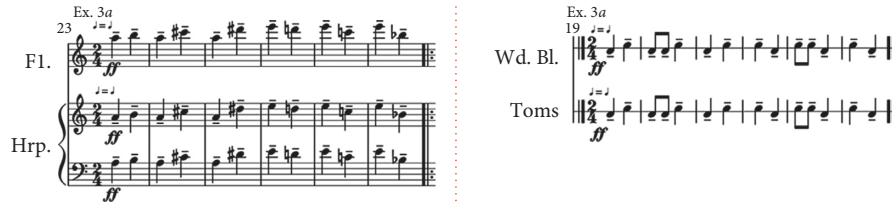


FIGURE 6: Example of homorhythmic, unison melody, with static expression. Left: flute and harp. Right: wood blocks and tom-toms.

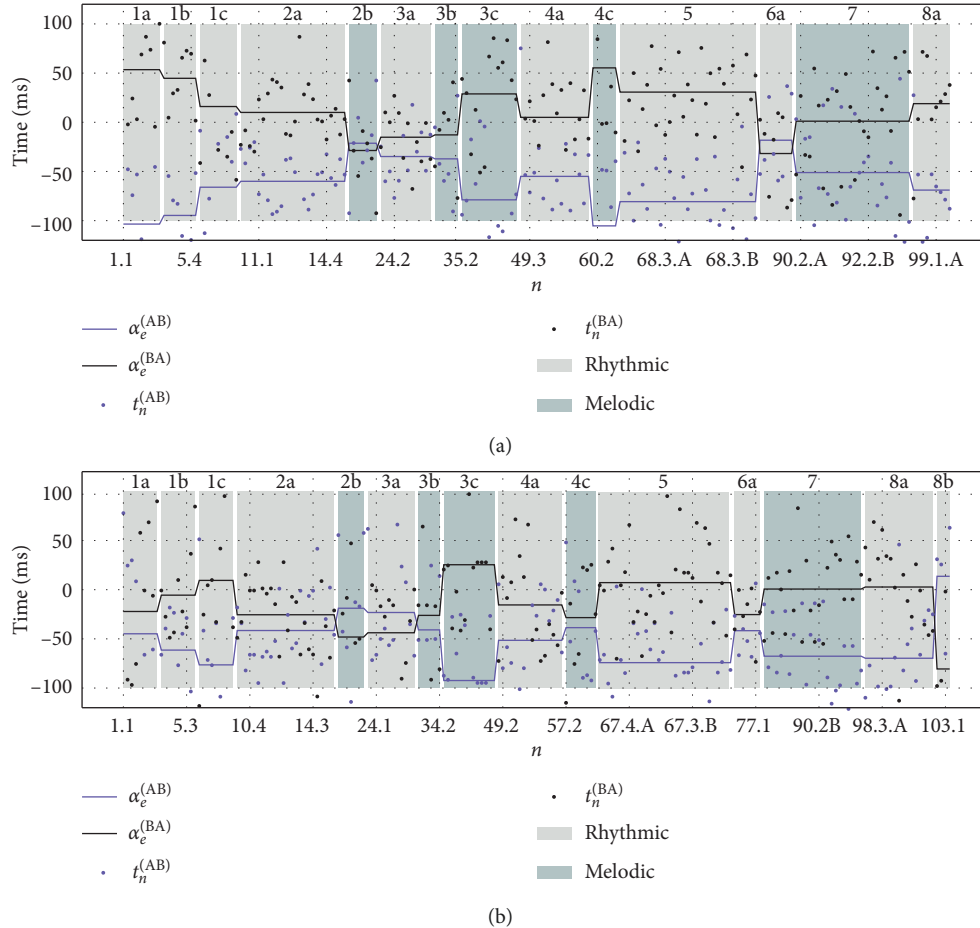


FIGURE 7: ISDs and asymmetry (averaged over each exercise) for couple (E). (a) Asymmetry for the couple E with 50 ms latency. (b) Asymmetry for the couple E with 67 ms latency.

TABLE 4: Latency values in the six conditions.

Condition	1	2	3	4	5	6
Delay (2-ways, ms)	28	33	50	67	80	134

directed quality of skilled musicians in managing to cope with the most adverse performing conditions. We can assume that such a situation is certainly not acceptable in a learning scenario. Nevertheless, the subjects proved their motivation and willingness to master the environment for the purpose of playing music remotely (Q5.2 and Q5.3).

In detail, Figure 8 shows the answers to the five postrepetition questions, regarding the focused concentration on the performance (Q1.2 and Q1.7, Figure 8(a) and Figure 8(b), respectively), the perceived coherence of the scenario (Q2.4, Figure 8(c)), the distraction factors (Q3.5,

Figure 8(d)), and the perceived quality of the performance (Q5.4, Figure 8(e)). The answer distribution highlights a negative effect of latency levels on the musicians involvement in the environment. However, these results must not be considered as conclusive; they rather highlight diverse inclinations and aptitudes of the subjects towards the delay issues. As general postexperiment comments indeed, participants C1 and D2 reported a lack of “musical connectedness,” despite the plausibility of the experience, which was also accounted by participants C2 and E1. It was reported that the decrease in involvement or flow, due to longer delays, increased the difficulty or impossibility to understand which was the cause of playing out of time. Of interest, this passage ended in an argument between the participants (couple E), whether the cause was ascribable to

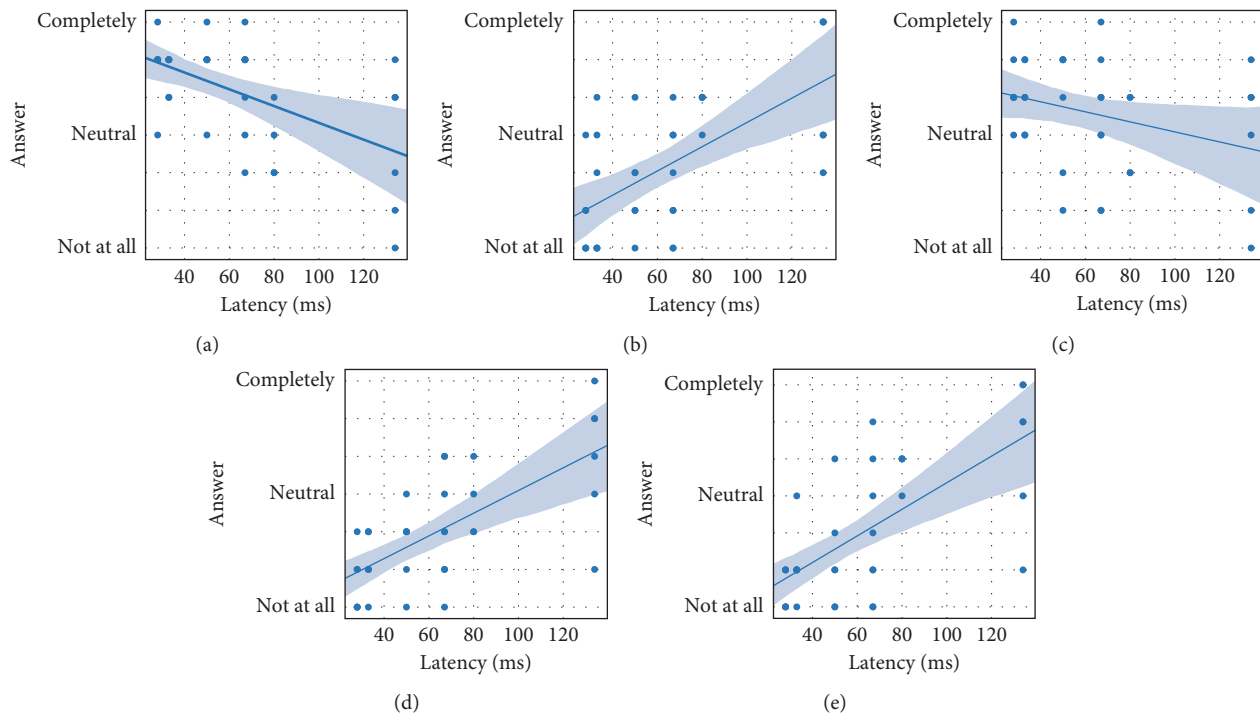


FIGURE 8: Answers to Q1.2, Q1.7, Q2.4, Q3.5, and Q5.4 in the postrepetition questionnaire with respect to the latency condition. Points represent individual answers to the questions. The solid straight lines represent the linear relationships, as obtained through linear regression, between the latency and the answers. The translucent band lines represent the 95% confidence interval for the regressions. (a) Answers to Q1.2. (b) Answers to Q1.7. (c) Answers to Q2.4. (d) Answers to Q3.5. (e) Answers to Q5.4.

the reduced commitment of the co-performer or the idiosyncrasy of the networked medium, thus reflecting a typical conflict and self-repair situation which can be compromised in telepresence systems [73]. As a final remark, it is interesting to look at the median values of the answers to the items on the interface awareness and quality (Section 3 of Table 2), which describe the perceived mastery of the interface, and are a subscale of the overall immersion of the system’s technology. The median values of the answers suggest that the musicians felt confident to concentrate on the musical task and make sense of the interface at hand. It is worth noting that the quality of the visual display does not seem to interfere or distract from performing (Q3.3), while the audio quality does (Q3.4). We hypothesize this is due to the poor quality of immersion of the visual representation (Q4.1), which makes it difficult for the performers to rely on vision to actively survey the performance (Q1.7). It is possible that this condition led the musicians to rather rely on the audio feedback in their performance. As a general comment, all the couples reported unanimously that the frontal screen resulted in a less natural interaction, as they normally use peripheral vision to monitor the co-performers on their side.

Taken together, the subjective evaluation of the NMP experience of the three duos returns a picture of a complex situation, wherein the issues at stake are multifaceted and systemic, especially with respect to the quality of the immersion and the sense of focused concentration awaited by musicians. From this viewpoint, a rather detailed reformulation of the

questions concerning the quality of the auditory and visual display would resolve the current, apparent ambiguity (e.g., the availability of certain information such as eye contact and breath attack, as a function of the system’s vividness and interactivity).

**4.2. Objective Evaluation.** In this section, we complement the subjective evaluation with the content analysis of the NMP performance. We show how measures of tempo trends can be used to interpret the performance strategies enacted by musicians to cope with NMP latency and in general to make sense of the medium behavior.

We compute the tempo trend  $\bar{\delta}(n)$  for all the recordings of the experiments. In Figure 9, we show two sets of annotations and corresponding tempo trends (scatter plot), smoothed trend (continuous lines), and linear approximation computed from  $\kappa$  (dashed lines) for musicians in Room 1 (blue) and Room 2 (orange).

Figure 9(a) shows the tempo trend of the two percussionists (couple C), playing with a latency of 134 ms. From the visual inspection, it can be observed a smoothed and highly correlated trend showing a high degree of synchronization and increase of tempo during the performance. This behavior suggests that with such high latency, the musicians were not able to follow each other, opting instead to a master-slave approach. The percussionists’ performance was particularly challenging and severely hampered by the presence of a high audio feedback due to the nature of the instruments.

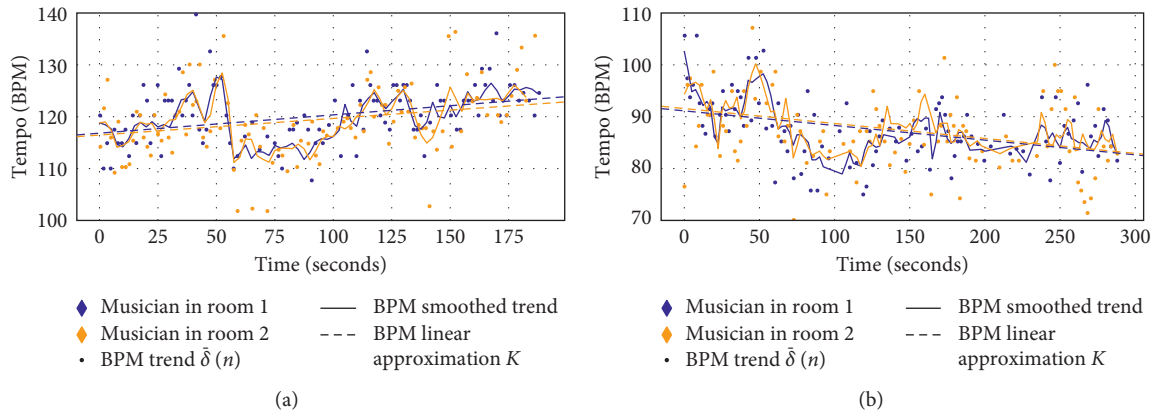


FIGURE 9: Tempo trend, in BPM, and its linear approximation for two sets of repetitions. (a) Tempo trend of the second repetition of couple C with 134 ms latency. (b) Tempo trend of the fifth repetition of couple D with 134 ms latency. (c) Legend of the figures.

A radically different situation is depicted in Figure 9(b), which shows the tempo trend of couple D performance (harp and flute), in the same latency condition of 134 ms; the two instrumentalists attempt to cope with the prohibitive latency condition by performing with a realistic interaction approach. The result, however, is a plain and progressive deceleration with respect to the reference tempo. In addition, in postexperiment debrief, the percussionist C2 commented that she mainly focused on keeping the internal tempo, while ignoring the co-performer's delayed performance. Both D1 and D2 confirmed that they were trying to follow each other's performance. The disparity of results can be accounted as well to the inherent difference between the instruments played by the two couples. Percussionists are effective and skilled followers, even at higher measures of delay [36], as it can be observed in the relative similar smooth fits of the tempo trend of both musicians, in Figure 9(a). D1, the harpist, and D2, the flautist, must confront greater challenges due to the constrictive relationship imposed by both the melodic and agogic constraints of the score. Unlike the percussionist, who has to focus mainly on the tactus, the harpist and the flautist have to preserve a synchronicity in pitch as well.

We also computed the ISDs  $t_n^{(AB)}$  and  $t_n^{(BA)}$  for all the repetitions and the asymmetry as the average ISDs over each of the 14 exercises in the score, which we will label as  $\alpha_c$ . Figure 7 shows the performance asymmetry of couple E (alto sax), where the  $x$ -axis is the common beats  $n$  and the  $y$ -axis shows the ISDs (in milliseconds) and the two asymmetries. The gray-scale bands represent the 14 exercises, from 1a to 8a, whereas the dark gray areas indicate those exercises rather centered on the melodic and expression synchronicity. The profile in purple represents the asymmetry of musician E1.

From the visual inspection of Figure 7(a), which refers to the latency condition of 50 ms, a few observations can be made. Musician E1 constantly anticipates the beats and musician E2 constantly follows him, except in the two regions of exercises 2b and 6a. In particular, at exercise 7, E1 anticipates E2 on average by the exact amount of latency, and hence from his side E2 is playing exactly on time, suggesting the overall negotiation of master/slave approach within the duo [74].

At 67 ms of latency, the duo seems to follow another strategy. As shown in Figure 7(b), especially at the beginning of the performance, from 1a to 3b, both musicians are anticipating, in the attempt to cope with the latency. Conversely, in exercises 5, 7, and 8a, they return to the master-slave approach, with the same roles. This passage well represents the idiosyncrasy experienced by the musicians, in the attempt to understand the medium behavior, and provides a clue of the relevance of the overall coherence of a given scenario in providing consistent and undeviating responses. In general, the time needed by the couple to estimate the latency and opt for the most appropriate interaction strategy is a clue of a disruptive effect which, in the current NMP environment, prevents the musicians from making a reliable judgment and ascribe the mistakes or the poor quality of their performance to their acts or to the medium.

## 5. Conclusions

The early results of the pilot experiment, described in this paper, offer a picture of the many entangled aspects that characterize the user experience in networked environments for music interaction. The low sample size of the duos involved clearly prevents us from stressing any conclusive statement. Conversely, the experimental environment draws attention to the complexity of the many experiential and technological variables that affect the effectiveness of a network music performance. In this respect, the conceptual framework situated in the chamber music practice and learning scenario acts as a magnifying glass for observing the constructive elements that create the plausible illusion of playing and learning music in geographically displaced environments.

Despite the clearly limited number of cases, the research activity carried out so far has important methodological implications. First of all, the experience of chamber music making is put in the foreground, with respect to the general issue of presence in the virtual environment and to the technology behind the system that enables it. Secondly, the use of objective quality metrics is very helpful for exploring



design issues through in-action and on-action reflections [75]. The rationale of a well-designed score is that of considering the significant mediating elements of a classical chamber music performance, and yet retaining control on the musical structures implied in the relationships of synchronicity, which represents the major drawback of networked systems.

We described how pairs of musicians manage to adjust their interaction and negotiate the performance, even in the poor sensory conditions of a basic NMP staging (i.e., monoaural capturing and rendering and screen-based frontal view representation of the remote co-performer). The metric of asymmetry is a measure of misalignment between the performed parts, thus providing the link with the beat and the score (i.e., the chamber music scenario). As it has been shown, it provides a promising descriptor of the interaction approaches at play within the performance. The shapes emerging in Figure 7 reveal interesting hypotheses which are worth investigating and, namely, concerning the different effects of rhythm and melody/expression articulatory score indications on the musicians' negotiation of the networked performance. If this is the case, the pedagogical implications for the distance learning scenario may require the design of specific exercises to train pitch and temporal acuity to cope with adverse NMP conditions.

For this purpose, we plan to revise the score in order to balance the types of exercises. We should also reflect on the presentation order of the exercises, in terms of abrupt changes and homogeneity of types, yet without compromising the overall musical meaningfulness of the stimulus in the next experimental sessions. In this respect, an additional and valuable source of information is represented by the "44 Duos for Violin" by Béla Bartók, a series of pieces composed for pedagogical purposes, specifically addressed to train motor responses to aural problems, rhythmic and structural features, interpretation, and music memory [76].

On the other hand, the current experimental procedure actually reflects the management of the rehearsal, which represents the intermediate type of performance, between the concert and the lesson. The duos were given a short time to rehearse before the experiment; therefore, they essentially played by reading the score at first sight. Occurring mistakes certainly have an effect on the computed asymmetry. If we want to investigate the concert type in the networked space, it is important that duos be given the score to work with well in advance. Objective quality metrics will represent a more valuable resource to quantify the performance. In the same fashion, the subjective evaluation of the coherence of the NMP scenario is expected to be more grounded in expectations from the real world. The systematic inquiry of the rehearsal and lesson types, instead, may require different approaches, and ethnographic observations, protocol analysis, and objective metrics should be carried out over a more extended period of time. Future works are planned, where experiments regarding latency will be carried out with an extensive number of participants. We also aim at investigating more systematically the occurrence of place illusion in the NMP

music learning scenario, by experimenting more immersive audio-visual feedback solutions, such as binaural rendering and full-body projections. The presence questionnaire is also undergoing a substantial revision. For example, items referring to the quality of immersion, visual and auditory, are being detailed, based on the comments collected: chamber musicians make use of several visual and auditory signals to communicate in ensemble, such as foot tapping and breath attack, which should be made available and apparent by the system. Finally, we are introducing biometric measurement techniques, to make the use of the questionnaire more reliable [25, 50].

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was conducted for the InterMUSIC project, which received the financial support of the Erasmus+ National Agency under the KA203 Strategic Partnership action under grant no. 2017-1-IT02-KA203-036770.

## References

- [1] J. Lazzaro and J. Wawrzynek, "A case for network musical performance," in *Proceedings of the 11th International Workshop on Network and Operating Systems Support for Digital Audio and Video, NOSSDAV '01*, pp. 157–166, ACM, New York, NY, USA, June 2001.
- [2] Á. Barbosa, "Displaced soundscapes: a survey of network systems for music and sonic art creation," *Leonardo Music Journal*, vol. 13, pp. 53–59, 2003.
- [3] L. Gabrielli and S. Squartini, *Wireless Networked Music Performance*, Springer, Berlin, Germany, 2016.
- [4] E. Lakiotakis, C. Liaskos, and X. Dimitropoulos, "Improving networked music performance systems using application-network collaboration," *Concurrency and Computation: Practice and Experience*, p. e4730, 2018.
- [5] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.
- [6] G. Hajdu, "Embodiment and disembodiment in networked music performance," in *Body, Sound and Space in Music and beyond; Multimodal Explorations*, C. Wöllner, Ed., pp. 257–278, Taylor & Francis, London, UK, 2017.
- [7] R. Mills, "Liminal worlds: presence and performer agency in tele-collaborative interaction," *Tele-Improvisation: Inter-cultural Interaction in the Online Global Music Jam Session*, pp. 145–166, Springer International Publishing, Cham, Switzerland, 2019.
- [8] F. Schroeder and P. Rebelo, "Sounding the network: the body as disturbant," *Leonardo Electronic Almanac*, vol. 16, no. 4-5, pp. 1–10, 2009.

- [9] J. Braasch, "The telematic music system: affordances for a new instrument to shape the music of tomorrow," *Contemporary Music Review*, vol. 28, no. 4-5, pp. 421-432, 2009.
- [10] F. Alpiste Penalba, T. Rojas-Rajs, P. Lorente, F. Iglesias, J. Fernández, and J. Monguet, "A telepresence learning environment for opera singing: distance lessons implementations over internet2," *Interactive Learning Environments*, vol. 21, no. 5, pp. 438-455, 2013.
- [11] M. Iorwerth, D. Moore, and D. Knox, "Challenges of using networked music performance in education," in *Proceedings of the UK 26th AES Conference*, vol. 8, Audio Education, August 2015.
- [12] K. Dye, "Student and instructor behaviors in online music lessons: an exploratory study," *International Journal of Music Education*, vol. 34, no. 2, pp. 161-170, 2016.
- [13] C. Johnson, "Teaching music online: changing pedagogical approach when moving to the online environment," *London Review of Education*, vol. 15, no. 3, pp. 439-456, 2017.
- [14] B. Kehrwald, "Democratic rationalisation on the network: social presence and human agency in networked learning," in *In Proceedings of the 7th International Conference on Networked Learning*, pp. 215-223, Aalborg, Denmark, May, 2010.
- [15] A. Margaryan, M. Bianco, and A. Littlejohn, "Instructional quality of massive open online courses (MOOCs)," *Computers & Education*, vol. 80, pp. 77-83, 2015.
- [16] C. Chafe and M. Gurevich, "Network time delay and ensemble accuracy: effects of latency, asymmetry," in *Proceedings of the 117th conference of the Audio Engineering Society Convention*, vol. 117, Audio Engineering Society, San Francisco, CA, USA, October 2004.
- [17] C. Chafe, J.-P. Cáceres, and M. Gurevich, "Effect of temporal separation on synchronization in rhythmic performance," *Perception*, vol. 39, no. 7, pp. 982-992, 2010.
- [18] S. Kolozali, M. Barthet, G. Fazekas, and M. Sandler, "Automatic ontology generation for musical instruments based on audio analysis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2207-2220, 2013.
- [19] C. Rottondi, M. Buccoli, M. Zaroni, D. Garao, G. Verticale, and A. Sarti, "Feature-based analysis of the effects of packet delay on networked musical interactions," *Journal of the Audio Engineering Society*, vol. 63, no. 11, pp. 864-875, 2015.
- [20] J. Forlizzi and K. Battarbee, "Understanding experience in interactive systems," in *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS'04*, pp. 261-268, ACM, New York, NY, USA, August 2004.
- [21] G. Novembre and P. E. Keller, "A conceptual review on action-perception coupling in the musicians' brain: what is it good for?," *Frontiers in Human Neuroscience*, vol. 8, p. 603, 2014.
- [22] P. Q. Pfordresher and C. Palmer, "Effects of hearing the past, present, or future during music performance," *Perception & Psychophysics*, vol. 68, no. 3, pp. 362-376, 2006.
- [23] R. I. Godøy and M. Leman, *Musical Gestures: Sound, Movement, and Meaning*, Routledge, New York, NY, USA, 2010.
- [24] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: factor analytic insights," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 3, pp. 266-281, 2001.
- [25] M. Meehan, S. Razzaque, M. C. Whitton, and F. P. Brooks, "Effect of latency on presence in stressful virtual environments," in *Proceedings of the IEEE Virtual Reality*, pp. 141-148, IEEE, Los Angeles, CA, USA, March 2003.
- [26] E. B. Nash, G. W. Edwards, J. A. Thompson, and W. Barfield, "A review of presence and performance in virtual environments," *International Journal of Human-Computer Interaction*, vol. 12, no. 1, pp. 1-41, 2000.
- [27] W. Woszczyk, J. Cooperstock, J. Roston, and W. Martens, "Shake, rattle, and roll: getting immersed in multisensory, interactive music via broadband networks," *Journal of the Audio Engineering Society*, vol. 53, no. 4, pp. 336-344, 2005.
- [28] R. Nordahl and N. C. Nilsson, *The Sound of BeingThere Presence and Interactive Audio in Immersive Virtual Reality*, Oxford University Press, Oxford, UK, 2014.
- [29] R. Skarbez, F. P. Brooks Jr., and M. C. Whitton, "A survey of presence and related concepts," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, p. 96, 2017.
- [30] T. W. Schubert, "A new conception of spatial presence: once again, with feeling," *Communication Theory*, vol. 19, no. 2, pp. 161-187, 2009.
- [31] M. Leman, P.-J. Maes, L. Nijs, and E. Van Dyck, "What is embodied music cognition?," in *Springer Handbook of Systematic Musicology*, R. Bader, Ed., pp. 747-760, Springer, Berlin, Heidelberg, 2018.
- [32] C. Sora, S. Jordà, and L. Codina, "Chasing real-time interaction in new media: towards a new theoretical approach and definition," *Digital Creativity*, vol. 28, no. 3, pp. 196-205, 2017.
- [33] S. Duffy, P. G. Healey et al., "Co-ordinating non-mutual realities: the asymmetric impact of delay on video-mediated music lessons," in *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, London, UK, July 2017.
- [34] C. Chafe, M. Gurevich, G. Leslie, and S. Tyan, "Effect of time delay on ensemble accuracy," in *Proceedings of the International Symposium on Musical Acoustics*, vol. 31, p. 46, Nara, Japan, March-April 2004.
- [35] G. Aschersleben, "Temporal control of movements in sensorimotor synchronization," *Brain and Cognition*, vol. 48, no. 1, pp. 66-79, 2002.
- [36] J. R. Cooperstock, "Interacting in shared reality," in *Proceedings of the 11th International Conference on Human-Computer Interaction*, Las Vegas, NV, USA, July 2005.
- [37] C. Knudsen, "Synchronous virtual spaces-transparent technology for producing a sense of presence at a distance," in *Proceedings of the 2001 Telecommunications for Education and Training Conference, TET*, Charles University Prague, Prague, Czech Republic, May 2001.
- [38] C. Hendrix and W. Barfield, "The sense of presence within auditory virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 5, no. 3, pp. 290-301, 1996.
- [39] M. Gurevich, D. Donohoe, and S. Bertet, "Ambisonic spatialization for networked music performance," in *Proceedings of the 17th International Conference on Auditory Display (ICAD2011)*, Budapest, Hungary, June 2011.
- [40] S. Farner, A. Solvang, A. Sæbo, and U. P. Svensson, "Ensemble hand-clapping experiments under the influence of delay and various acoustic environments," *Journal of the Audio Engineering Society*, vol. 57, no. 12, pp. 1028-1041, 2009.
- [41] J. Malloch and M. M. Wanderley, "Embodied cognition and digital musical instruments: design and performance," in *The Routledge Companion to Embodied Music Interaction*, M. Lesaffre, P.-J. Maes, and Leman, Eds., pp. 438-447, Routledge, Abingon, UK, 2017.
- [42] C. Chafe, "I am streaming in a room," *Frontiers in Digital Humanities*, vol. 5, no. 27, 2018.
- [43] Y. Iwaya and B. F. Katz, "Distributed signal processing architecture for real-time convolution of 3d audio rendering for

- mobile applications,” in *Proceedings of the International Conference on Virtual Reality and Augmented Reality*, pp. 148–157, Springer, London, UK, October 2018.
- [44] B. Boren and M. Musick, “Spatial organization in musical form,” in *Proceedings of the International Computer Music Conference*, pp. 33–38, ICMC, Daegu, South Korea, August 2018.
- [45] J.-P. Cáceres and C. Chafe, “Jacktrip: under the hood of an engine for network audio,” *Journal of New Music Research*, vol. 39, no. 3, pp. 183–187, 2010.
- [46] C. Drioli, C. Allocchio, and N. Buso, “Networked performances and natural interaction via LOLA: low latency high quality A/V streaming system,” *Information Technologies for Performing Arts, Media Access, and Entertainment*, pp. 240–250, Springer, Berlin, Germany, 2013.
- [47] P. Holub, J. Matela, M. Pulec, and M. Šrom, “Ultragrid: low-latency high-quality video transmissions on commodity hardware,” in *Proceedings of the 20th ACM International Conference on Multimedia*, pp. 1457–1460, ACM, New York, NY, USA, October–November 2012.
- [48] J. V. Draper, D. B. Kaber, and J. M. Usher, “Telepresence,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 40, no. 3, pp. 354–375, 1998.
- [49] J. J. Cummings and J. N. Bailenson, “How immersive is enough? A meta-analysis of the effect of immersive technology on user presence,” *Media Psychology*, vol. 19, no. 2, pp. 272–309, 2016.
- [50] M. Slater, “How colorful was your day? Why questionnaires cannot assess presence in virtual environments,” *Presence: Teleoperators and Virtual Environments*, vol. 13, no. 4, pp. 484–493, 2004.
- [51] M. Slater, “Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3549–3557, 2009.
- [52] M. Leman, *Embodied Music Cognition and Mediation Technology*, Mit Press, Cambridge, MA, USA, 2007.
- [53] J. W. Davidson and J. M. M. Good, “Social and musical coordination between members of a string quartet: an exploratory study,” *Psychology of Music*, vol. 30, no. 2, pp. 186–201, 2002.
- [54] D. Glowinski, G. Gnecco, S. Piana, and A. Camurri, “Expressive non-verbal interaction in string quartet,” in *Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 233–238, IEEE, Geneva, Switzerland, September 2013.
- [55] A. Chirico, S. Serino, P. Cipresso, A. Gaggioli, and G. Riva, “When music “flows”. state and trait in musical performance, composition and listening: a systematic review,” *Frontiers in Psychology*, vol. 6, p. 906, 2015.
- [56] B. G. Witmer and M. J. Singer, “Measuring presence in virtual environments: a presence questionnaire,” *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, pp. 225–240, 1998.
- [57] L. Comanducci, M. Buccoli, M. Zanoni et al., “Investigating networked music performances in pedagogical scenarios for the intermusic project,” in *Proceedings of the 23rd Conference of Open Innovations Association (FRUCT)*, pp. 119–127, IEEE, Bologna, Italy, November 2018.
- [58] L. Badino, A. D’Ausilio, D. Glowinski, A. Camurri, and L. Fadiga, “Sensorimotor communication in professional quartets,” *Neuropsychologia*, vol. 55, pp. 98–104, 2014.
- [59] C. Alexandraki and R. Bader, “Anticipatory networked communications for live musical interactions of acoustic instruments,” *Journal of New Music Research*, vol. 45, no. 1, pp. 68–85, 2016.
- [60] S. Vandemoortele, K. Feyaerts, G. De Bièvre, M. Reybrouck, G. Brône, and T. De Baets, “Gazing at the partner in musical trios: a mobile eye-tracking study,” *Journal of Eye Movement Research*, vol. 11, no. 2, p. 6, 2018.
- [61] M. Yoshie, K. Kudo, T. Murakoshi, and T. Ohtsuki, “Music performance anxiety in skilled pianists: effects of social-evaluative performance situation on subjective, autonomic, and electromyographic reactions,” *Experimental Brain Research*, vol. 199, no. 2, pp. 117–126, 2009.
- [62] S. Duffy and P. Healey, “A new medium for remote music tuition,” *Journal of Music, Technology and Education*, vol. 10, no. 1, pp. 5–29, 2017.
- [63] A. Williamon and J. W. Davidson, “Exploring co-performer communication,” *Musicae Scientiae*, vol. 6, no. 1, pp. 53–72, 2002.
- [64] E. Geelhoed, D. Prior, and M. Rofe, “Designing a system for online orchestra: microphone evaluation and cost-benefit analysis,” *Journal of Music, Technology and Education*, vol. 10, no. 2-3, pp. 213–230, 2017.
- [65] B. Boren and A. Genovese, “Acoustics of virtually coupled performance spaces,” in *Proceedings of the 24th International Conference on Auditory Display-ICAD2018*, pp. 80–86, Michigan Technological University, Houghton, MI, USA, June 2018.
- [66] D. Markovic, F. Antonacci, A. Sarti, and S. Tubaro, “Soundfield imaging in the ray space,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2493–2505, 2013.
- [67] L. Bianchi, F. Antonacci, A. Sarti, and S. Tubaro, “Model-based acoustic rendering based on plane wave decomposition,” *Applied Acoustics*, vol. 104, pp. 127–134, 2016.
- [68] S. Serafin, M. Geronazzo, C. Erkut, N. C. Nilsson, and R. Nordahl, “Sonic interactions in virtual reality: state of the art, current challenges, and future directions,” *IEEE Computer Graphics and Applications*, vol. 38, no. 2, pp. 31–43, 2018.
- [69] S. Delle Monache, M. Buccoli, L. Comanducci et al., “Time is not on my side: network latency, presence and performance in remote music interaction,” in *Proceedings of the XXII CIM Colloquium on Music Informatics-Machine Sounds, Sound Machines*, pp. 152–158, Udine, Italy, November 2018.
- [70] B. G. Witmer, C. J. Jerome, and M. J. Singer, “The factor structure of the presence questionnaire,” *Presence: Teleoperators and Virtual Environments*, vol. 14, no. 3, pp. 298–312, 2005.
- [71] G. Robillard, S. Bouchard, P. Renaud, and L. Cournoyer, “Validation canadienne-française de deux mesures importantes en réalité virtuelle: l’immersivité tendencieuse questionnaire et le presence questionnaire,” in *Proceedings of the 25e Congrès Annuel de la Société Québécoise pour la Recherche en Psychologie (SQRP)*, Trois-Rivières, Canada, November 2002.
- [72] B. Bartók, *Mikrokosmos, 153 Progressive Piano Pieces, New Definitive Edition*, Boosey & Hawkes Music Publishers Limited, London, UK, 1987.
- [73] S. Duffy and P. Healey, “Using music as a turn in conversation in a lesson,” in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 35, pp. 2231–2236, Queen Mary University of London, London, UK, August 2013.
- [74] E. Carôt and C. Werner, “Network music performance-problems, approaches and perspectives,” in *Proceedings of the Music in the Global Village Conference*, vol. 162, pp. 23–10, Budapest, Hungary, September 2007.

- [75] D. A. Schön, *The Reflective Practitioner: How Professionals Think in Action*, Basic Books, New York, NY, USA, 1983.
- [76] S. M. Genovefa, "The pedagogical significance of the Bartók duos," *American String Teacher*, vol. 12, no. 3, pp. 22–29, 1962.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

