TIME IS NOT ON MY SIDE: NETWORK LATENCY, PRESENCE AND PERFORMANCE IN REMOTE MUSIC INTERACTION

Stefano Delle Monache
Dept. of Architecture and Arts
Iuav University of Venice
sdellemonache@iuav.it

Michele Buccoli, Luca Comanducci, Augusto Sarti
DEIB, Politecnico of Milano University
name.surname@polimi.it

Giovanni Cospito, Enrico Pietrocola, Filippo Berbenni
Conservatory of Music G. Verdi of Milano
name.surname@consmilano.it

ABSTRACT

A pilot test on the sense of presence and the quality of the interaction in Networked Music Performance is presented. Subjective measures, based on a presence questionnaire, are combined with objective quality metrics, in order to stress the contribution of temporal factors (i.e., network latency) on the musical experience in the mediated environment. Preliminary results in the scope of chamber music practice are presented.

1. INTRODUCTION

Computer-aided musical collaborations between geographically-displaced musicians have been subject of extensive investigation from a variety of perspectives, since the late '90s. Early categorizations of computer systems for musical interaction have been proposed in [1], based on the temporal (synchronous vs. asynchronous) and spatial (colocated vs. remote) dimensions of the performance. The research and development has focused on the technical and perceptual issues (i.e., network delay and audio quality) affecting the on-line, simultaneous performance between musicians located and playing in remote rooms [2].

A substantial body of work has been produced in that area that has been crystallized under the acronym of NMP, namely Networked Music Performance. Gabrielli and colleague provide a valuable picture of the state of the art of NMP research and projects [3, Chapter 2, 3].

From a different angle, a renovated interest in remote collaborative environments has been growing in the area of audio-video (AV) streaming and conferencing systems for educational purposes [4]. NMP technologies and tools are increasingly being available on the market, and proposed as viable standards in blended and distance learning [5].

The EU funded project InterMUSIC [1] (Interactive Environment for Music Learning and Practicing, 2017 - 2020) aims to bridge the approaches established in NMP research with the opportunities of distance learning and education.

A project-based and practice-led approach to research is aimed at distilling, by the end of the project, an effective and operational environment, and especially a systematized knowledge in the form of best practices and guidelines for the implementation of remote environments for music interaction and education. This involves to author three online pilot courses in music theory and composition, chamber music practice, and vocal training, by means of the implementation of Massive Open Online Courses (MOOCs). With MOOCs, the interaction between teacher and students is usually one-to-many and asynchronous. On the other side, distance learning of the music practice can be delivered through one-to-one synchronous interaction between the teacher and the student.

One of the most popular tool for NMP is designed to provide low-latency interaction by means of dedicated connections and high-performing hardware [6]. In the common scenario, however, students do not have access to such networks and need to connect with the teacher by means of general purpose connections and hardware, which introduce processing and transmission latency.

With regard to NMP, the goal of the project is to investigate the user experience in order to optimize and improve the tools currently available. In this paper, we introduce a study, still in progress, aimed at understanding how temporal factors (i.e., network latency) affect the sense of presence, and the quality of the performance of chamber music duos involved in remote collaboration, i.e music making.

We ask 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. A qualitative assessment through questionnaires on the sense of presence and the perceived quality of the performance [7] is combined with quality metrics of the objective performance [8]. A follow-up study will be devoted instead to the investigation of spatial representations, auditory and visual.

The premise is that effective music-making and communication rely on the availability of auditory and visual cues (i.e., sonic gestures) [9], which are inevitably constrained in NMP, and in telepresence environments in general. We turn the traditional engineering approach to NMP research upside down, and seek for design strategies to

[1] http://intermusicproject.eu/. The Consortium is composed of the Conservatory of Music "G. Verdi" of Milano (Coordinator), the Polytechnic University of Milan, the RDAM Royal Danish Academy of Music of Copenhagen, the LMTA Lithuanian Academy of Music and Theatre of Vilnius, the AEC Association Européenne des Conservatoires, Académies de Musique et Musikhochschulen.

Figure 1: Instrumentalist positioning in room 1, with frontal view of the co-performer displaced in room 2.

compensate, and facilitate a plausible music experience in the mediated environment.

In this respect, we are investigating how simulated network delay, and diverse modes of audio-visual spatial representation, separately, affect the subjective experience of being present and together in the shared reality environment. 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 [10].

The paper is organized as follows: in Section 2 we provide the background context on top of which we are building our study; in Section 3 we elaborate on the concept of presence in the embodied cognition framework; Section 4 introduces the pilot experiment, and the methodology; we discuss the preliminary observations collected, in Section 5.

2. BACKGROUND

Both remote music performance and tuition face two major issues, that is dealing with temporal and spatial factors that are crucial for the effective and expressive performers and tutor-student communication.

Whereas remote music performance demands instantaneous and immediate responses, musical interaction in remote tuition rather reflects the flexibility of turn-taking in conversations [11]. In both situations, visuo-spatial forms of communication and behaviors are exploited to support mutual understanding and negotiation between peers.

Properly designed spatial elements in remote interactive environments may facilitate the compensation of time-dependent misalignments in the communication. In this respect, the design of the stage and the performance scene implies not only the physical displacement of various equipments in the rooms, but also the choice of appropriate solutions in both visual and sound feedback display [4, 11]. Various off-the-shelf strategies, such as the use of spatialized audio to increase the sense of presence [12], life-size and near-life-size visual display to preserve a coherent distance perception in virtual proximity [13], projections and video loop techniques to support synchronous interaction [14], have been proposed. These aspects are not considered in the current experimental research and will be subject of separate investigation.

In the following subsections we illustrate some relevant network latency effects on remote music performance, and discuss the available environments and tools, from the InterMUSIC perspective.

2.1 Temporal factors in remote interaction

Temporal factors refer to latency issues in data transfer, inherent to any NMP due to many factors, which can be broadly divided into processing time and network latency [3, Chapter 3]. The former includes the time for digital-to-analog and analog-to-digital conversion, for the acquisition and rendering of AV signals, their buffering and their packetization/depacketization for transmission. The latter is the latency introduced by the transmission of the packets, and it depends on the properties of the network connection and of the streams to be transmitted [2].

The temporal factors in remote interaction affect the quality of the music performance, and require musicians to cope with the delayed interaction. Music tolerance to network latencies have been assessed in a variety of conditions, and quantified in a wide range of delay thresholds [2]. In [8], Rottondi and colleagues conducted a set of experiments on NMP and performed a signal-based analysis of the recorded performances to draw some conclusions on the effect of the latency and of other musical factors on the quality of the performance. For instance, they showed that adverse network conditions ($20 - 60$ ms) affect the tempo of the performance, by inducing a pronounced tendency to deceleration, computed as explained in Section 4.4.

Carôt and Werner identify different strategies that musicians can follow to cope with the latency, depending on the style and the tempo of the music performance at hand [15]. The best scenario, mostly feasible for low latencies, is the Realistic Interactive Approach, when the performers flawlessly interact together. With latencies higher than 25ms, musicians tend to follow a Master-Slave or a Laid-Back Approach. In the former, one musician leads the performance and others follow her lead; in the latter, one musician plays slightly behind the rhythm to compensate for the delay. These approaches are also common in co-presence performances, when a performer plays a lead role or behind the rhythm.

Jointly with latency, other factors have been observed to affect the quality of a NMP. Barbosa and Cordeiro observe that slow instruments' attack times lead to higher deceleration, but also higher synchronization between musicians [16], and Rottondi and colleagues observe a correlation between deceleration and pieces' rhythmic complexity and instruments' noisiness [8]. Rhythmic and timbral factors will be investigated in future experiments.

2.2 Environments for remote interaction

In the scope of the InterMUSIC project, we intend to use LOLA [2], developed by Conservatorio di Musica Giuseppe Tartini of Trieste, in collaboration with GARR Consortium [6], and UltraGrid [3], developed by CESNET's Laboratory of Advanced Networking Technologies (Sitola) [17].

LOLA requires to use dedicated hardware and the European National Research and Education Network to provide low-latency, large bandwidth connection. LOLA runs as a Windows program with a GUI, and offers several codecs for AV signals, AV recording functionality, and a set of tools for remote assistance. Due to its features, LOLA has been widely used for artistic remote performances [6]. However, LOLA is designed to work in best conditions, and it is not open-source for the community.

UltraGrid is a cross-platform program which also implements several AV codecs. Since it is based on command line, it is more difficult to use for non-tech-savy users. The code of UltraGrid is public [4], hence it is possible to implement additional functionalities, e.g., the support for spatial factors. UltraGrid can reach an end-to-end latency as low as 50 ms, which may be too high for some scenarios.

In this pilot test, we use LOLA because of its low-latency. We intend to use UltraGrid in future experiments for spatial or other factors.

3. PRESENCE: KEEPING THE GOOD COMPANY

Network constraints and high quality sound reproduction have been considered the main factors affecting the plausibility of music interaction in networked performance. Research strictly focused then on the sensory and control dimensions involved in the experience of playing together remotely [18].

In InterMUSIC, we approach the problem of providing an effective communication and interaction environment for networked music practice and teaching, by looking at presence studies [19, 10] in the embodied cognition framework [20, 7]. The basic assumption is that the experience of music emerges in interaction, as complex network of predictive models of observable patterns and intentional states, that are acquired through knowledge and skills [21].

The concept of presence is subject of ongoing investigation as potential construct from which deriving appropriate measures of the effectiveness of virtual environments. Several theoretical and operational models have been proposed over the last two decades (see [19] for a comprehensive survey), also in the attempt to linking such a construct to task performance. A conclusive demonstration has not been provided yet [10].

In the scope of InterMUSIC research, we adopt the definitions of presence proposed in [19, 7], that is presence as *the cognitive feeling of being in a place*, intended as *the perceived realness of the mediated experience*. In this study, we refer to the presence model by Schubert and colleague [7], emerging from the factor analysis of 246 answers to a 75-item survey of questions taken from several questionnaires. Due to space constraints, we do not elaborate further on this definition, instead we discuss the basic components of the presence experience, as operational constructs that have been assessed experimentally in established questionnaires [19].

The sense of presence is tied to action in the (mediated) environment, and results from the interpretation of the mental model derived. Given the focus of the chamber music practice task, the suppression of conflicting sensory stimuli and the focused allocation of attentional resources is of primary importance.

3.1 Components of the presence experience

Presence encompasses three major categories of constructs: i) the subjective experience of being there (spatial presence, involvement, realness); ii) evaluations of the immersive technology (interface awareness and quality); iii) evaluations of the interaction (predictability/anticipation) [7].

Spatial presence refers to the emerging relation between the mediated environment as a space and the instrumentalist's own body. **Involvement** or flow retains the attention side of the presence experience, that is the relative concentration and focus on the real and the mediated environments, in terms of keeping the compatible information from the real environment and suppressing the incompatible one. **Realness** encompasses reality judgments with respect to the meaningfulness and coherence of the experience with the expectations from the real-world.

Interface awareness and quality take in account distraction factors, and in general provides a clue of the mastery of the interface in the specific activity at hand. As additional construct, **quality of immersion** considers the vividness, coherence and fidelity of sensory factors (i.e., auditory and visual cues, and valid actions) as supported by the technology.

Predictability refers to the possibility to anticipate what will happen next, in terms of activation of motor representations as a consequence of perceiving while playing.

The peculiarity of NMP, and in particular in the context of chamber music practice in InterMUSIC, makes the remote interaction environment very constrained, since the music making task is very specific and demanding, in terms of active survey. Put in design terms, this represents an advantage, potentially providing clearer user requirements. Much of expected outcomes are aimed at collecting and organizing the instrumentalists expectations.

4. THE EXPERIMENTAL STUDY

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 infrastructures. How does the experience of network

[2] LOLA (LOw LAtency audio video streaming system) `https://lola.conts.it`

[3] UltraGrid `https://lola.conts.it/`

[4] UltraGrid code `https://github.com/CESNET/UltraGrid`

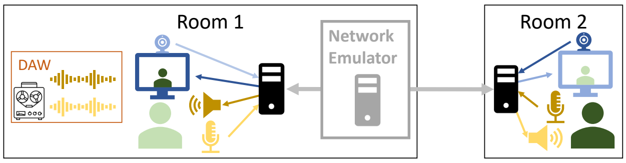


Figure 2: Setup of the pilot experiment, where two musicians perform together through a Network Emulator.

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

Table 1: Instrumentation and location of the instrumentalists in room 1 or room 2.

delay and interruption affect the involvement and realness, the predictability, the interface awareness and quality, and the overall quality of the musical performance? The same question applies separately to AV spatial representations in remote interaction and will be investigated in a forthcoming study. The outcomes are expected to provide relevant design implications in the staging of classrooms dedicated to remote music practice and teaching.

In Section 4.1 we describe the setup of the experiment, and in Section 4.2 the participants and the stimuli for the experiment. Performances' quality and sense of presence are assessed through subjective questionnaire, reported in Section 4.3, and objective metrics, explained in Section 4.4. The scores of the stimuli and the presence questionnaire are publicly available [5] . A further research goal in the long term is to provide evidence of links between presence and performance in remote music interaction.

4.1 Apparatus

Figure 2 represents the experimental setup. Two musicians perform in two rooms with no direct sound interferences.

Musicians interact by means of contact microphones (monaural acquisition), loudspeakers (monaural rendering), cameras and screens. The spatial arrangement of visual displays and loudspeakers reflects the current practices established in remote music practice and tuition, that is a frontal view of the co-performer in order to improve eye contact [6, 22]. Figure 1 shows the staging in Room 1.

The hardware equipment is connected to two Windows computers, which are two high-end Intel/Nvidia powered workstations with i7 esa/octa core processors, using PCIe audio cards, according to both LOLA and UltraGrid hardware and software requirements. The computers communicate together through a Gigabit ethernet connection by means of a server with two Ethernet interfaces acting as a Network Emulator to add latency [6] . The server is placed in Room 1 to be easily accessible during the tests and to ease troubleshooting in case of network issues.

Audio output and input of the performer in Room 1 is redirected to a Digital Audio Workstation to record the performance (from the perspective of Room 1). These recordings are used to compute content-based metrics of quality,

[5] `https://tinyurl.com/intermusicCIM2018`
[6] Using Linux command `tc`

Rhythm	Melody	Expression (dynamics, articulation, agogic)	Ex.
homorhythm	octave or unison	static	3a
homorhythm	opposite direction	static	8a
heterorhythm	opposite direction	alternation	5
heterorhythm	homodirection	climax	7
phasing	octave or unison	static	6b
slicing	slicing	dynamic	4b
imitation	imitation	imitation	2b
ostinato	pedal tones	static	3b

Table 2: Types of musical structures in rhythm-melody-expression relationship.

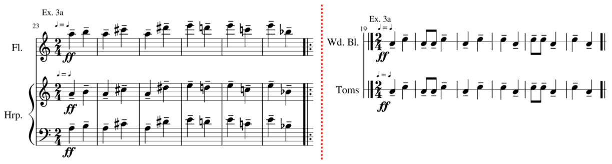


Figure 3: Example of homorhytmic, unison melody, with static expression. Left: Flute and harp. Right: Wood blocks and tom-toms.

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

Table 3: Latency values in the six conditions.

as described in Section 4.4.

4.2 Participants and task

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 are all musicians with at least five years of academic musical practice. Each duo had already a certain familiarity of minimum two weeks of rehearsal. Table 1 shows the instrumentation and respective location per each duo in rooms 1 and 2 (see Figure 2).

The stimuli were designed in order to encompass diverse basic structures of musical interaction with respect to time management and communication strategies. 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.

In this respect, we looked at Béla Bartók's *Mikrokosmos* piano pieces [23], which represent a valuable methodological compendium of exercises in meaningful *rhythm-melody-expression* relationships. In Table 2, we pinpoint eight types of structures per musical dimension. These can be combined in diverse expressive relationships of musical synchronicity. A few examples are reported and referenced in the fourth column to guide the reader in the understanding of the score.

Figure 3 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 musical stimuli have a duration of 3 minutes, and a reference tempo of 112 BPM.

Each duo had to perform the exercise, under six different conditions of emulated network delay, reported in Table 3. As described in Section 2.1, a certain latency is unavoidable due to time for audio processing. By setting the network latency to 0 ms, we estimated the two-way processing time as 28 ms, which is the first condition in Table 3. We use two sets of latency values: a deterministic set, and a set corresponding to rhythmic figures, given the reference tempo. Specifically, 33, 67 and 134 ms correspond to one 64th, 32th or 16th respectively. The idea behind this choice is to investigate whether having a musically-meaningful latency will help performers to cope with it. However, due to the few examples, statistical tests in this regard have not been conducted yet.

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, as that of performing the exercise for 6 times, in physically displaced locations, and through the telepresence environment. No information about the nature of the six network delay conditions was disclosed. 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 session. These are extensively reported in Section 4.3. 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.

4.3 Presence questionnaire

The questionnaire used in the study was constructed by merging three reference questionnaires on presence [24, 7, 25], and selecting the most appropriate items with respect to the music performance task. 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 subdivided into five *groups*; and a post-repetition questionnaire with five questions extracted from the general one. We report the post-experiment questionnaire in Table 4 (with the median, mean and standard deviation of the answers).

The questionnaire is constructed around three main concepts: i) Presence (G1); ii) Immersion (G2, G3, G4); iii) Quality of the Performance (G5). G1 on presence is further composed of items on three main constructs, that is *spatial presence* (Q1.2-4), *involvement* (Q1.3, Q1.6-9), *realness* (Q1.1, Q1.5). G2 encompasses items on *predictability and interaction* (Q2.x), G3 on *interface awareness and quality* (Q3.x). G4 focuses on the *quality of immersion* (Q4.x), and

Question	**Median**	**Mean**	**Std**
1) Involvement and realness			
1.1) How much did your musical experience in the remote environment seem consistent with your real world experiences?	4.0	4.40	1.36
1.2) In the remote environment I had a sense of "being there".	3.5	4.17	1.77
1.3) The sense of playing in the remote environment was compelling.	5.0	4.50	1.50
1.4) I had a sense of playing in the remote environment, rather than performing something from outside.	4.0	3.80	1.17
1.5) How realistic did the remote environment seem to you?	4.5	4.33	0.75
1.6) How aware were you of the real world surroundings around you during the performance?	4.5	4.00	1.91
1.7) How completely were you able to actively survey the musical environment using vision?	4.0	3.83	1.57
1.8) How completely were you able to actively survey the musical environment using audition?	5.0	4.67	0.94
1.9) The delay affected the sense of involvement.	4.5	4.00	1.63
2) Predictability and Interaction			
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) I was able to anticipate the musical outcome in response to my performance in the remote environment.	5.0	5.00	1.00
2.4) The environment was responsive to actions performed by my partner.	4.5	4.25	1.48
2.5) 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
2.6) It was difficult to cope with the distance performance.	4.0	3.83	1.57
3) Interface awareness and quality			
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
4) Quality of the immersion			
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
5) Quality of the music performance			
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

Table 4: Statistics of the answers to the post-experiment questionnaire defined in Sec. 4.3. We indicate in italics the post-repetition questions.

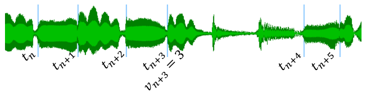
G5 is dedicated to the subjective assessment of the *quality of the performance* (Q5.x).

The post-repetition questionnaire contains five items (indicated in italics in Table 4) referring to the three concepts, with a special emphasis on the experience of delay.

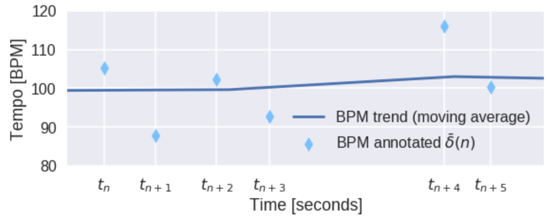
4.4 Objective quality metrics

Beside the subjective questionnaire, we conducted a set of content-based analysis on the recordings of the study. The

(a) The stimulus.



(b) The signal annotated with t_n and v_n.



(c) The computed BPM trend $\bar{\delta}(n)$.

Figure 4: Representation of the annotation procedure for computing objective content-based metrics.

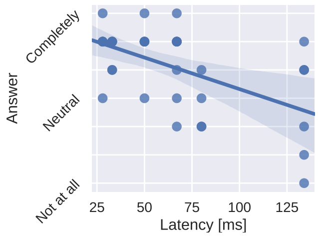
goal of this analysis is to obtain an objective metrics of the quality of the performance. In the literature, researchers have proposed different metrics, related to the rhythmic trend of the performances [2]. In Fig. 4 we show a representation of the annotation procedure that leads to the computation of these metrics. Given the stimulus represented by a score (Fig. 4a), the first step is to annotate, in the recordings, the instants when an onset occurs on the beat, as $t_1, t_2, ..., t_N$ (Fig. 4b). 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 Fig. 4b, 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 $(t_1, v_1), ..., (t_N, v_N)$ 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 4c. In order to smooth the noise in the annotated BPM due to execution and annotation imprecision, we smooth $\bar{\delta}(n)$ using a moving average filter, with a fixed width of 5 seconds, as shown in Fig. 4c.

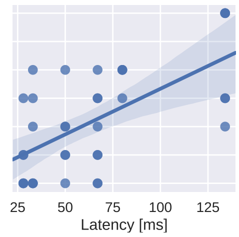
From on $\bar{\delta}(n)$ we can compute several metrics [2] related to the symmetry between two performers, or the imprecision/regularity of the execution, or the slope of the linear approximation of $\bar{\delta}(n)$. The latter provides a compact descriptor of the trend of the tempo, since $\kappa = 0$ when the tempo remains steady for the whole performance and it assumes positive or negative values in case of acceleration or deceleration, respectively. While in the InterMUSIC project we intend to use all the aforementioned metrics, in the rest of this paper we will focus on the latter, due to its interpretability and graphical representation.

5. RESULTS AND DISCUSSION

Given the limited number of collected sessions, we provide the reader with a narrative of the information that we are able to extract from both subjective and objective evaluation. We consider only three sessions out of five, since two sessions were either not completed or deeply biased.



(a) Answers to Q1.3



(b) Answers to Q5.4

Figure 5: Answers to Q1.3 and Q5.4 in the post-repetition questionnaire with respect to the latency condition.

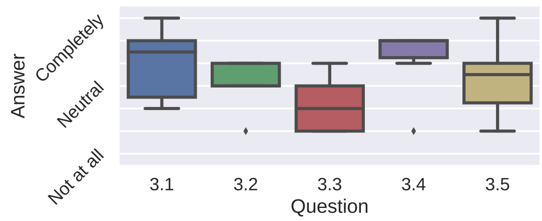


Figure 6: Distribution of the answers to the questions related to G3 *Interface awareness and quality* (see Table 4).

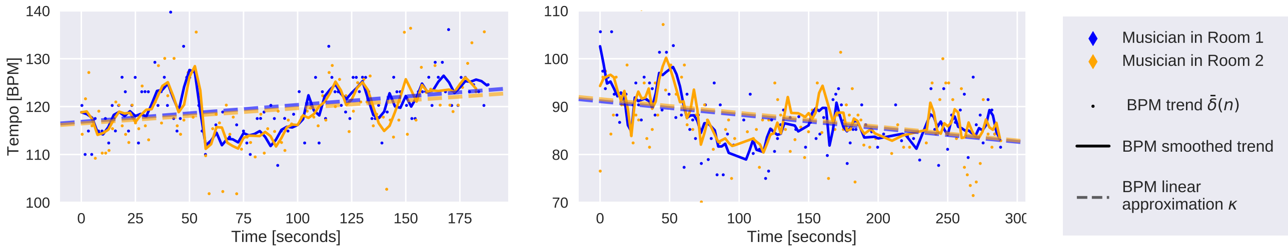
5.1 Subjective evaluation

The subjective evaluation is aimed at measuring the general sense of presence in NMP performance, and at observing whether and how delay affects the presence components, that is the attention allocation to the real and mediated environment, the coherence and fidelity of the overall experience, in addition to the perceived quality of the performance.

Answers to the post-repetition questionnaire return predictable, yet interesting results. For instance, in Fig. 5 we show the answers to two post-repetition questions on involvement and the subjective experience of delay: Q1.3 *The sense of playing in the remote environment was compelling* (Fig. 5a) and Q5.4 *The delay affected the sense of involvement* (Fig. 5b). The answers' distributions highlight a negative effect of latency levels to musicians' involvement in the environment.

From the answers to the post-repetition questionnaire in Table 4, it emerges a general distress caused by the latency and a willingness to adapt and find ways to cope with it. In post-experiment comments, 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. Of interest, the decrease in involvement or flow, due to longer delays, resulted in the difficulty or impossibility to understand who or which was responsible of playing out of time.

Focusing on G3 *interface awareness and quality*, we show the distribution of answers in Fig. 6. The median values of the answers suggest that musicians feel confident to concentrate on the musical task, adapt and master the interface at hand. It is worth noting that the visual display quality seems not to interfere or distract from the performing (Q3.3), while the audio quality does (Q3.4). We believe this is due to the poor quality of immersion of the



(a) Signal-based metrics of the second repetition of couple C, with 134 ms latency.

(b) Signal-based metrics of the fifth repetition of couple D, with 134 ms latency.

(c) Legend of the figures

Figure 7: BPM trend and its linear approximation for two sets of repetitions.

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, since they normally use peripheral vision to oversee the co-performers on their side.

5.2 Objective evaluation

We compute the tempo trend $\bar{\delta}(n)$ for all the recordings of the experiments. In Figure 7 we show two sets of annotations and corresponding tempo trends (scatter plot), smoothed trend (continuous lines) and linear approximation computed from κ (dashed lines) for musicians in room 1 (blue) and 2 (orange).

Figure 7a shows the results regarding the two percussionists (couple C), playing with a latency of 134 ms. The smoothed trend shows that there is a high degree of synchronization between the performers, and the trends are highly correlated, making them to increase the tempo during the performance. This is probably due to the fact that with such a latency the musicians were not able to follow each other and instead opted for a master-slave approach. The percussionists performance was particularly challenging because of the nature of their instruments that caused a high audio feedback, which further distracted them from the performance.

Figure 7b shows the results for the harp and flute players (couple D) playing with a latency of 134 ms . While the latency value is the same as in Figure 7a, the results are dramatically different. By listening to the recordings, we note that the two instrumentalists attempt to perform with a realistic interaction approach, but the prohibitive latency condition led them to strongly slow down over time with respect to the reference tempo.

Our analysis of the compensation strategies followed by the two couples are further supported by the comments that they provided after the end of the session. While the percussionist C2 stated that she focused on the internal tempo, ignoring the partner's delayed performance, both D1 and D2 stated that they were trying to follow each other's performance.

The discrepancies in the two performances are also due to the inherent difference in the musicians' instruments. Percussionists are able to efficiently follow themselves even at higher levels of latency [13], which is shown by the relative similar smooth fits of the tempo trend of both musicians, as shown in Fig. 7a. D1, the harpist, and D2, the flautist, must confront bigger challenges due to the constrictive relationship imposed both by the melodic and agogic constraints of their stimuli. Unlike the percussionist, who have to focus mainly on the tactus, the harpist and the flautist have to preserve a synchronicity in pitch as well.

6. CONCLUSIONS

NMP practices are subject of renovated interested in the wider scope of music teaching. Within the activities of the InterMUSIC project, we are currently tackling the most demanding aspects of music interaction in remote environments. We seek to gain a wider understanding of NMP issues in the context of chamber music, and provide design solutions that cope with temporal and spatial factors affecting the remote musical experience.

In this work, we presented the design of a pilot experiment, which included the technical setup, the creation of stimuli and presence questionnaire and the introduction of objective metrics of quality for performance. While in the first experiment we focused on temporal factors, in future work we plan to collect a valubale number of sessions and start exploring diverse design strategies in spatial representations. This will allow us to address issues raised by the participants of the experiments, such as the use of peripheral vision or the need of spatialized audio. In addition, we intend to take a wider variety of objective metrics into consideration [2], as well as to introduce physiological measures of presence [19].

7. ACKNOWLEDGMENT

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 number: 2017-1-IT02-KA203-036770.

The authors would like to thank Daniele Moro, Sebastian Troia and Marco Fumagalli for their help in the setup of the Network Emulator. We would also like to thank the participants of the pilot test for their effort and their availability to be contacted for further experiments.

8. REFERENCES

[1] Á. Barbosa, “Displaced soundscapes: A survey of network systems for music and sonic art creation,” *Leonardo Music Journal*, pp. 53–59, 2003.

[2] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, “An overview on networked music performance technologies,” *IEEE Access*, vol. 4, pp. 8823–8843, 2016.

[3] L. Gabrielli and S. Squartini, *Wireless Networked Music Performance*. Springer, 2016.

[4] 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.

[5] M. Iorwerth, D. Moore, and D. Knox, “Challenges of using networked music performance in education,” in *AES Conference:UK 26th Conference: Audio Education (August 2015)*, 8 2015.

[6] C. Drioli, C. Allocchio, and N. Buso, “Networked performances and natural interaction via lola: Low latency high quality a/v streaming system,” in *Information Technologies for Performing Arts, Media Access, and Entertainment*, pp. 240–250, Springer, 2013.

[7] T. Schubert, F. Friedmann, and H. Regenbrecht, “The experience of presence: Factor analytic insights,” *Presence: Teleoperators & Virtual Environments*, vol. 10, no. 3, pp. 266–281, 2001.

[8] C. Rottondi, M. Buccoli, M. Zanoni, D. Garao, G. Verticale, and A. Sarti, “Feature-based analysis of the effects of packet delay on networked musical interactions,” *J. Audio Eng. Soc*, vol. 63, no. 11, pp. 864–875, 2015.

[9] R. I. Godøy and M. Leman, *Musical gestures: Sound, movement, and meaning*. Routledge, 2010.

[10] 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.

[11] S. Duffy, P. G. Healey, *et al.*, “Co-ordinating non-mutual realities: The asymmetric impact of delay on video-mediated music lessons.,” in *Proc. of the 39th Annual Conference of the Cognitive Science Society*, 2017.

[12] C. Hendrix and W. Barfield, “The sense of presence within auditory virtual environments,” *Presence: Teleoperators & Virtual Environments*, vol. 5, no. 3, pp. 290–301, 1996.

[13] J. R. Cooperstock, “Interacting in shared reality,” in *Proc. HCI International, 11th Int. Conf. on Human-Computer Interaction*, 2005.

[14] C. Knudsen, “Synchronous virtual spaces–transparent technology for producing a sense of presence at a distance,” in *Proc. of The 2001 Telecommunications for Education and Training conference, TET*, 2001.

[15] E. Carôt and C. Werner, “Network music performance - problems, approaches and perspectives,” in *in International School of new Media, Institute of Telematics, University of Lbeck. Music in the Global Village - Conference*, 2007.

[16] A. Barbosa and J. Cordeiro, “The influence of perceptual attack times in networked music performance,” in *Audio Engineering Society Conference: 44th International Conference: Audio Networking*, Nov 2011.

[17] P. Holub, L. Matyska, M. Liška, L. Hejtmánek, J. Denemark, T. Rebok, A. Hutanu, R. Paruchuri, J. Radil, and E. Hladká, “High-definition multimedia for multiparty low-latency interactive communication,” *Future Generation Computer Systems*, vol. 22, no. 8, pp. 856–861, 2006.

[18] W. Woszczyk, J. Cooperstock, J. Roston, and W. Martens, “Shake, rattle, and roll: Gettiing immersed in multisensory, interactiive music via broadband networks,” *Journal of the Audio Engineering Society*, vol. 53, no. 4, pp. 336–344, 2005.

[19] 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.

[20] T. W. Schubert, “A new conception of spatial presence: Once again, with feeling,” *Communication Theory*, vol. 19, no. 2, pp. 161–187, 2009.

[21] 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, 2018.

[22] S. Duffy and P. Healey, “A new medium for remote music tuition,” *Journal of Music, Technology & Education*, vol. 10, no. 1, pp. 5–29, 2017.

[23] B. Bartók, *Mikrokosmos, 153 Progressive Piano Pieces, New Definitive Edition*. London: Boosey & Hawkes Music Publishers Limited, 1987.

[24] B. G. Witmer and M. J. Singer, “Measuring presence in virtual environments: A presence questionnaire,” *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[25] G. Robillard, S. Bouchard, P. Renaud, and L. Cournoyer, “Validation canadienne-française de deux mesures importantes en réalité virtuelle: l’immersive tendencies questionnaire et le presence questionnaire,” *25e Congrès Annuel de la Société Québécoise pour la Recherche en Psychologie (SQRP), Trois-Rivières*, 2002.