Investigating Networked Music Performances in Pedagogical Scenarios for the InterMUSIC Project

Luca Comanducci, Michele Buccoli,
Massimiliano Zanoni, Augusto Sarti
Politecnico di Milano
Milan, Italy
{name.surname}@polimi.it

Stefano Delle Monache
Iuav University of Venice
sdellemonache@iuav.it

Giovanni Cospito, Enrico Pietrocola,
Filippo Berbenni
Conservatory of Music G. Verdi of Milan
Milan, Italy
{name.surname}@consmilano.it

***Abstract*—With the big improvement of digital communication networks, Networked Music Performances (NMP) received a great interest from music live performance and music recording industry. The positive impact of NMP in pedagogical applications, instead, has been only preliminary explored. Within the InterMUSIC project, we aim to investigate NMP from a pedagogical perspective, that has considerable differences with respect to music performances, and to develop tools to improve distance learning experiences. In this paper, we introduce a conceptual framework designed to be the foundation for all the experiments conducted in the project. We also present two preliminary experiments that investigate the sense of presence of geographically-distant musicians in a distance learning scenario. We discuss the comments provided by the musicians as a set of requirements and guidelines for future experiments.**

I. Introduction

The rapid evolution of technology and the consequent increasing speed of digital communication networks allows to improve the communication experiences with the dramatic reduction of the virtual distances. Among the others, the process, facilitates the emergence of new scenarios of interaction among geographically-displaced people. Since the late ’90s, a community composed of musicians, technicians and scientists has been investigating how technology can enable music performances leading to the definition of Networked Music Performance (NMP), which *occurs when a group of musicians, located at different physical locations, interacts over a network to perform as they would if located in the same room* [1].

NMP has immediate applications in the performative scenario enabling musicians to rehearse from remote distances and to perform in geographically-distributed concerts [16]. As part of the broader category of audio-video streaming and conferencing systems, NMP has been having interesting applications in educational area for blended and distance learning [2].

The EU funded project InterMUSIC(Interactive Environment for Music Learning and Practicing, 2017 - 2020) aims to develop and improve tools for distance learning of music and to collect such tools in integrated remote environments for music interaction and education. Within the project, we have chosen three online pilot courses: music theory and composition, chamber music practice, and vocal training. Tools for courses are developed following two main paradigms.

In the first paradigm, students from any part of the world access the course and find the teaching material provided by the teacher, with the possibility to interact with their colleagues and professors. This paradigm, namely Massive Open Online Courses, is designed for one-to-many asynchronous communication and it is widely used also by elite universities [4].

In the second paradigm, students use NMP softwares for attending master-student lessons or rehearse together. This paradigm has stricter requirements in terms of synchronicity and therefore only allows one-to-one (or few-to-few) synchronous communication. In this paper, we focus on the NMP-based paradigm and we discuss the requirements related to the specific pedagogical scenario.

The literature about NMPs has investigated numerous factors which may affect several aspects of the performances. Affecting factors include the unavoidable network latencies [22], timbral properties of the employed instruments [5] and rhythmic properties of the performed pieces [19]; while affected aspects range from objective quality of the performance [20], to perceptual metrics of musicians’ comfortability [30]. This investigation has led a highly popular tool for NMP, which allows low-latency interaction by using high-end hardware connected through an ad-hoc inter-university network [27].

In order to widen the audience of NMP, specific tools need to be available also for general purpose connections and hardware, hence addressing higher processing and transmission latency. Nevertheless, with the idea of using NMPs in the educational context, the goal is not the NMPs’ objective or subjective quality, but rather those aspects and factors that can guide students to improve their technique or enable a comfortable remote rehearsal.

The aspects we need to address for the success of pedagogical NMPs are not known a priori, leading us to make some assumptions. For example, we may assume that the acoustic proprieties of the environment affect or alienate the sense of presence of the performers, or we may require that the networked environment should not worsen students’ level of stress. In order to validate our assumptions, we need to conduct perceptual experiments with musicians to test different conditions. As an example, we test how musicians perform in environments with different acoustic proprieties, and we measure musicians’ stress level using sensors for biometric signal recording [34]. With the purpose to set up the confrontation with music teachers and students, during the experiments, we also collect comments. Useful comments are transformed into technical requirements for the project.

With this goal in mind, in this paper we identify the main entities involved in the performances and the interaction among them and we formalize them in a framework. The role of the framework is to provide an abstraction for the experiments that we need to conduct. The framework is designed to be generic for any kind of networked or physical performance, and we plan to extend it in order to provide a comprehensive semantic description of rehearsal and teaching activities for pedagogical scenarios. We present the architecture and details of the framework in Section III.

Using this framework, we are able to design the perceptual experiments for conducting investigations on NMPs in the pedagogical scenarios. In Section IV we describe two pilot experiments we conducted to investigate the role of visual cues in musical interaction, and the influence of latency in the sense of presence of musicians, respectively. The preliminary results of the experiments and the next steps of our investigation are presented in Section V.

II. Background and Related Work

In the literature, a wide investigation on NMPs has been devoted to understand the influence of different latency conditions on the performance. We discuss the main results in Section II-A. In Section II-B we briefly introduce the investigation regarding the network topologies and architectures, while in Section II-C we provide an overview of the studies on rooms' acoustic factors. Lastly, in Section II-D we list the main softwares employed for NMPs and their main features.

A. Temporal factors

Temporal factors refer to every aspect of the infrastructure that cause the presence of end-to-end delay between the musicians present in different locations. We can broadly divide the causes of end-to-end delay into two main factors, i.e., the signal processing delay and the network delay [18].

The former comprises the delays caused by the whole signal chain, consisting of acquisition hardware (e.g. soundcards), the encoding/decoding and fragmentation processes. The latter comprises all the possible delays caused by the network transmission between transmitter and receiver due for example to network congestion.

The level of latency is a factor that dramatically affects the NMP experience and as such was extensively analyzed in the literature. In [19] the authors present a series of content-based experiments that show how high latency values in the range of $20 - 60$ ms cause a decrease in the quality of the performance expressed by a tendency of the musicians to tempo deceleration.

In [22], [21], [20] the authors take into account the concept of *Temporal Separation*, which represents the time needed for the action of one person to reach another one in a setting where both are acting together. A set of experiments are devised, which evaluate the ability of several couples subjects in performing a clapping rhythm together while being in two separated rooms and while varying the transmission latency in the range of 3-78 ms. The results of such experiments prove that the best performance results are obtained when the transmission latency between the subjects is comparable to a Temporal Separation value corresponding to a setting where the two subjects are in the same room. The authors also observed that unnaturally low latency values cause an acceleration in the musicians' tempo.

B. Network factors

The choices about all the aspects of the network architecture depend on the aspects needed in the NMP framework. Network architecture used for NMP comprises both decentralized peer-to-peer(P2P) and client-server [17].

In P2P architectures each participant has to send its audio/video data to all the other musicians. This poses great issue for what concerns the scalability of the NMP system, causing a trade-off between the number of people connected and the quality of the audio-video content, which must be degraded in order to gain enough bandwidth to connect all the musicians.

Client-server architectures address the scalability issue, since they are based on a server that receives the individual streams sent by each musicians, mixes it and then sends it back. However, the bandwidth requirements remain rather high, and the server causes an additional delay, since it adds a two-way communication with each participant.

C. Acoustic and Spatial factors

The analysis of the impact of acoustic and spatial factors in NMP and distance learning has not been as deep and thorough as the ones performed on the analysis of the tempo effects.

In [25] the authors performs a set of experiments in semi-anechoic chambers and add different levels of artificial reverb in order to simulate a natural environment. The experiments show that the performances of the musicians give no noticeable differences that could be directly explained by the change in reverberation.

In [26], the author present an experiment with two musicians performing handclapping (similarly to [22]) considering three acoustic conditions: real reverberant, virtual anechoic and virtual reverberant. The results show that anechoic conditions cause more synchronicity issues than the reverberant ones.

In [24] the authors present a study that concerns the application of Ambisonics techniques to NMP problems. Athough they do not develop a fully functional framework for NMP, they prove the feasibility of implementing 3D spatial audio for NMP tasks.

D. Available softwares

In this Section we list the main tools that have been developed for NMP; for an extensive list we refer the reader to [17].

JackTrip [29] was developed by the SoundWIRE research group at CCRMA in order to support bi-directional music performances. 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[27] 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). LOLA is based on low-latency audio/video acquisition hardware and on the optimization of all the steps needed to transmit audio/video contents through a dedicated network connection. Because of its low-latency properties, we used it in the preliminary InterMUSIC experiments presented in [30]. As a drawback, the project is not open source and it is not optimized for generic network connections.

On the other side, UltraGrid [28] is an open-source software that allows audio/video low latency transmission. While its performance are still far from those achieved by LOLA, it is more flexible for generic hardware and networks and it allows contributors to implement new functionalities. For this reason, we are considering the use of UltraGrid in the next stages of the InterMUSIC project.

III. Framework

In this Section, we introduce a framework for the design of perceptual experiments on musical performances. The framework is composed of five main entities, which interact with each other in numerous ways, depending on the kind of experiment we want to conduct and performance we aim to observe. The framework can be considered as a first formalization of the problem of investigating NMP, and we aim to develop it further during the project.

We show in Figure 1 a schematic representation of the framework and the basic relations among entities. A **performance** occurs when two or more **subjects** musically interact together through a **medium**. Subjects can be musicians during a rehearsal, as well as teachers and students. In order to consider a large number of probable scenarios, a performance can occur with all the subjects in the same room (*local performance*), with all the subjects geographically distant (*networked performance*) or with part of the subjects in the same place and part of the subjects geographically distant (*mixed performance*). Subjects interact by means of a *medium*. In the case of local performances, the medium is a *physical medium*, such as simple air propagation. In the case of networked performances, the medium is a *networked medium*, such as an Internet connection and the NMP software/hardware equipment used in order to connect the two subjects. In the case of mixed performance both physical medium and networked medium are involved.

In all the scenarios subjects perform in an **environment** with specific timbral (acoustics of the room) and spatial properties (location of the subjects 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 he/she is playing as the *real environment* and the environment he/she perceives relative to the geographically-distant subjects as the *virtual environment*. For example, in Fig. 3, we show a set of frames from one of the experiments we conducted. Fig. 3a shows a harp player in her *real* environment; her partner's perception of this environment, i.e., the *virtual environment*, is shown in Fig. 3b.

In order to analyze the performance, it is crucial to run a **data recording** stage, using different devices to capture the multimodal signals. The factors and aspects that will be possible to analyze from the performance depend on the properties of the devices, e.g., whether they are *video* or *audio devices*, or where they are placed. In the following sections we describe in detail the aforementioned entities.

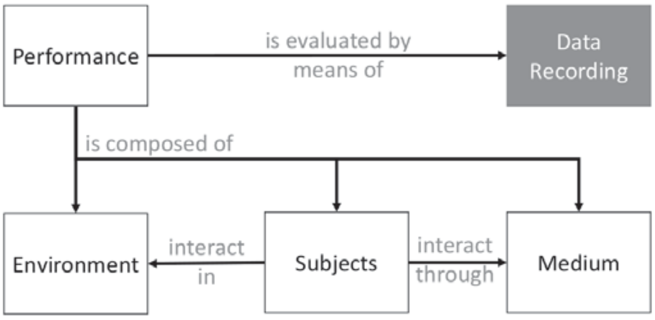


Fig. 1. A graphical representation of the proposed framework with main entities and interaction

A. Performance

In our experiments we consider two types of performances: *performed piece* or *taught lesson*. The performance is the entity at the highest hierarchical level and is composed of the subjects, the environments and the mediums. Main properties of the performance are date and time, location(s), type of performance, metadata (e.g., composer of the piece, tempo, meter, key signature, score, duration, etc.) and composition in form of symbolic representation (MIDI or musicXML) [3].

Some properties of the performance depend on the nature of its sub-entities. For example, if the subjects are two musicians, the performance can be a *rehearsal* or a *concert*, while if one of the subject is a teacher, the performance is defined as a *lesson*. As discussed above, the property of a performance also depends on the type of medium: we have a networked or mixed performance if we have at least one networked medium.

B. Subjects

Subjects involved in the performance may belong to a variety of possible roles, such the one of musician, student, teacher or conductor. They are identified by name, age, experience and musical background, however it is important to take into account that these data evolve over time and so must be referred to the day of the performance. Other properties inherently related to a subject are the assigned *score* and the used instrument. Scores can be also described by means of their musicological properties (e.g., dynamics or rhythmic complexity), while instruments can be described by means of a content-based timbral analysis (e.g. attack time) [19]. It is possible to analyze such properties as aspects that may affect the final quality or success of the performance.

C. Environment

We refer to the spatial and acoustic properties of both the physical place where a subject performs and the perception of it by the other subjects as physical and virtual environment, respectively. The properties of the virtual environment depend on the location and specifics of the audio/video capture and rendering devices, such as the microphone and speakers connected to the NMP equipment. For example, using an

TABLE I. DETAILS OF THE EXPERIMENT IN CO-PRESENCE

Entity	Properties
Performance	Co-presence performance. Parts arranged from Bartok pieces as described in [cit].
Subjects	Two brothers, violin and cello players. Violin player: male, 25 years, 17 years musical experience; Cello player: male, 19 years, 10 years musical experience
Environment	Recording and mastering studio in the Conservatory of Milan; acoustically equipped with bass traps. Musicians sit side by side, with peripherical vision.
Medium	Condition 1: Air propagation. Condition 2: Visual occlusion applied by means of a a screen with progressive layers of canvas applied to decrease transparency and visibility.
Data recording	Interview to the participants and free comments.

array of microphones we can capture the acoustic scene [31] and render it using arrays of speakers through spatial audio techniques [32].

Properties of the environment are also the possible processing applied to the audio or video signals. For example, by applying some kind of reverberation to the incoming audio signal, a subject will have a different perception of the virtual environment.

We also collect the information about the interaction of the subjects with the environment, such as the position of the musicians in the room, the details of the audio/video acquisition (e.g., microphone on the instrument vs. fixed position microphone) and the relative position of musician and devices.

D. Medium

The medium refers to the connection between the environments and, hence, the subjects. In the case of a networked performance, the medium collects information regarding the employed software for NMP and its settings, the network architecture and specifics, like bandwidth and latency. In the case of a local performance, like a traditional lesson, we collect information such as the distance between the subjects, which measures the acoustic latency between them, and describe possible visual / acoustic occlusions that may be placed between the musicians.

E. Data recording

In order to observe the experiment and draw meaningful conclusions, we need to record the properties of the performance that are interesting for our analysis. For the data recording stage, we consider multimodal signals and their processing byproduct as well as a questionnaire filled by the subjects. In the former case, we can extract objective metrics of the performance, while in the latter we consider subjective results; both are important to assess the outcome of the experiment.

With regard to the multimodal signals, the audio recording of the performance, from the two environments, are clearly useful to assess the quality of the performance or possible modifications in the timbral or rhythmic properties. Beyond that, video recordings are also useful to annotate saccadic movements during the interaction between the subjects [33], and we aim to capture biometric signals to objectively estimate the subjective distress of the performers [34].

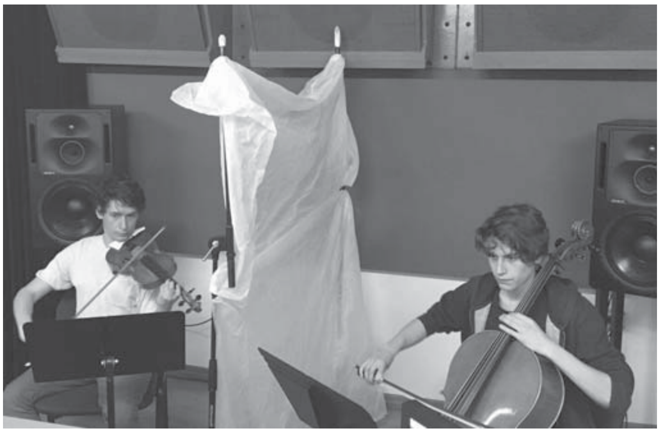


Fig. 2. The two musicians in experiment 1, with partial occlusion and blurred effect

IV. PRELIMINARY EXPERIMENTS

In this Section, we describe two preliminary experiments we conducted in Spring 2018 following the conceptual framework presented in the previous section. The first experiment was a rehearsal between two musicians in the same room, where we inserted visual occlusions to test their ability to interact in adverse conditions. The second experiment comprised a set of rehearsals between five couples of musicians in two separate rooms, connected via a network emulator to test the sense of presence and the quality of performance at different latency conditions. The results of the two experiments are discussed in the next section.

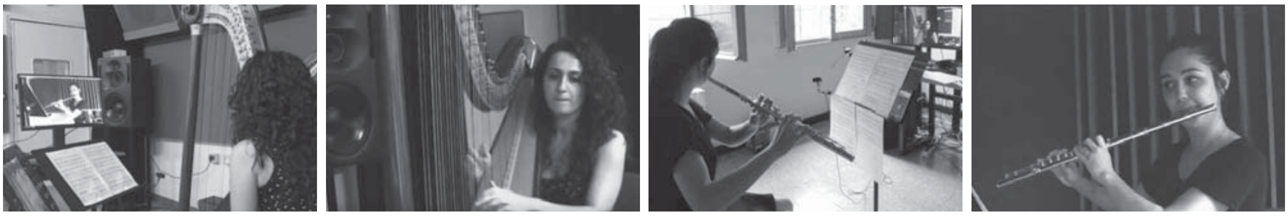
A. Co-presence performance with visual occlusion

We invited a string duo to make a pilot test of the environment, of the perceptual questionnaire and of the musical pieces that we intended to use for the second experiment. We also included a test on adverse conditions of the medium (visual occlusion) to qualitatively assess the importance of video feedback in NMPs.

We show a summary of the test conditions in Table I. The two subjects were a violin and a cello players. They were brothers and had a long-time experience in playing together. Due to these facts they were equipped with a wide set of well-established visual and audio cues developed in order to interact with each other. For the musical parts, we created a composition that highly requires mutual interaction and understanding, using simultaneous attacks, alternate scales, changes of tempo, unison playing, sustained loop, etc. The parts are described in detail in [30] and publicly available.

The performers were asked to play in two conditions: with no visual occlusion, and with partial visual occlusion, by inserting a tulle panel between the two instrumentalists, thus providing a blurring effect on their figures (as shown in Figure 2). The aim was to observe their behavior, adaptation strategies, and emerging non-verbal communication (bodily and musical gestures).

The data recording of the experiment was based on a subjective questionnaire on their sense of presence and free comments on the experience.



(a) A harp player performing in her real environment

(b) A harp player as visually perceived in the virtual environment

(c) A flute player performing in her real environment

(d) A flute player as visually perceived in the virtual environment

Fig. 3. View of real environments and virtual environments in the networked experiment

B. *Sense of presence in NMPs*

Five couples of musicians, playing different combinations of instruments, took part in the experiment. The couples had familiarity in playing together since at least two weeks. The details of the experiment are reported in Tab. II, for a more in-depth description, we refer the reader to [30].

Each session of the experiment comprised a single couple, the two musicians were placed in different rooms, standing in front of a screen, in order to be connected both with respect to video and audio. An example of the setup for the couple consisting of harp and flute players is shown in Fig. 3. In Fig. 3a and Fig. 3c it is possible to see the viewpoints of the musicians consisting of the scores and the screens connected to the other room.

The couples played the same parts considered in the previous experiments. The medium involved two computers equipped with LOLA connected through a network emulator, by means of which we could change the transmission latency. A single session consisted of the couple playing the same part in six repetitions, where each time we simulated a different latency level in a range between $28\mathrm{ms}$ and $134\mathrm{ms}$, as detailed in [30]. It is important to note that the latency levels weren't presented sequentially to the musicians (e.g. in a decreasing or increasing order), but were instead selected with no particular criterion for each repetition.

We asked the musicians to fill two different subjective questionnaires. The first one was presented after each repetition and consisted of five questions, selected in order to analyze their perception of the performance with respect to the latency just experienced. The second, reported in [30], was presented to the musicians at the end of the whole session and consisted of 27 questions, investigating the various aspects of the experience such as their sense of presence and the perceived general quality of the performance.

V. Results and discussion

Though limited in the amount of involved people and performances, the experiments described in the previous section were helpful to start the discussion about the pedagogical applications of NMPs with musicians, and to collect useful comments and suggestions that will guide our investigation.

A. *Objective, subjective and biological metrics*

In the two experiments, we investigated the sense of presence and quality of performance of the couple of subjects in case of visual occlusion (co-presence experiment) and different network latency conditions (networked performance). The acquisition and evaluation was performed using subjective and objective metrics [30].

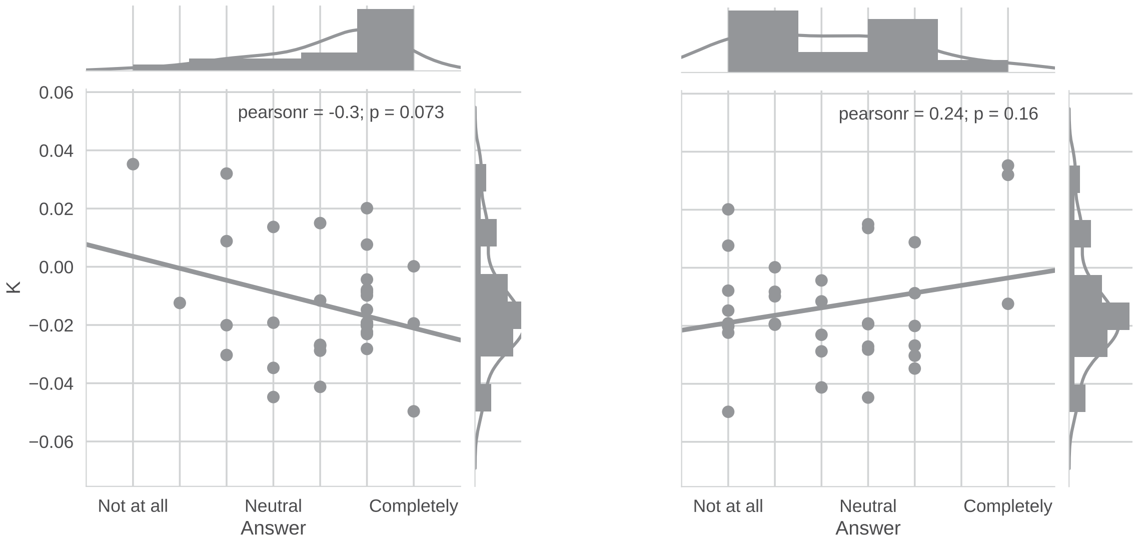
With regard to the former, we used a post-experiment 27-item questionnaire divided in five main topics, such as *Predictability and Interaction*, or *Quality of the Music Performance*. After each phase of the networked experiments, we asked a subset of five questions to evaluate the impact of different latency conditions in the questionnaire.

With regard to the latter, we acquired the audio recordings of the networked performance, manually tracked the beat, extracted a BPM trend and computed a degree of acceleration/deceleration K (*tempo slope*) from it. The assumption behind the choice of K is that the more two musicians decelerate during the performance (i.e. high negative K), the lower the perceived quality of the performance. Other metrics used in NMP literature include the pacing, the regularity or the imprecision of the performance [17].

It is interesting to observe the comparison between the subjectively-perceived quality of the performance and the objective metrics computed from the corresponding audio recording. In Fig. 4 we show the comparison with the answers to two quality-related questions, i.e., *the sense of playing in the remote environment was compelling* (Fig. 4a) and *the delay affected the sense of involvement* (Fig. 4b).

TABLE II. Details of the nmp experiment

Entity	Properties
Performance	Five NMP performances. Parts arranged from Bartok pieces.
Subjects	Five couples of musicians with different combinations of instruments. Average age: 21.9 years. Musical experience: at least 5 years
Environment	Two rooms: two mastering studios in the Conservatory of Milan; acoustically equipped with bass traps. Musicians sit in front of a screen and a webcam and monophonic audio input/output.
Medium	Software LOLA with video and audio transmission. Network emulator with different latency condition and fixed jitter.
Data recording	Audio recording of the performance from one room, audio/video recording using LOLA; perceptual questionnaire.



(a) Correlation between subjective answers to the question *the sense of playing in the remote environment was compelling* in the x-axis and objective metrics values

(b) Correlation between subjective answers to the question *the delay affected the sense of involvement* in the x-axis and objective metrics values

Fig. 4. Correlation between subjective answers in the x-axis and tempo slope K in the y-axis

Due to the few samples we obtained, it is not possible to draw statistically-meaningful conclusions. The preliminary observations, however, show an interesting trend that we intend to investigate further. The musicians seem to be more compelled with the remote environment when the levels of K are lower, i.e., when the tendency to slow down is more accentuated. Analogously, their sense of involvement in the performance show little positive correlation with the tempo trend.

Two musicians, indeed, can easily keep the tempo by making one subject following the lead of the other, using a master-slave approach to cope with latency [23]. In this case musicians would not improve their musical skills to interact with partners. From these results, we can infer that the assumption that low K leads to a lower perceived quality is not proven, and therefore K is not suitable to be used as an objective metrics of the subjective satisfaction of subjects.

In the context of the project, we plan a future investigation that will be devoted to find a content-based metrics that is coherent with the pedagogical purposes of NMPs and suitable for providing a useful feedback for the students.This study will involve acquiring biometric signals to estimate the level of distress during the performance and to investigate whether the NMPs contribute to increase such level.

B. Auditory and visual feedback

After the experiments, we asked musicians about the strategies they adopt for musical coordination and interpretation, which they report to be based on breathing signaling and communicative gestures to keep synchronization, especially for attacks and the duration of sustained notes. In the co-presence experiment, the no-sight condition deeply affected the expressiveness of the performance. In full visual occlusion, the performers relied mostly on acoustic cues to keep the tempo, with the apparent effect of an acceleration during the performance.

This aspect is also investigated in the remote performance by asking in the perceptual questionnaire how much the visual and auditory display quality interfered or distracted them from performing. In Fig. 5 we display a histogram of the answers in a 7-point likert scale. While the influence of the visual feedback is limited (Fig 5a), the auditory feedback is predominant for the performance.

During the time of free comments, the subjects explained that the video feedback was less relevant due to the low degree of synchronicity introduced by the latency, which led them to look less for a visual interaction with their partner. In order to translate these comments into requirements for a NMP tool, we need to assess the importance of visual and auditory interaction in performances and rehearsal. A level of auditory interaction can be estimated using a measure of asymmetry between the audio recordings of the two subjects, as computed in [20]. With regard to visual interaction, we intend to acquire a video recording of the performance using cameras that capture when the subjects are watching at their partners on the screen. By annotating both intentional and saccadic movements, we aim at estimating whether a higher level of interaction between subjects corresponds to higher satisfaction in the performance and, possibly, how to design the visual and auditory feedback in our platform.

C. Peripheral visual feedback

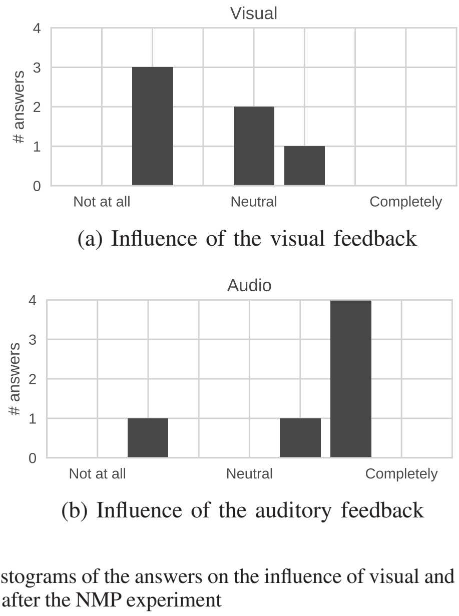
In the previous subsection, we discussed the importance of visual feedback in co-presence and networked performances.

During the co-presence experiment (Section IV-A), we observed how performers where comfortable with blurred visual of their partners. They reported us that most of the information for synchronization is contained in the perception of motion from the partners, rather than in the full view.

In the networked experiment the subjects were placed in front of a webcam and a monitor in order to improve eye contact, as in [11]. Some participants of the networked performances commented that one of the reason of the influence of the visual feedback was the lower importance of direct visual interaction. In their typical disposition during rehearsal and performance they are placed in front of an audience rather than in front of each other and, therefore, rely mostly on peripheral vision for interacting.

As next steps, we intend to test the role of peripheral vision by trying different arrangement of video equipment and subjects during the NMPs. This involves to test different virtual environments and find the most promising for pedagogical purposes. The most suitable arrangements may involve use multiple visual feedbacks, e.g., one frontal for eye contact and one for catching peripheral motion. This would mean dealing with multiple video streams, which demands a larger bandwidth. Reducing the demand in terms of bandwidth is nonetheless desirable in order to improve the spread of our tools. We intend to investigate on strategies to decrease the need of bandwidth.

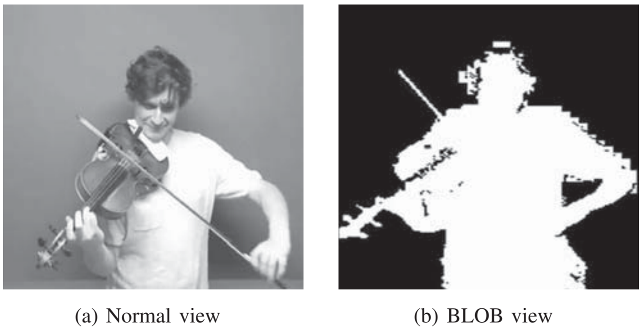
Both LOLA and Ultragrid implement several coding algorithm to reduce the bandwidth, while slightly increasing the processing time for coding/decoding the video streams [27], [28]. A higher saving would be to detect and transmits only the silhouette of the musicians as a binary large object (BLOB) [6]. In Fig. 6 we show a comparison between a normal take of a musician (Fig. 6a) and a BLOB view (Fig. 6b). Being a binary image, the BLOB view is extremely lighter, while keeping most of the information on the motion. As future directions, we intend to investigate whether the BLOB or other motion-related representations can effectively convey the information needed by musicians for rehearsing together.



(a) Influence of the visual feedback

(b) Influence of the auditory feedback

Fig. 5. Histograms of the answers on the influence of visual and auditory feedback after the NMP experiment



(a) Normal view (b) BLOB view

Fig. 6. Comparison between normal and BLOB view of a musician during a NMP

D. Mono vs Stereo vs 3D Audio

Monophonic audio acquisition and rendering is commonly used in NMPs [30]. Some studies employ headphones to avoid audio loops [19], while others use monophonic speakers and echo-cancellation algorithms to avoid feedbacks [27].

The directionality of the sound is crucial in case of performances with many musicians, where it helps locating the sources and improving the ability to focus on the instruments separately.

We intend to investigate the influence of sound directionality in NMPs by testing different audio conditions and verifying which one the musicians prefer. A first condition is to use panning with headphones and stereo speakers to place the sound sources. A second condition is to use binaural rendering for headphones and more accurate methods for the stereophonic location of the sounds. Lastly, we want to use an array of speakers in different arrangement to allow more accurate rendering of sound fields.

It is also worth mentioning that it is not natural for musicians to be wearing headphones during a performance, as they affect their ability to hear their own sound. Moreover, to properly locate the sound sources in the virtual environment, the binaural rendering requires a proper head tracking algorithm to constantly update the location of sources with respect to the orientation of the listener [15].

There are further scenarios where a 3D acquisition of the sound field is required. Most instruments have indeed a clear pattern of radiance and listening to them from different angles leads to sensible changes of the timbral properties [13]. In a teacher-student scenario, a teacher may be interested to have a richer information to assess the progress of their students. We aim to employ state-of-the-art techniques for the plenacoustic analysis and rendering, as in [7].

E. Matching acoustics of environments

In the networked experiment we did not take the influence of room acoustics for the performance into consideration. In other experiments, the performers play in acoustically insulated and semi-anechoic rooms to remove this influence [19]. Musicians have difficulty playing in semi-anechoic rooms, as the perceived timbre of instruments change dramatically making hard for them to recognize the quality of their performance [35].

In real-case scenarios, it is likely that musicians are going to perform over network from two acoustically-different environments. In this case, they would listen to their own sound colored by their environment and receive their partner audio signal from a different acoustic, creating a misalignment between the two environments.

In future work, we want to take the effect of environment acoustic into account by testing different combinations of real rooms' acoustics and synthetically-processed acoustics [36]. This will help us understanding if musicians require techniques to address the issue of different acoustic environments for having a realistic performance. These techniques may involve to first blindly assess the acoustic conditions of the two environments and then de-convolve the partner's audio stream [14].

F. Measure for latency compensation and virtual conductor

One of the main factor investigated in the NMP literature was the influence of network latency in the quality [17], [27], as we also did in the networked experiment. The new 5G network is showing promising results for enabling low-latency connections, that may enable NMPs applications even for generic users [12]. Nevertheless, a certain amount of latency is likely to be present for NMPs using general purpose hardware and connections.

For this reason, researchers have been developing strategies to cope with the latency. In [23] the authors identify a set of strategies such as the laid-back approach, which involves for a musician to play slightly behind the beat, and the delayed feedback approach, which involves to equip one of the musician with a delayed feedback of their own sound in order to synchronize it with their partners sound. The presence of a conductor has also been shown to increase the tolerance to delay, thanks to the shared cue provided to the performers [8].

In the project, we aim to develop algorithms that help musicians compensating for the delay. We plan to use a beat tracker to dynamically track the rhythm of the performance of the musicians, as in [9], and provide it in advance to musicians. This will allow musicians to follow a virtual conductor, similarly to what proposed in [10].

VI. Conclusion

Using Networked Music Performances for pedagogical scenarios requires a deep investigation on the topic in order to find metrics, factors and aspects that may help musicians to improve their musical skills.

In this paper, we introduced a framework for conducting perceptual experiments to continue this investigation for the purposes of the project InterMUSIC. We then presented two experiments conducted using the framework and the preliminary results that we observed.

Starting from these results, we described the areas of investigation that we intend to follow, with the final goal of developing a platform for NMPs in pedagogical scenarios that also work with general-purpose hardware.

Beyond this area, in future work we intend to further develop the formalization of the framework into an ontology, which will be integrated in the NMP tools, in order to collect and analyze a number of semantically-annotated rehearsal or lessons [5].

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 the participants of the pilot tests for their effort and their availability to be contacted for further experiments.

References

[1] J.Lazzaro and J.Wawrzynek, "A case for network musical performance", in *Proc. NOSSDAV Workshop*, Jun. 2001, pp. 157–166.

[2] M. Iorwerth, D. Moore and D. Knox, "Challenges of using Networked Music Performance in education", in *Proc. AES Conference*, Aug. 2015, pp. 157–166.

[3] T. Eerola and P. Toiviainen, *MIDI Toolbox: MATLAB Tools for Music Research*, Finland: University of Jyväskylä, 2004.

[4] N. Lushnikova, P. Chintakayala and A. Rodante, "Massive Open Online Courses from Ivy League Universities: Benefits and Challenges for Students and Educators" *in proc. of the XI International Conference Providing continuity of content in the system of stepwise graduate and postgraduate education*, Nov. 2012

[5] Ş. Kolozali and M. Barthet and G. Fazekas and M. Sandler, "Automatic Ontology Generation for Musical Instruments Based on Audio Analysis", in *IEEE Transactions on Audio, Speech, and Language Processing*, Oct. 2013, pp. 2207–2220

[6] A. Camurri and T. Moeslund, Chap. "Visual gesture recognition" in *Musical Gestures: Sound, Movement, and Meaning*, Edited by R. I. Godøy, and M. Leman, Routledge, 2010

[7] A. Canclini, L. Mucci, F. Antonacci, A. Sarti, S. Tubaro, "A Methodology for estimating the ratiationpattern of a violin during the performance" *in Proc. of the European Signal Processing Conference (EUSIPCO)*, 2015

[8] A. Olmos, M. Brulé, N. Bouillot, M. Benovoy, J. Blum, H. Sun, N. W. Lund, and J. R. Cooperstock, "Exploring the role of latency and orchestra placement on the networked performance of a distributed opera" *in 12th annual international workshop on presence*, 2009

[9] M. Goto, "An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds" in *Journal of New Music Research*, 30:2, pp. 159-171, 2010

[10] A. Nijholt, D. Reidsma, R. Ebbers and M. ter Maat, "The Virtual Conductor: Learning and Teaching about Music, Performing, and Conducting", in *proc. of the IEEE International Conference on Advanced Learning Technologies*, 2008

[11] S. Duffy, and P. Healey, "A new medium for remote music tuition", in *Journal of Music, Technology & Education*, 2017, pp 5–27

[12] UK 5G Innovation Network, Video: World's First 5G Distributed Music Concert, Web: https://uk5g.org/discover/read-articles/video-worlds-first-5g-distributed-music-concert/

[13] A. Canclini, L. Mucci, F. Antonacci, A. Sarti and S. Tubaro, "Estimation of the radiation pattern of a violin during the performance using plenacoustic methods" *in Audio Engineering Society Convention 138*

[14] A. Canclini, D. Markovi, L. Bianchi, F. Antonacci, A. Sarti, S. Tubaro, "A geometrical approach to room compensation for sound field rendering applications" *in Proc. of the European Signal Processing Conference (EUSIPCO)*, 2014

[15] L. Bonacina, A. Canclini, F. Antonacci, M. Marcon, A. Sarti, S. Tubaro, "A low-cost solution to 3D pinna modeling for HRTF prediction" in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016

[16] A. Barbosa, "Displaced soundscapes: A survey of network systems for music and sonic art creation", *Leonardo Music Journal*, pp. 53–59,2003,MIT Press

[17] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti "An Overview on Networked Music Performance Technologies", *IEEE Access*,4,2016

[18] E. Lakiotakis, C. Liaskos, X. Dimitropoulos, "Improving networked music performance systems using application-network collaboration", *Concurrency and Computation: Practice and Experience*,Wiley Online Library, 2018

[19] 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", *Journal of the Audio Engineering Society*, 63(11), 864-875.

[20] C. Chafe, J-P. Caceres, M. Gurevich, "Effect of temporal separation on synchronization in rhythmic performance", *Perception*, 39(7): 982-992, 2010

[21] C. Chafe and M. Gurevich, "Network Time Delay and Ensemble Accuracy: Effects of Latency, Asymmetry", *Proc. of the AES 117th Conf.*, SF, 2004

[22] C. Chafe, M. Gurevich, et al. "Effect of Time Delay on Ensemble Accuracy" *Proc. 2004 Intl. Soc. Musical Acoustics*, Nara, 2004

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

[24] M. Gurevich, D. Donohoe and S. Bertet, "Ambisonic Spatialization for Networked Music Performance", *International Community for Auditory Display*, 2011

[25] A. Carôt, C. Werner and T. Fischinger, "Towards a comprehensive cognitive analysis of delay-influenced rhythmical interaction", *ICMC*, 2009

[26] S. Farner, A. Solvang, A. Sæbø and Peter Svensson, A. Carôt, C. Werner and T. Fischinger, "Ensemble hand-clapping experiments under the influence of delay and various acoustic environments",*Journal of the Audio Engineering Society*, 57(12), pp 1028–1041, 2009

[27] 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*, Springer, 2013 pp.240–250,

[28] P. Holub, L. Matyska, M. Liška, L. Hejtmánek, J. Denemark, T. and Rebok, A. Hutanu, R. Paruchuri, J. Radil, and E. Hladká "High-definition multimedia for multiparty low-latency interactive communication", *Future Generation Computer Systems*, 22(8), pp.856–861, 2006, Elsevier

[29] J.-P. Cáceres, C. Chafe, "JackTrip: Under the Hood of an Engine for Network Audio", *Proceedings of International Computer Music Conference*, Montreal, 2009.

[30] S. Delle Monache, M. Buccoli, L. Comanducci, A. Sarti, G. Cospito, E. Pietrocola and F. Berbenni, "Time is not on my side: network latency, presence and performance in remote music interaction." *Proceedings of the XXII Colloquium on Musical Informatics (CIM)*, Udine,20-23 November, 2018.

[31] D. Markovic and F. Antonacci and A. Sarti and S. Tubaro, "Soundfield Imaging in the Ray Space", *IEEE Transactions on Audio, Speech, and Language Processing*, 21(12), pp. 2493-2505, 2013

[32] L. Bianchi, F. Antonacci, A. Sarti, S. Tubaro, "Model-Based Acoustic Rendering based on Plane Wave Decomposition", *Applied Acoustics*, Elsevier, 2016(104), pp. 127–134

[33] S. Vandemoortele, K. Feyaerts, G. De Bièvre, M. and Reybrouck, G. Brône, T. De Baets, "Gazing at the partner in musical trios: a mobile eye-tracking study", *Journal of Eye Movement Research*, Vol.11(2),2018 pp.6

[34] Yoshie, Michiko and Kudo, Kazutoshi and Murakoshi, Takayuki and Ohtsuki, Tatsuyuki, "Music performance anxiety in skilled pianists: effects of social-evaluative performance situation on subjective, autonomic, and electromyographic reactions", *Experimental Brain Research*, vol.199(2),22 Aug.2009,pp.117

[35] Wieslaw Woszczyk and William Martens, "Evaluation of virtual acoustic stage support for musical performance", *Journal of the Acoustical Society of America*, Vol. 123(5), 2008, pp. 3089

[36] Matthew Boucher, David Pelegrin-Garcia, Bert Pluymers, and Wim Desmet, "Auralization as a tool for evaluating an acoustical instrument", *Proc. of the Third Vienna Talk on Music Acoustics*, 1619 Sept. 2015, University of Music and Performing Arts Vienna, pp. 19