Impact of Source Panning on a Global Metronome in Rhythmic Networked Music Performance

Robert Hupke, Jürgen Peissig
Leibniz University Hannover
Institute of Communications Technology
Hannover, Germany
{hupke, peissig}@ikt.uni-hannover.de

Andrea Genovese, Sripathi Sridhar,
Agnieszka Roginska
New York University
New York, USA
{genovese, sripathi.sridhar, roginska}@nyu.edu

***Abstract*—Achieving synchronicity and tempo stability during a Network Music Performance (NMP) is not straightforward under heavy network latency conditions. Previous work on Global Metronomes has shown that it is possible to provide a universal time-reference signal to the connected nodes of an NMP. This paper illustrates the effects of using a global metronome and signal source-panning on a rhythmic performance. An experiment was conducted to evaluate the objective musical outcome and the subjective impressions of using these elements, applied to several pairs of Djembe percussion duets, under different tempo and latency conditions. The objective analysis in terms of tempo stability and synchrony, conducted from the perspective of an hypothetical audience, suggest that the use of a metronome achieves stability improvements at medium and higher latency levels, while the use of signal panning has been effective in improving the metronome efficiency. Subjective analysis data shows that the use of the metronome becomes challenging for higher delays while it was perceptually non-intrusive for the musicians for lower and medium delays.**

I. Introduction

High speed communication networks have enabled near real-time musical collaboration between physically remote locations. At a time in which creative and cultural professionals are looking for innovative solutions to connect with each other, it is worthwhile to offer innovative solutions to improve both, the musical interplay and the participants' feeling of being connected in a shared environment. In this context, the main goal is to improve the Quality of Experience (QoE) for the joint interaction within a networked music performance (NMP) towards an experience closer to a regular co-located musical performance. A significant point of interest in such NMP systems is the rhythmical nature and stability of musical interactions, which involve salient challenges such as network latency. As new technologies and strategies are explored as solutions for coping with signal delay, it is important for the technical components to avoid causing annoyance to the musicians and diverting the focus away from the performance. The design of and study of these solutions should therefore link the objective outcomes to the subjective experience of the users.

Previous studies on the effects of latency in NMPs [1]–[5], mostly focusing on rhythmic patterns performed by hand-clapping in pairwise interactions, have examined objective qualities like tempo stability, synchronicity and temporal separation. These experiments showed that delays shorter than $10\,\mathrm{ms}$ to $15\,\mathrm{ms}$ cause a performance tempo acceleration due to the subjects' intrinsic tendency to anticipate. The best synchronicity, while maintaining a stable tempo, can be achieved in a range between $10\,\mathrm{ms}$ to $25\,\mathrm{ms}$. In the "usability range" between $25\,\mathrm{ms}$ to $65\,\mathrm{ms}$, a tendency of tempo deceleration becomes discernible and coping strategies become applicable, whereas delays beyond this range lead to a significant deterioration of the performance. While there are validated general trends that can be understood from these results, these findings also represent a somewhat atypical musical task. Ecologically viable musical interactions have been further explored by linking rhythmic complexity and tonal instrument category to tempo variation. In [6] it was found that complex rhythms and higher spectral flatness (guitars, drum) lead to stronger deceleration patterns. Other studies connecting tempo and latency showed that the issue of objective and perceived temporal synchronization can be affected by the musical properties of the genre, signal attack, hierarchy of musical interaction, and musician's familiarity with network performance environments [7]–[10].

The observations of tempo deceleration at high delays leads to the assumption that the usage of a metronome can counteract the tempo drift and improve synchrony provided that a common pulse can be established and the delay is not too severe. A proposed solution came from [11], where a metronomic feedback pulse was dynamically adjusted to the variable delay network conditions, allowing the musicians to keep synchrony at the cost of tempo stability. A different, centralized, approach was proposed in [12] where a shared software environment would compute and broadcast a global pulse using the timestamps of incoming packets. Musicians would also be able to adjust the delay of the counterpart's stream to align the beat with a precise off-synch, provided the score tempo and network latencies are constant. This solution can achieve stability and synchrony, however, the real-time component of the interaction is somewhat hindered and does not generalize well to all kinds of musical interactions. A similar, perhaps opposite, strategy is to add artificial self-delay to align a musician's own feedback sound to the incoming transmission [13]. A more generalizable distributed approach for a global metronome was first proposed in [14], [15], where

ISSN 2305-7254

a satellite Global Positioning System (GPS) signal is received at each NMP location and used to generate a synchronized metronome pulse. The authors previously studied how this system can be used both as globally shared metronome [16] and as a system to calculate the holistic latency between a sender and receiver [17]. The system's impact on rhythmic scenarios were first explored in [18] where it was determined that the metronome was helpful to stabilize temporal trends in the presence of high latency levels.

While musical correctness and tempo stability can measure the objective success of the NMP, the QoE from the point of view of a musician also relies on psychological constructs of involvement and medium co-presence [19]. Delle Monache et. al. [9] categorize the "presence experience" with groups of constructs, such as *spatial presence and involvement*, *perceived realness*, and *interface awareness*, where the latter construct addresses the subjective impact of technical factors on attention and focused concentration. Post-experiment questionnaires interestingly revealed that poor audio quality can be perceived as interfering or distracting from performance. Under this premise, it is unclear whether the aforementioned delay coping mechanisms and interfaces would be detrimental, more than beneficial, to the performer's subjective standpoint. For example, self-delay strategies can help to minimize the temporal auditory dissonance between the connected pair [5], [18], but this option seems less viable for traditional instruments which present immediate haptic and acoustic feedback, leaving the question open on whether self-delay can lead to confusions and be detrimental in particular situations.

In the presence of multiple sound sources, literature indicates that spatial source separation (e.g. stereo panning) can improve cognitive attention and segregation of the auditory scene [20], [21]. In the context of NMPs, the nature of a paired interaction can be described as a *divided-attention* tasks between own sound (self-delayed or acoustic), partner sound, and possibly a metronome signal. Hence, it is possible to hypothesize that spatially separating these auditory events may aid the musicians' performance allowing more room for directing the cognitive attention where necessary and possibly improving the *immersion* character of the experience. The use of stereo panning in NMPs is often mentioned in projects addressing virtual NMP environments and interfaces [22], [23]. However, to the knowledge of the authors, the use of source-panning in NMPs is relatively unexplored through quantitative analysis.

There is room for further exploration of improvement strategies in rhythmic NMPs in regards to the QoE of the participants involved. By relating subjective user evaluations against the measurable outcome of a musical performance, it is possible to assess the ultimate impact of these strategies on the musicians' interplay, and achieve insights that can point to more intuitive, higher-quality, NMPs. Thanks to modern internet speed and the rich seminal work present in literature, it is also now possible to shift away from atypical musical tasks towards more traditional music scenarios.

This document presents an experiment designed as part of a larger series of collaborative studies between New York University (NYU) and Leibniz University Hannover (LUH). The previous work in section II discusses the framework infrastructure design for our studies on NMPs strategies and virtual collaborative environments as well as the implementation and testing of a global metronome device. Section III illustrates an experiment designed to evaluate the effectiveness of two latency-coping strategies. Specifically, it is aimed to further evaluate the impact of the global metronome and the introduction of signal panning techniques in an NMP interplay. These strategies are tested over a tonal percussion duet consisting of Djembe drummers playing under different tempo and artificial latency conditions. Results are discussed in section IV for measurable objective metrics, such as tempo stability and rhythmical alignment, and subjective trial evaluations on the quality of the interplay and the cognitive load brought by the use of said strategies.

II. Previous work

The joint academic collaboration between NYU and LUH brings together the *Holodeck* [24] and *LIPS* [25] projects, both centred on the development and study of novel immersive and interactive displays which can be used for augmented or virtual auditory applications. More specifically, this collaboration aims to study innovations and strategies for distributed collaborative musical immersive environments using a real-time data exchange framework, capable of transferring and recording different types of data through a central network server [26]. The resulting multimodal ecosystem aligns with the concept of the Internet of Musical Things, where immersive musical interactive displays are enabled by computing networks [27].

A. Framework

The framework infrastructure set in place between the two remote nodes is based on *CoreLink* [26], a network communication infrastructure capable of handling real-time transmission of several data types from a transmitter node to any number of subscribing receivers, via a central server. The system provides an API for locally encoding and decoding the various data streams at each node, and for customizing the data exchange according to network speed capabilities and local rendering needs. CoreLink is mainly accessed internally at NYU via a dedicated low-latency fibre-optic network, but it can be also accessed externally through special permissions. All data passing through the server can be optionally recorded, processed or analyzed through high-performance computing machines.

Fig. 1 schematizes the way the collaborative studies exchange data. Each local node collects and encodes the various stream types and sends them to the central server, which in turn distributes them to a subscriber node for decoding and rendering. The main audio types that can be exchanged consist of uncompressed multichannel audio (similarly to Jacktrip [28]), video, and motion capture (mocap) skeleton data [29]. All the data is embedded into a single timestamped transmission which is decoded and parsed at each receiving node

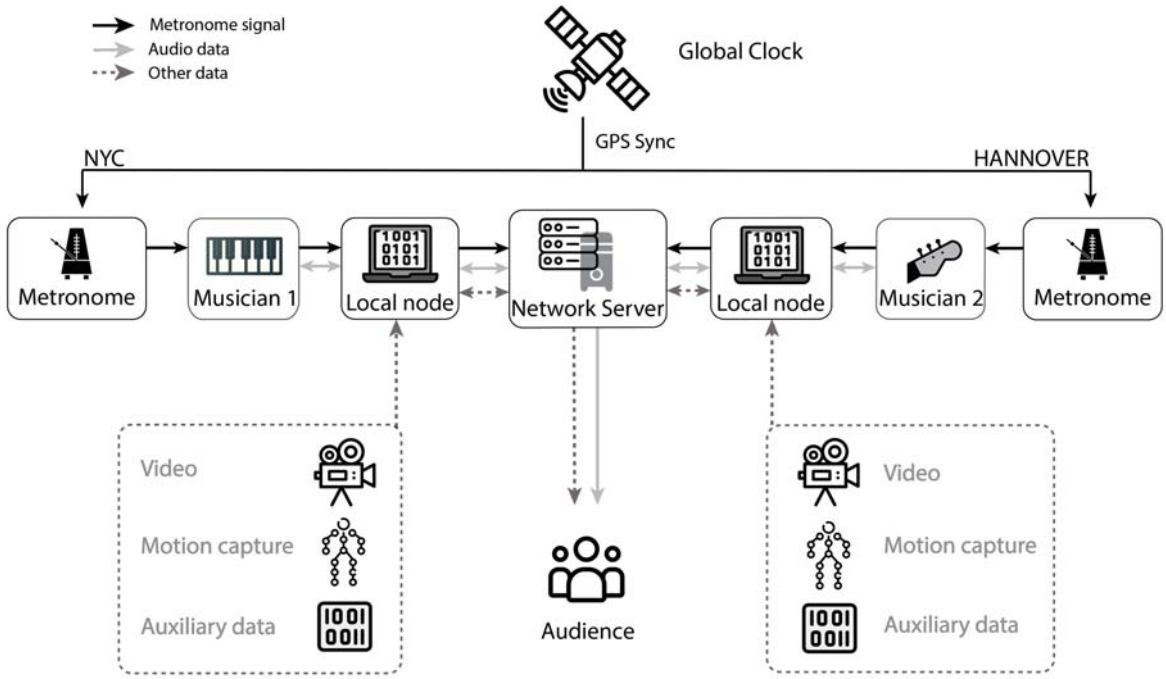


Fig. 1. NYU-LUH Networked Music Performance Framework

for local rendering and reproduction. The framework can be expanded by adding a Global Metronome unit which provides a universal synchronization signal to aid musical performance, and to possibly re-align the audio stream for distribution to a third-party audience. The metronome synchronization signal is computed from a satellite GPS signal, universally reachable from any part of the globe.

Rendering metadata can also be embedded as an auxiliary stream. Such data can be used for further adaptation of the local reproduction at a node to match the reproduction at the other nodes. This is a desirable feature for the creation of a shared distributed environment, where participants can transmit metadata for establishing a common virtual room. Some example include virtual object positions, acoustic filters, NMP parameters, spatial audio filters, EQ parameters.

B. Virtual Presence

From a QoE perspective, classical interaction of two participants in a NMP can be evaluated by plausibility, indicating the degree of how close the NMP is to real music participation (also for an audience). Literature on ensemble behaviour suggests that a visual connection between musicians is one of the key elements for maintaining a sense of plausible musical experience [30]. When a visual link is absent, an important layer of communication used by ensembles to communicate timing, musical expression, and interpretation, is missing [31]. Video streams have been often included in NMPs to study telepresence [32], but their immersive qualities are limited and the large data transmission rates imply higher latencies. Motion capture data is relatively low-bandwidth, as only skeleton positional points are transmitted, since the rendering happens locally at each node.

The use of embodied avatars in VR/AR NMP environments could help restore a sense of plausible agency while also maintaining the flexibility to create a cohesive environment, either symmetric (virtual common environment) or asymmetric and cohesive to the physical space of performance (locally adapted mixed reality). Another scale of assessment for the quality of network interaction is that of virtual co-presence [19], [23], [33], defined as "the sense of *being together* in a shared environment". While it is unclear if better co-presence may lead to objectively more accurate musical performances, it is possible to assume that the subjective judgement of the musical interplay might improve.

Generally, a relatively small number of contributions in the NMP field deal with motion-capture but more data is expected by future research involving both audio and embodied avatars. Within the context of our data-exchange framework, informal pilot experiments were able to establish a two-way real-time transmission of mocap skeleton data. Each receiving node was able to use the incoming data to render a virtual avatar character in a game-engine. Although the exact total latency between capture and rendering of motion capture data over this network is unclear, it is expected to be higher than that of audio alone. Future experiments will investigate on the impact of this offset and explore dedicated synchronization strategies. Other open questions that may be addressed involve particular body and instrument tracking scenarios, immersive spatial audio displays, and correlation studies between subjective quality ratings (plausibility, co-presence, etc.) and musical outcomes.

C. Global Metronome

The insertion of a global clock signal in the framework was motivated by several reasons. A global clock can primarily serve as a kind of metronome for remote musicians, which allows direct bidirectional interplay while maintaining a shared time-reference [16]. Furthermore, the same clock signal can be used to calculate signal latency between nodes [17], and as a time-alignment synchronization reference for distribution of the music stream to an audience by the central network server.

A Raspberry Pi (RPi) was used to implement the metronome as described in [14]. The basic concept of the global metronome is to use the GPS signal to synchronize the system time of the devices at all nodes of the NMP. Since the digital-to-analog audio converter of the Raspberry is unreliable and several tested methods for generating an analog audio signal did not produce a reliable, jitter-free audio output, an additional audio interface was used for accurately audio click generation. The authors showed in [16] that by the use of an audio interface (Focusrite Scarlett 2i2) an accuracy with a standard deviation of $\sigma = 56\,\mu$s can be achieved for the audio click signal of two GPS-synchronized metronomes. A method was presented by the authors in [17] to determine delay times in the signal processing chain between geographically remote nodes by the use of the metronome. Although it was mainly developed as a delay compensating tool for psychoacoustic investigations in NMPs, by using a direct Jacktrip connection [28] for the audio stream between NYU and LUH, it was shown that the metronome is suitable for the measurement of one-way and round-trip delay times in NMPs. In our particular case, a round-trip delay of $100\,\mathrm{ms}$ (one-way latency $50\,\mathrm{ms}$) was measured for a distance of $6200\,\mathrm{km}$ between our institutions.

D. Experimental investigations on the Metronome

In previous experimental studies, the impact of the metronome on the ensemble accuracy in terms of subject ratings, tempo stability and imprecision (as described in [10]) was investigated by several rhythmic experiments on NMP in [18]. Artificial delays up to $91\,\mathrm{ms}$ were inserted into the audio transmission between two subjects playing a complementary rhythm on a practice pad for drummers. Subjects were instructed to play the rhythm using a *leader-follower* interplay strategy [2], with and without the presence of metronome and the presence of self-delay matching the artificial delay amount.

The results of the experiment show that the global metronome leads to a stabilization of tempo acceleration caused by the delay, while the imprecision level stays constant up to a latency threshold of about $28\,\mathrm{ms}$ to $36\,\mathrm{ms}$ before deteriorating. It was found that alignment to the musical counterpart was easier when self-delay was applied. The interplay role also had an impact on both imprecision levels and subjective ratings, where whoever assumed the role of "leader" produced better results and higher ratings than the "follower". Subjective ratings and post-experiment interviews also suggested that, at the higher latency levels, playing with the metronome "felt more challenging" and it might have had a counterproductive effect.

Although the experiment was carried out with musically-proficient students, it was unclear how professional musicians who are used to perform together would react to the insertion of the global metronome. Another question is raised whether the use of the metronome in NMPs would also be accepted in music genres which normally do not use a metronome. Furthermore, the author points out that only a very simplified rhythm was used and that further investigations with different groups of instruments and more complex polyphonic arrangements should be carried out.

III. Experimental Study

The main motivation behind the experiment, presented in this contribution, was to observe the impact of the global metronome mechanism within a more ecologically viable ensemble configuration than the previously used hand-claps or drum pads, and playing a more complex rhythm. Furthermore, a headphone stereo-panning condition was included, in order to explore the effects of spatial source separation on cognitive focus and subjects' interplay ratings. An experiment was conducted at NYU to observe the objective and subjective effects of the global metronome and signal stereo-panning (and the combination of both), under different latency levels and target performance tempo.

A. Musical Components

Rhythmic interplay operates in a well-structured temporal framework, making it a challenging condition to explore and a paradigm better capable of highlighting the possible benefits of a global metronome. To these ends, the choice of instrumentation fell on a pair of *Djembe* drums. *Djembe* is a hand percussion acoustic instrument of west-african origin and an instrument which is typically used in the context of ensemble playing [34], [35]. Typical Djembe music is performed in ensembles that vary by region and context. A variety of rhythms exist that are associated with events of social and historical significance where most rhythms have at least two accompaniments, if not more. Since each drum is typically tuned to different pitches, the different djmebes take on unique roles within the performance [35].

A simulation of this paradigm is used on a small scale in this experiment, with two participants playing the *Sofa* rhythm [36], as seen in Fig. 2. This particular rhythm was chosen

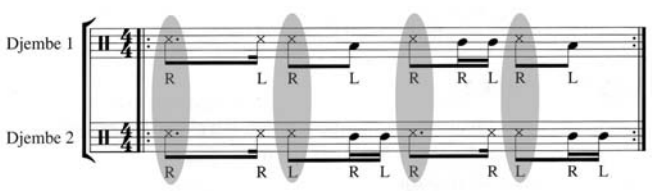


Fig. 2. Sofa rhythm- Rhythmic patterns for the two Djembe performers. The synchronization onsets (gray highlights) are used to determine the beat tempo.

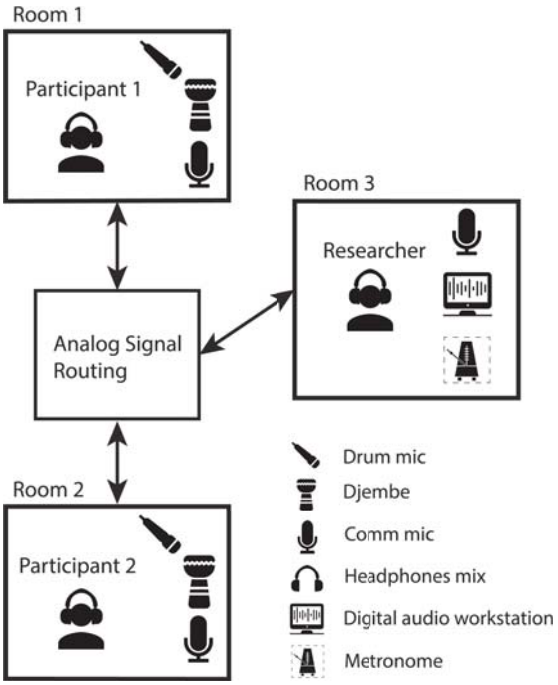


Fig. 3. Experimental setup. The researcher in room 3 monitors and routes the signals, controls the simulated NMP environment conditions and records the performance using a digital audio work station.

because it is relatively simple, while nevertheless being a reasonable approximation of a realistic performance in truncated form. Here the two participants' rhythms are different but complementary, and have points of synchronization at eighth notes and dotted eighth notes (gray highlight). The sequence in 4/4 time is one bar long, and repeated eight times in each trial of the experiment. This results in a recording of about 20 s in duration, subject to variation caused by the target tempo of the trial.

B. Methodology

Three subject pairs were recruited among undergraduate and graduate Music Performance students in the Department of Music and Performing Arts Professions at NYU Steinhardt. Subjects had between 8 and 16 years of musical experience, with an average of 11.2 years. Of the three subject pairs, two pairs of participants were not familiar with each other. One pair was very familiar and had performed together multiple times previously. Since the experiment required participants to engage in rhythmic interplay with another participant, one of the participants was allowed to take part in two instances of the experiment, playing opposite roles at each participation. Thus there were five unique participants among six subjects. At each experiment instance, Participant 1 was assigned to the *djembe 1* sequence from Fig. 2, and participant 2 to the *djembe 2* sequence.

The musicians were placed in two acoustically isolated rooms with no visual contact nor real-time feedback of the other participant's playing. The subjects used open back headphones in order to more clearly hear their own signal acoustically. A researcher was located in a separate room monitoring and controlling each trial. The experiment setup is depicted in Fig. 3 and shows how the researcher room controls the signal routing, manipulating the interplay conditions according to the trial variables and recording the performance. A talk-back microphone was also placed in each room for communication between trials during the experiment.

For this experiment, *source-panning* and *metronome* were the main components under investigation. In summary, the four conditions were the following:

- *Condition 1*: w/o panning, w/o metronome
- *Condition 2*: w/o panning, w/ metronome
- *Condition 3*: w/ panning, w/o metronome
- *Condition 4*: w/ panning, w/ metronome

The four conditions were explored under the influence of two independent variables, namely artificial *delay* and *tempo*. Each instance of the experiment thus consisted of 32 trials (repeated twice) comprised of unique combinations of *tempo*, *delay*, *source-panning*, and *metronome* presented in random order.

The metronome signal was routed from the researcher's workstation to each musician. It is important to note that the *leader* and *follower* strategy [18] was not followed in this experiment. Instead, a four measure metronome count-in track signalled the start of every trial. This track was modified according to the tempo of the trial, and depending on whether the current condition involved the enabling of the metronome. In the event that the trial did not have an active metronome (*condition 1*), the track was faded out by the end of the fourth measure. Conversely, the same metronome track was maintained through the eight measure sequence for trials whenever the metronome condition was active (*condition 2*).

In a similar vein, source panning was either enabled or disabled at each trial. Whenever the panning was disabled, the incoming transmission, and - if enabled - the metronome, were mixed in mono format over the headphones. If the panning was instead enabled (*condition 4*), the co-participant signal would be panned to 60° right and the metronome to 60° left in order to create a spatial separation of the two sources. In the case where the metronome was off then the co-participant panning would still take place (*condition 3*), allowing more room for the acoustic signal of the participant's own instrument to pass through the open-back headphones.

Artificial latency values of 10 ms, 25 ms, 50 ms and 100 ms were chosen to represent a range of varying levels of NMP playing difficulty. Since the baseline signal routing latency was measured to be 6 ms, a software delay was added to the music streams to achieve the trial latency values desired. The metronome signal was not affected by the added delay. The performance tempo levels were set at 90 bpm and 120 bpm. These were chosen to examine whether the performance dynamics and impact of the strategies vary with tempo. Differently than previous experiments, and due to the strong instantaneous haptic and acoustic response of a Djembe drum, no self-delay condition was deemed necessary.

Before every trial, participants had the opportunity to test their headphone levels, get familiar with the material and practice with the co-participant. Every trial was repeated twice in order to increase the sample size for objective analysis. At the end of each trial, participants were tasked with completing a short post-trial questionnaire for specific ratings on the current NMP conditions (Table I). An additional post-experiment questionnaire for more general feedback was compiled at the end of the session. The total experiment time was found to be around 90 minutes, including the time taken to fill out the post-experiment questionnaire for general feedback.

TABLE I. QUESTIONS AND LIKERT SCALES FOR THE END-TRIAL SUBJECTIVE QUESTIONNAIRE.

Question	Low(1)	High(5)
How would you rate the quality of musical interplay?	Very poor	Very good
How hard was it to maintain synchrony with your co-performer?	Impossible	Very easy
How difficult was it to distinguish the metronome sound from the co-performer's sound?	Could not distinguish	Very easy
Were the panning conditions useful or a distraction to your performance?	Very distracting	Very helpful

IV. ANALYSIS AND RESULTS

A. Objective Metrics

In order to perform an objective evaluation of the recorded audio signals, the played rhythm was analyzed based on the metrics presented in [10]. For one particular subject pair, the trials that used the source-panning condition (*condition 3* and *4*) had to be discarded due to a technical error where the system did not activate. This resulted in a total number of 320 recordings for analysis. The first step was to detect the individual beat onsets in the recordings. For this, the recordings were analysed by an automatic threshold peak detection. By windowing each detected event, a tangent through the maximum slope of the audio sample's envelope was calculated. The intercept of the tangent and the x-axis was selected as onset for each analyzed audio event and was stored as a timestamp. This procedure enabled a stable detection of the sound events, even with different playing styles of the musicians. For a more detailed description of the detection method we refer the reader to [17].

To validate the detected rhythm all timestamped onsets were assigned a note value by categorizing detected events into eighth and sixteenth notes, based on the time distance between consecutive notes. After this, a beat-pattern search was performed in order to detect potential performance mistakes and avoid the detection of wrong onset events. As the playing styles of the individual musicians were very different, 35 onset detection series had to be manually edited to account for certain individual hits not recognized by the algorithm. Eight recording files were discarded due to mistakes, incomplete performances and pauses within the performance. This resulted in a total of 312 recordings being used for evaluation. For this analysis, the metrics were computed starting from the second bar of the recordings, in order to remove some performance adjustment irregularities that were found in the first bar of some recordings. Finally, 28 timestamps were detected and evaluated on the synchronization points per trial.

Because of the interchanging rhythm of both subjects (Fig. 2), all onsets occurring at the time of the n-th quarter note within the beat were selected as the synchronization point t_n. The *Inter-Onsets Intervals (IOIs)*, defining the time in seconds between two consecutive beats, were calculated as $IOI_n = t_{n+1} - t_n$ and converted into beats per minute (bpm) by $\bar{\delta}(n) = 60/IOI_n$. On the basis of the IOIs we considered the following metrics: *pacing*, *tempo slope* and the *mean lag* between subjects. This last metric was adapted from the one of *asymmetry* [10] to consider absolute lag between players rather than the direction of the lag, since it was deemed a more useful perspective in the absence of clear leader-follower roles.

Pacing π is the average length of the IOIs calculated over a whole trial of N onsets as

$$\pi = \frac{1}{N} \sum_{n=1}^{N} IOI_n. \tag{1}$$

By calculating the inter subject time difference (ISD), which is the time differences between subject A and subject B ($t_{AB,n} = t_{A,n} - t_{B,n}$) it is possible to make an assessment about the performance synchrony between the two subjects. By calculating the mean time of the absolute ISD, the metric *mean-lag* is obtained:

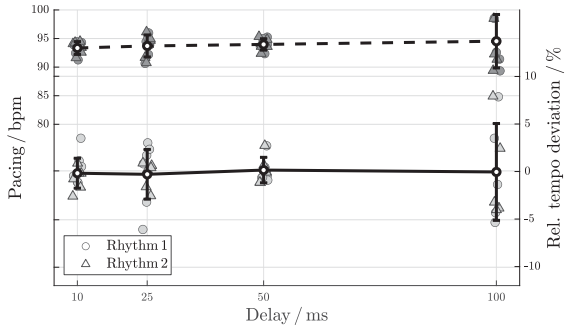
$$\alpha = \frac{1}{N} \sum_{n=1}^{N} \left| t_{A,n} - t_{B,n} \right|. \tag{2}$$

The *tempo slope* κ can be estimated by calculating the slope of a linear regression trough each $\bar{\delta}(n)$. This metric indicates the subjects tendency to accelerate or decelerate over the whole trial. To contextualize the severity of the tempo change in relationship to the initial performance tempo, this metric is reported as a percentage factor.

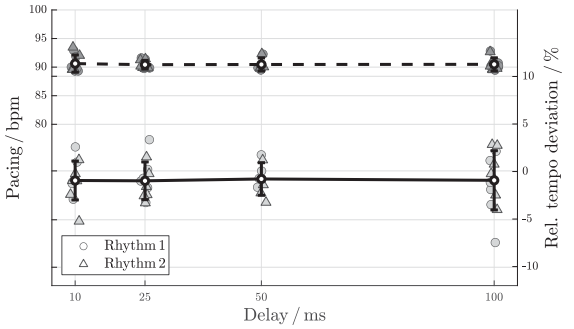
B. Objective Analysis

In this section we show the evaluation of all the four investigated conditions, with the aim to identify the objective effect of the introduced delay compensating strategies on the musical interaction. The results are separately presented for the different investigated tempos ($90\,\mathrm{bpm}$ and $120\,\mathrm{bpm}$), since different effects can be observed. For better comparison, the tempo slope κ is shown as the relative deviation from the starting tempo in percentage.

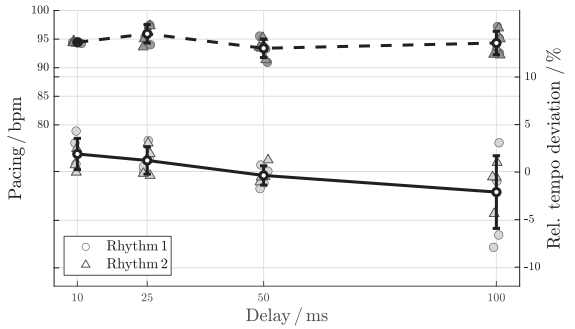
In Fig. 4 the relative tempo deviation (solid line) and pacing (dashed line) are shown for all four experiments with the predefined tempo of $90\,\mathrm{bpm}$. The error bars represent the



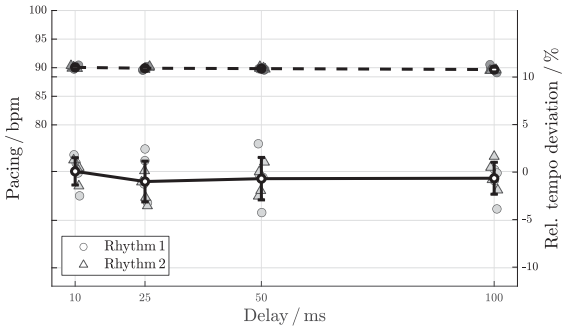
(a) *Condition 1*: w/o panning w/o metronome



(b) *Condition 2*: w/o panning w/ metronome

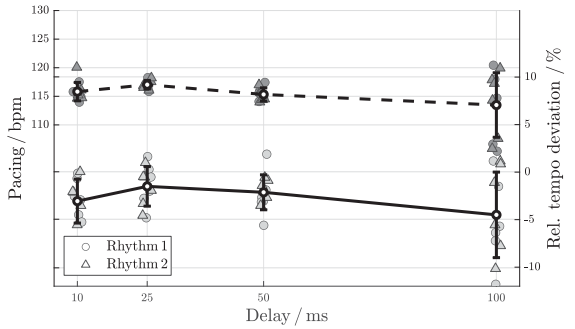


(c) *Condition 3*: w/ panning w/o metronome

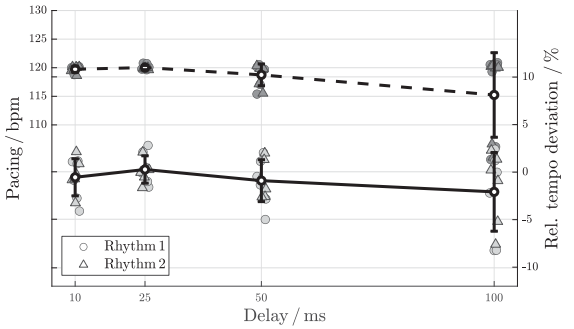


(d) *Condition 4*: w/ panning w/ metronome

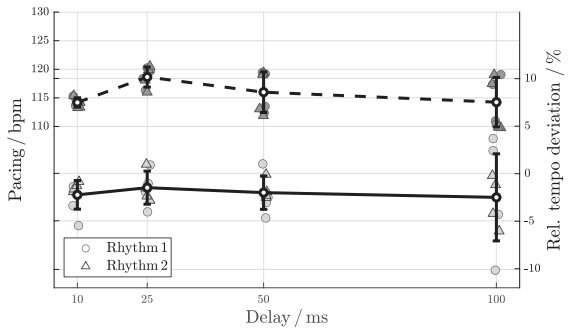
Fig. 4. Pacing (dashed line) and relative deviation from starting tempo (solid line) for all four tested delays at a predefined tempo of 90 bpm. Error bars show the mean and standard deviation. Actual values are separated for both rhythms (circle, triangle) and are randomly displaced by a small margin for better illustrating their density.



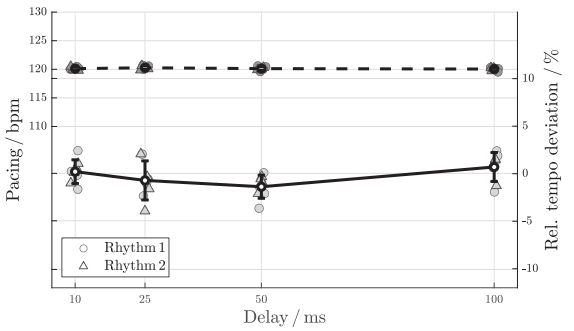
(a) *Condition 1*: w/o panning w/o metronome



(b) *Condition 2*: w/o panning w/ metronome

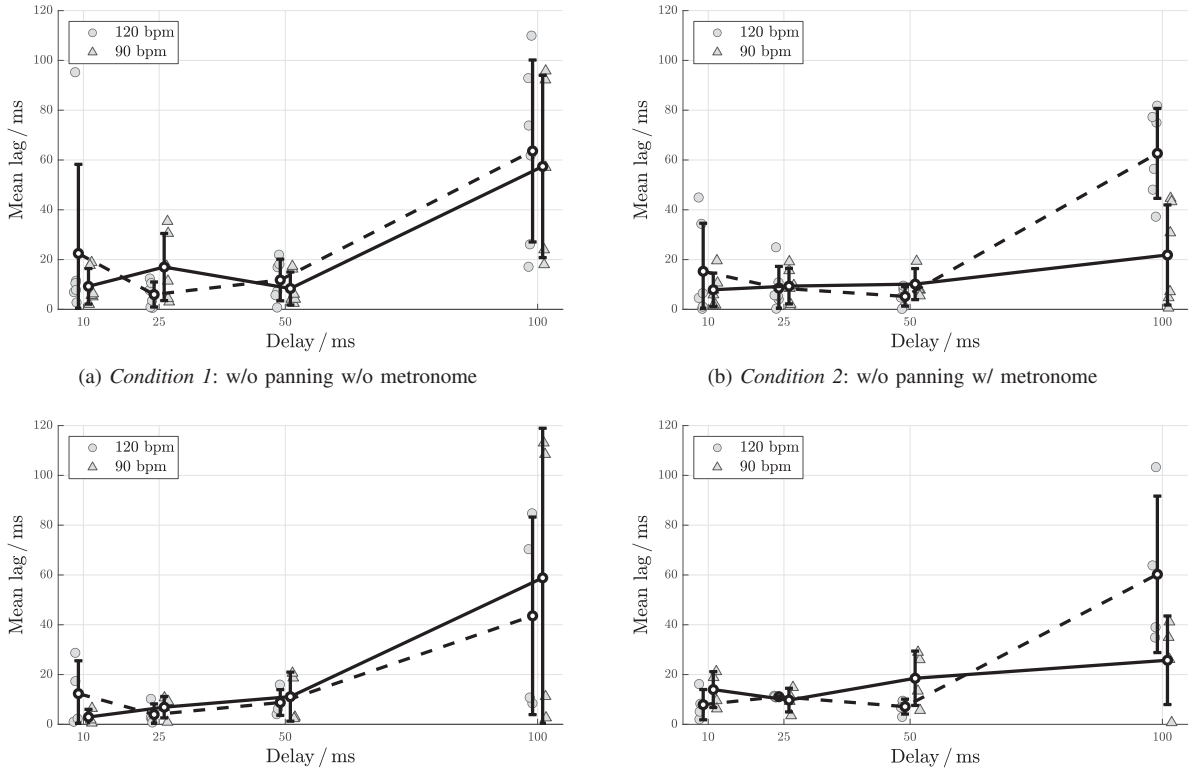


(c) *Condition 3*: w/ panning w/o metronome



(d) *Condition 4*: w/ panning w/ metronome

Fig. 5. Pacing (dashed line) and relative deviation from starting tempo (solid line) for all four tested delays at a predefined tempo of 120 bpm. Error bars show the mean and standard deviation. Actual values are separated for both rhythms (circle, triangle) and are randomly displaced by a small margin for better illustrating their density.



(a) *Condition 1*: w/o panning w/o metronome

(b) *Condition 2*: w/o panning w/ metronome

(c) *Condition 3*: w/ panning w/o metronome

(d) *Condition 4*: w/ panning w/ metronome

Fig. 6. Mean lag for all four tested delays. Error bars show the mean and standard deviation of both predefined tempos 90 bpm (solid line) and 120 bpm (dashed line). Actual values are separated for both tempos (circle, triangle) and are randomly displaced by a small margin for better illustrating their density.

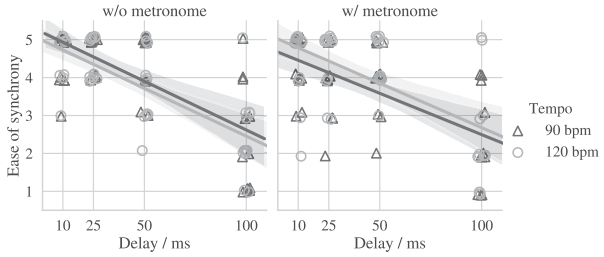
mean and the standard deviation of all trials per delay. The actual data computed for each delay level is shown separated for both djembe rhythms as scatter plots (triangle, circle). It can be seen from the data of all four conditions that the pacing stays nearly constant over all tested delays. The participants did not seem to be able to reach the given tempo of 90 bpm without the use of the metronome, as for *condition 1* (Fig. 4a) a mean tempo of 93.8 bpm, and for *condition 2* (Fig. 4c) a mean tempo of 94.5 bpm has been reached. Against the expectations, no tempo deceleration for higher delays can be observed in *condition 1* (Fig. 4a), which correlates with a small increase of pacing instead of an expected decrease at the delay of 100 ms.

By using the metronome, it can be seen that for both conditions, with and without panning (Fig. 4b and Fig. 4d), the subjects were able to follow the predefined tempo. By using additional stereo panning together with the metronome (Fig. 4d), the tempo deceleration of *condition 3* was stabilized and small decreases in the variances of both pacing and tempo deviation can be stated, even for high delays. No discernible differences between the two rhythms seem to have occurred.

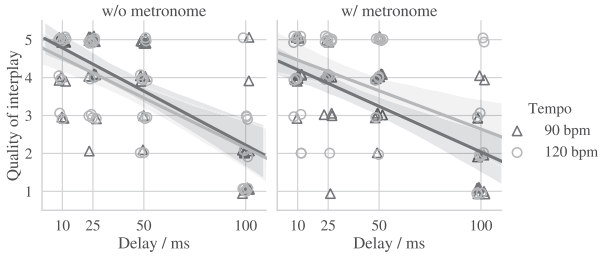
Comparing these results to the tempo of 120 bpm, it can be noticed that there are different tendencies on pacing and tempo acceleration. In Fig. 5 all four conditions of the 120 bpm trials are depicted. In comparison to the previous case, the musicians played at a slower average tempo over the individual delay times without the use of the metronome. Thus, an average tempo of 115.4 bpm for the experiment without metronome (Fig. 5a) and 115.7 bpm for the experiment with panning (Fig. 5c) are shown. However, the influence of the metronome seems to be smaller than the 90 bpm case, as a tempo deceleration was observed for the condition with metronome (Fig. 5b). Nevertheless, the usage of the metronome seems to lead to an adjustment of the given starting tempo. While the metronome alone seems to have just a small effect on tempo stabilization, by using the stereo panning together with the metronome, nearly constant values of pacing and tempo acceleration were achieved over the whole range of delays (Fig. 5d).

In Fig. 6 the *mean lag* for all four conditions is depicted separately for both tempos. As it can be seen from Fig. 6b and Fig. 6d, the metronome has a higher impact on synchrony on the trials at 90 bpm (solid line) than those at 120 bpm (dashed line), especially for high delays. For small and medium latencies in the range 10 ms to 50 ms, no meaningful impact of the metronome on the rhythmic alignment was found, since the *mean lag* values were relatively low also on the conditions with no metronome. However, a small decrease in the variance is found when the metronome was active.

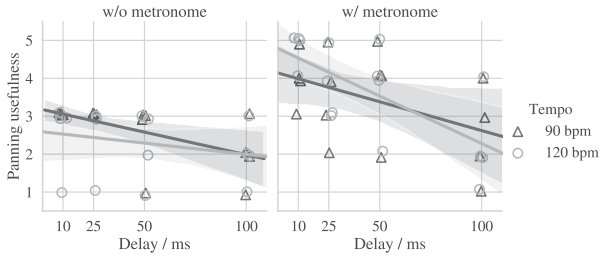
Nonetheless, at the delay level of 100 ms, the maximum recorded *mean lag* deviation for 90 bpm was reduced from 113 ms (Fig.6c) to 43 ms (Fig. 6b), and 41 ms (Fig. 6d) when panning was active. This corresponds to a decrease of



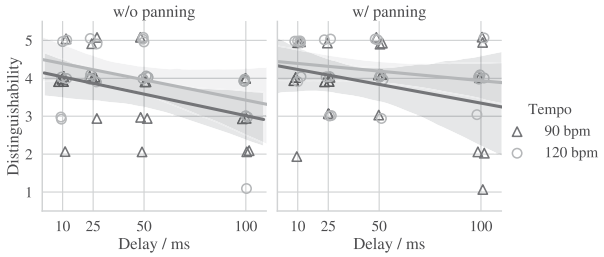
(a) Rated ease of synchrony



(b) Rated quality of interplay



(c) Rated usefulness of panning



(d) Rated distinguishability of metronome

Fig. 7. End trial questionnaire responses- a) Rated Ease of Synchrony, b) Rated Quality of Interplay, c) Rated Usefulness of Panning, and d) Rated Distinguishability of Metronome. In subfigures a), b), and c), plot on the left shows ratings for trials without metronome, and on the right with metronome. In subfigure d), plot on the left is without panning, right with panning. Added lines are linear regression estimates with 95% bootstrapped confidence intervals. Triangles and circles correspond to $90\,\mathrm{bpm}$ (dark line) and $120\,\mathrm{bpm}$ (light line). All points are randomly offset along the x and y axes to illustrate density.

misalignment from about a $1/16^{\text{th}}$ note to less than a $1/32^{\text{th}}$ note. No remarkable influence of source panning alone is noted.

C. Subjective Evaluation

Two questionnaires were drafted in order to explore the performer's subjective impressions on the quality of the performance in relation to the independent variables, and to interpret the ratings in relation to other objective results collected from the recorded signals. Participants were instructed to answer some evaluation questions at the end of each trial condition within an experiment session. The questions where rated on a 5-points likert scale (Table I) and aimed to rate the interplay quality, difficulty, spatial segregation and whether the metronome and latency strategies were distracting or useful, during the trial conditions.

Results of the end-trial questionnaires detailed in Table I are analyzed with respect to variations in *delay*, *tempo*, *metronome*, and *panning*. As expected, there is an observable trend in Fig. 7 where higher latencies negatively influence the scores for the perceived amount of effort and perceived interplay quality. No discernible positive or negative effects of the metronome and performance tempo were observed on the ease of synchrony (effort) ratings (Fig. 7a), nor the quality of interplay ratings (Fig. 7b).

The introduction of source panning tends to be seen as more helpful at lower latency values, and less helpful for higher delay values (Fig. 7c). However, there is a trend of panning being rated as more useful when the metronome is active indicating that the strategy was not deemed effective (sometimes detrimental) when just applied to a single stream to allow more of the personal acoustic sound to filter through the headphones. No significant effect of panning is observed on the question regarding source discrimination, where it was asked if it was difficult to distinguish the metronome sound from the co-performer stream (Fig 7d).

A second kind of questionnaire was compiled by the participants at the end of the experiment session, in order to contextualize particular pair combinations and obtain general feedback on the investigated elements. This final set of questions collected information on the participant's musical expertise, familiarity with NMPs, familiarity with the experiment partner, comments about the usefulness of the panning and the metronome and whether any other spontaneous delay coping strategy was elicited during the interplay. As part of this post-experiment questionnaire, subjects were asked if they used different cognitive and performance strategies to cope with latency on their own. Some subjects reported alternating between "locking in" with the other participant and the metronome, while others, particularly subjects with the most reported musical experience, focused on aligning with their internalized rhythm. One of the subjects described using "hocketing", a technique that involves playing a displaced rhythm with respect to the other performer [37], in order to try to cope with higher latency conditions.

V. Discussion

Summarizing the objective analysis, the results indicate that the use of source panning and the use of the metronome have a positive effect on tempo stabilization. Although the

individual impact of the metronome on tempo stability was more influential than that of source panning, best results were achieved when both were combined. Although in this experiment, source separation was only performed in a rudimentary way and the sound scene was well structured with only two musicians, results indicate that further investigations regarding source separation especially in NMPs with multiple participants should be conducted.

The effects of the metronome are in accord with our earlier observations in [18], which showed that the metronome achieves a tempo stability over the whole range of tested delay times, while enabling a more constant time-alignment (synchrony) of the subjects up to a delay time between $28\,\mathrm{ms}$ to $36\,\mathrm{ms}$. The results of this study show that the synchrony of the interplay hardly varies within a range between $10\,\mathrm{ms}$ to $50\,\mathrm{ms}$ and seems to be nearly independent from the usage of the metronome. However, the range was increased by about $20\,\mathrm{ms}$ compared to [18]. This leads to the assumption that we can be positive towards a NMP between NYU and LUH, as we would have to cope with a one-way delay of at least $50\,\mathrm{ms}$ [17].

It is somewhat surprising that the different starting tempos have led to a different tendency in pacing. Without the use of the metronome, the musicians could not maintain the given starting tempo and the tendency for both given tempos was contrary. A possible explanation for this might be the chosen group of instruments and rhythm. The effect could be also caused by the small group of subjects, and even though the subjects were professional instrumentalists, they had little experience in NMP. Alternatively, a too short adjustment time to the starting tempo, before each trial began, may have also caused this effect.

Tempo also had an impact for the synchrony lag at high-delays, where $90\,\mathrm{bpm}$ produced better time-alignments than $120\,\mathrm{bpm}$. A possible reason for this trend might be explained by the easier applications, at slower tempos, of the "locking-in" strategies that the musicians reported in the post-experiment questionnaire. As expected, our study demonstrated that the musical abilities of the subjects had a major impact on the interaction at higher delays. As reported by the musicians, they were able to intuitively adapt to the given situation and use strategies they had learned by playing together.

In the post-experiment questionnaire two of the six participants mentioned that "the metronome was distracting to their performance in the presence of high latency". Some participants also stated that the use of a metronome is "not typical in their style of Djembe performance", potentially affecting the perceived usefulness of the tool. In most cases participants did not notice panning. One subject noted that the panned stereo mix was unnatural. When relating the subjective results to what was found in the objective data, it is interesting to note that despite the metronome not having been perceived as particularly useful, the objective data shows improvements on stability. Similarly, source panning was also not perceived as particularly useful at higher delays, or even noticed. Yet, it brought some minor improvements when paired to the metronome. These disparities between the two layers of evaluation suggests that unconscious improvements to the performance can be achieved without necessarily distracting the performer. Arguably, the technological aid should be as non-invasive as possible, and non-awareness may be desirable.

It is also shown that when panning was applied alone and no metronome was present, the perceived effect was not rated as useful, sometimes distracting, but no objective detrimental effect was observed. Conversely, the panning was mostly rated as very useful when the metronome was on, this time in agreement with the objective result.

Future investigations - with a larger subject pool and more latency resolution in the range between $50\,\mathrm{ms}$ to $100\,\mathrm{ms}$, will expand the subjective inquiry into the internalized coping strategies and in which conditions were those more influential. Learning effects also are worthy of investigations as some musicians may adapt faster according to expertise type and level. This particular experiment can be expanded to include polyphonic ensembles and the use of three-dimensional spatial audio panning techniques for immersive displays. Additional instrumental and musical configurations are also due to be explored in order to make broader conclusions on the effects of the global metronome in relation to rhythm complexity, pitched tonality, and interplay roles. For a final validation, the strategies would also have to be assessed outside of laboratory conditions, throughout actual NMP concerts, in order to fully gather ecologically viable feedback by potential users.

VI. Conclusions

In this publication we illustrate the question of how could more realistic Networked Music Performances be studied and addressed in order to link the success of the outcome to the subjective quality of experience. We presented a potential framework of what a future NMP might look like and the setup of the cooperation between NYU and LUH, which uses a satellite signal to create a synchronized global metronome track on both sides of the NMP.

Through an experiment, this paper explores the effects of a global metronome and source panning, on three pairs of Djembe percussionists. These are discussed as possible assistance tools that may achieve objective and subjective improvements over rhythmic interplays on NMPs. Objective analysis indicates that source panning on the metronome can have a positive effect on tempo stabilization, although the influence of the metronome is much higher. Furthermore, it was shown that the effect of the metronome on the rhythmic alignment may depend on the playing tempo. Subjective evaluations show clear trends of lower quality ratings with higher delays, but no effect of tempo. When rating the perceived usefulness of the source panning, the ratings where not always in agreement with the objective outcome, suggesting that the two layers of evaluation are not always co-dependent.

Our experiment was useful to introduce the question of source-panning within the context of rhythmical NMPs, and

also provided additional insights into the objective and subjective impact of the global metronome strategy. The conclusions encourage further work into this line of studies, where deeper explorations on the experience of the musicians can highlight how attention and focus spans during the interactions. On a broader scale, future work over the NYU-LUH collaboration framework will explore the relationship between objective accuracy and quality of experience, using immersive technologies and novel strategies.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to all the performers who participated in the study and the collaborators who helped to draft the experimental design and setup. All icons used for the images available under Free and CC-BY licenses, we thank inipagistudio, Pixel perfect, Those Icons, Smashicons and Freepik (www.flaticon.com), Dale Humphries, Vectors Point, Arthur Shlain and Ian Rahmadi Kurniawan from the Noun Project, for making their work freely available.

REFERENCES

[1] C. Chafe and M. Gurevich, "Network time delay and ensemble accuracy: Effects of latency, asymmetry," in *Audio Engineering Society Convention 117*. Audio Engineering Society, 2004.

[2] C. Chafe, J.-P. Caceres, and M. Gurevich, "Effect of temporal separation on synchronization in rhythmic performance," *Perception*, vol. 39, no. 7, pp. 982–992, 2010.

[3] M. Gurevich, C. Chafe, G. Leslie, and S. Tyan, "Simulation of Networked Ensemble Performance with Varying Time Delays: Characterization of Ensemble Accuracy," *Significance*, 2001.

[4] S. Farner, A. Solvang, S. Asbjørn, and U. P. Svensson, "Ensemble Handclapping experiments under the influence of delay and various acoustic environments," *AES: Journal of the Audio Engineering Society*, vol. 57, no. 12, 2009.

[5] E. Chew, A. Sawchuk, C. Tanoue, and R. Zimmermann, "Segmental tempo analysis of performances in user-centered experiments in the distributed immersive performance project," in *Proceedings of the Sound and Music Computing Conference, Salerno, Italy*, 2005.

[6] 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.

[7] C. Bartlette, D. Headlam, M. Bocko, and G. Velikic, "Effect of Network Latency on Interactive Musical Performance," *Music Perception: An Interdisciplinary Journal*, vol. 24, no. 1, pp. 49–62, 2006.

[8] A. A. Sawchuk, E. Chew, R. Zimmermann, C. Papadopoulos, and C. Kyriakakis, "From remote media immersion to distributed immersive performance," in *Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence*, 2003, pp. 110–120.

[9] S. Delle Monache, L. Comanducci, M. Buccoli, M. Zanoni, A. Sarti, E. Pietrocola, F. Berbenni, G. Cospito, and M. Geronazzo, "A presence- and performance-driven framework to investigate interactive networked music learning scenarios," *Wireless Communications and Mobile Computing*, vol. 2019, 2019.

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

[11] J.-P. Cáceres, R. Hamilton, D. Iyer, C. Chafe, and G. Wang, "To the edge with china: Explorations in network performance," in *ARTECH 2008: Proceedings of the 4th International Conference on Digital Arts*, 2008, pp. 61–66.

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

[12] N. Bouillot, "njam user experiments: Enabling remote musical interaction from milliseconds to seconds," in *Proceedings of the 7th international Conference on New interfaces For Musical Expression*, 2007, pp. 142–147.

[14] R. Oda and R. Fiebrink, "The Global Metronome: Absolute Tempo Sync For Networked Musical Performance," *Proceedings of the International Conference on New Interfaces for Musical Expression*, vol. 16, 2016.

[15] R. K. Oda, "Tools and techniques for rhythmic synchronization in networked musical performance," Ph.D. dissertation, Princeton University, 2017.

[16] R. Hupke, L. Beyer, M. Nophut, S. Preihs, and J. Peissig, "A rhythmic synchronization service for music performances over distributed networks," in *Fortschritte der Akustik : DAGA 2019, 45. Jahrestagung für Akustik, Rostock*, 2019.

[17] R. Hupke, S. Sridhar, A. Genovese, M. Nophut, S. Preihs, T. Beyer, A. Roginska, and J. Peissig, "A latency measurement method for networked music performances," in *Audio Engineering Society Convention 147*, Oct 2019.

[18] R. Hupke, L. Beyer, M. Nophut, S. Preihs, and J. Peissig, "Effect of a global metronome on ensemble accuracy in networked music performance," in *Audio Engineering Society Convention 147*, Oct 2019.

[19] B. Jung, J. Hwang, S. Lee, G. J. Kim, and H. Kim, "Incorporating copresence in distributed virtual music environment," in *Proceedings of the ACM symposium on Virtual reality software and technology*, 2000, pp. 206–211.

[20] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

[21] B. G. Shinn-Cunningham and A. Ihlefeld, "Selective and divided attention: Extracting information from simultaneous sound sources," in *ICAD*, 2004.

[22] D. El-Shimy and J. R. Cooperstock, "Reactive environment for network music performance." in *NIME*, 2013, pp. 158–163.

[23] A. Genovese, M. Gospodarek, and A. Roginska, "Mixed realities: a live collaborative musical performance," in *Audio for Virtual, Augmented and Mixed Realities: Proceedings of ICSA 2019; 5th International Conference on Spatial Audio; September 26th to 28th, 2019, Ilmenau, Germany*, pp. 159–164.

[24] "The NYU Holodeck," http://holodeck.nyu.edu/, accessed: 2020-06-14.

[25] "The LIPS Project — Homepage," http://www.lips-project.de/, accessed: 2020-06-18.

[26] "NYU Corelink — Homepage," http://https://corelink.hpc.nyu.edu/, accessed: 2020-06-14.

[27] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet, "Internet of musical things: Vision and challenges," *IEEE Access*, vol. 6, pp. 61 994–62 017, 2018.

[28] Juan-Pablo Cáceres and Chris Chafe, "Jacktrip: Under the hood of an engine for network audio," *Journal of New Music Research*, vol. 39, no. 3, pp. 183–187, 2010.

[29] "Optitrack motive," https://optitrack.com/software/, 2020, accessed: 2020-04-11.

[30] G. Hajdu, "Embodiment and disembodiment in networked music performance," in *Body, Sound and Space in Music and Beyond: Multimodal Explorations*. Routledge, 2017, pp. 257–278.

[31] M. Iorwerth and D. Knox, "Playing together, apart: Musicians' experiences of physical separation in a classical recording session," *Music Perception: An Interdisciplinary Journal*, vol. 36, no. 3, pp. 289–299, 2019.

[32] C. Alexandraki and D. Akoumianakis, "Exploring new perspectives in network music performance: The diamouses framework," *Computer Music Journal*, vol. 34, no. 2, pp. 66–83, 2010.

[33] K. L. Nowak and F. Biocca, "The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 12, no. 5, pp. 481–494, 2003.

[34] R. Polak, "A musical instrument travels around the world: Jenbe playing in bamako, west africa, and beyond," *The World of Music*, vol. 52, no. 1/3, pp. 134–170, 2010.

[35] T. Y. Price, "Rhythms of culture: Djembe and african memory in african-american cultural traditions," *Black Music Research Journal*, vol. 33, no. 2, pp. 227–247, 2013.

[36] M. Markus and J. Galeota, "Beginning djembe : essential tones, rhythms and grooves," Boston, MA, 2016.

[37] J. K. Nketia, "The hocket-technique in african music," *Journal of the International Folk Music Council*, vol. 14, pp. 44–52, 1962.