

A Networked Music Performance Trio: A Case Study of Latency Mismatch

Aktera Teranontamongkol

Dept. Physics, Engineering and Technology
University of York
York, UK
sai.tera@york.ac.uk

Helena Daffern

Dept. Physics, Engineering and Technology
University of York
York, UK
helena.daffern@york.ac.uk

Gavin Kearney

Dept. Physics, Engineering and Technology
University of York
York, UK
gavin.kearney@york.ac.uk

Kenneth I. Brown

Dept. Physics, Engineering and Technology
University of York
York, UK
kenneth.brown@york.ac.uk

Abstract—Remote musical collaboration has never been more accessible. Despite some limitations, NMP allows musicians to perform together in real time over Wide Area Networks (WANs) regardless of physical location. Latency, or time delay, is one of the key underlying factors in NMP and largely determines whether real-time collaboration is practical. Studies suggest that One Way Trip (OWT) mouth-to-ear latency below 30 ms between remotely connected pairs is perceived as natural and does not require compensation, whereas adopting a leader-follower strategy can increase tolerance up to 50 ms, if not higher. As collaborative musical performance commonly extends beyond pairs in real-world settings, this study's interest therefore lies in the introduction of an additional musician.

This study contributes a case study of the effects of latency mismatch in trio NMP collaborations. It reflects a real-world use case of NMP, where latency between three peers can often be asymmetrical. For example, a pair of musicians (peers A and B) located in London, UK, may be connected to a third peer in Edinburgh, Scotland. Physical distance is often a major determinant of latency, as greater distances typically involve more network hops, resulting in increased delay. An experiment was conducted in controlled environments with a musical trio separated into different isolated live rooms within the music studios at the University of York. Without visual cues, as intended in the study design, they performed a series of musical tasks together using only binauralised audio cues under different controlled latency conditions between the threshold of 15 ms to 35 ms. Results indicate how a group of three musicians coped with symmetrical and asymmetrical latency conditions, investigating whether latency mismatch dominates or otherwise influences the effects of the underlying base latency.

Index Terms—Networked music performance, latency assessment, ensemble synchronisation, spatial audio

I. INTRODUCTION

Networked Music Performance (NMP) solutions, together with low-latency audio streaming technologies, have seen a clear increase in attention in recent years, with the release of new NMP software such as Bonza Music and FarPlay, as well as the continued development of existing platforms such as JamKazam, Soundjack, and JackTrip, making it more accessi-

ble than ever to the general public and broadening its range of use cases. This trend provides evidence that NMP is moving in the right direction, addressing a fundamental challenge in time-sensitive, synchronisation-dependent activities such as musical performance, and many other potential applications.

Having said that, there are yet a number of limitations to take into account in order to produce a successful NMP session, one of which is audio latency. Not only in NMP implementations, one of the sources of audio latency is present in every system and technology where audio is processed digitally, generally determined by audio buffer size (block size) and sample rate.

$$\text{latency (s)} = \frac{\text{buffer size}}{\text{sample rate}} \quad (1)$$

For example, a buffer size of 128 samples at a sample rate of 48,000 Hz will introduce 2.67 ms of latency to the system. Smaller buffer sizes reduce the time the system waits before processing the next audio buffer, thereby decreasing overall latency. In NMP systems, overall latency is influenced by several factors, including audio buffer sizes of the sound card (or audio interface) and within the NMP software, as well as network buffer size, jitter buffer size, and network latency, which is a direct function of the physical distance between peers and the number of network hops.

Previous studies of NMP concerning the effects of latency show a similar trend of tempo deceleration and synchrony deterioration in duo performance, often described in terms of the Ensemble Performance Threshold (EPT), defined as “the level of delay at which effective real-time musical collaboration shifts from possible to impossible” [1], at a One-Way Trip (OWT) latency of approximately 25–30 ms [2]–[8], or up to 40 ms [9], [10], while some work also reports tempo acceleration at latencies below 10 ms [3]. Given that most, if not all, studies in this area have focused on pairs of participants, there remains a gap in the understanding of latency effects in NMP

scenarios involving more than two performers, despite such configurations being common in real-world applications such as band rehearsal and general musical performance. Although some studies have considered larger ensembles, these have not necessarily involved real-time interaction between multiple performers. For example, [11] evaluated 5G architectures for NMP using up to eight emulated performers, with prerecorded audio from eight performers played back through eight NMP devices. Nevertheless, research involving more than two real performers remains limited. This study is therefore motivated to broaden the understanding of the effects of latency in three-performer scenarios.

Introducing a third performer increases the complexity of latency relationships, as latency can become mismatched or asymmetrical between peers, unlike in two-performer configurations, where symmetrical latency is typically assumed, as is the basis of most previous studies. Figure 1 shows a simple trio configuration in which Peers 1 and 2 are co-located and share a base latency of 15 ms, while Peer 3, located in London, experiences a 30 ms latency, resulting in a mismatch relative to Peers 1 and 2. As NMP inherently involves geographically distributed participants, such latency mismatches are common in practice. Accordingly, this study investigates the following research questions:

- **RQ1:** To what extent does latency mismatch influence the effects of the underlying base latency and performers' subjective experience in a three-performer NMP ensemble?
- **RQ2:** At what level of latency mismatch does this influence become apparent in the effects of the underlying base latency and performers' subjective experience?

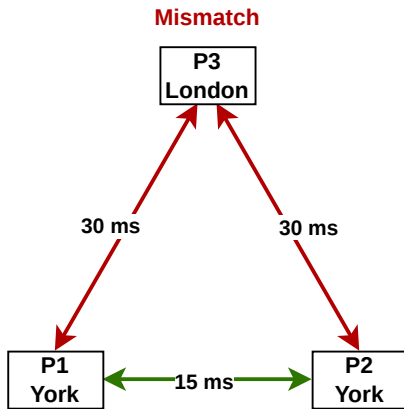


Fig. 1. A diagram illustrates an example of a latency mismatch scenario in a trio NMP collaboration.

In addition to aspects concerning latency, this study also employs an Immersive Network Music Performance (INMP) system, given the growing interest in INMP software such as Bonza Music, which incorporates spatial audio technologies, specifically binaural rendering together with head tracking, to provide participants with the perception of performing within a co-located simulated acoustic space. The benefits of binaural audio in NMP applications have been reported in

multiple previous studies, including [12], which suggests that Extended Reality Environments (XREs), combining virtual environments with spatial audio, can enhance participants' sense of immersion and perceived quality of experience of NMP without necessarily resulting in a corresponding improvement in measurable musical performance. Further evidence is provided by [13], where the majority of performers preferred binaural rendering with head tracking over a standard stereo setup for NMP, as well as for self-practising with backing tracks. Previous work has also shown that the use of Impulse Responses (IRs) with lower Reverberation Times (RTs) can result in lower objective asynchrony, suggesting improved timing precision, likely due to increased clarity [4]. Yet it remains unclear whether employing lower-RT IRs (e.g., 0.5 s), as suggested by previous work, yields greater overall synchrony than conventional NMP systems with mono or stereo output. Such a comparison, however, is beyond the scope of this study.

II. METHODOLOGY

A. Participants

A single group participated in this case study. The participants comprised two PhD students and one Technical Specialist in Immersive Audio at the AudioLab, Department of Physics, Engineering and Technology, University of York. The group consisted of a guitarist, a pianist, and a vocalist, who were existing bandmates and occasionally performed together. All participants had prior experience with NMP and were familiar with performing under latency conditions, while their research and professional work reflected a strong understanding of virtual and augmented reality, immersive audio, and psychoacoustics. The participants were not informed of the study hypotheses. Although the limited sample size restricts the generalisability of the findings, and the participants' prior experience with NMP may have influenced the results, this study serves as an initial case study in this area.

B. Experiment Setup

The experiment was conducted in a controlled environment, using isolated live rooms within the music studios at the University of York, which are acoustically treated to some extent. Rather than relying on a specific NMP software and network connectivity, which can exhibit latency-varying, the experiment was conducted entirely offline. REAPER served as the main workstation, receiving inputs from the three participants, introducing controlled latency conditions, recording the performances for subsequent Music Information Retrieval (MIR) analysis, and distributing binauralised return output signals to the participants' open-back headphones as if the participants were connected remotely over the Internet. This also enabled a high degree of precision and repeatability in latency control. Sixteen bidirectional analogue audio lines interconnected the three live rooms, allowing flexible audio routing between rooms, with Live Room 3 serving as the central REAPER workstation. As shown in Figure 2, each participant was assigned to a separate live room. The

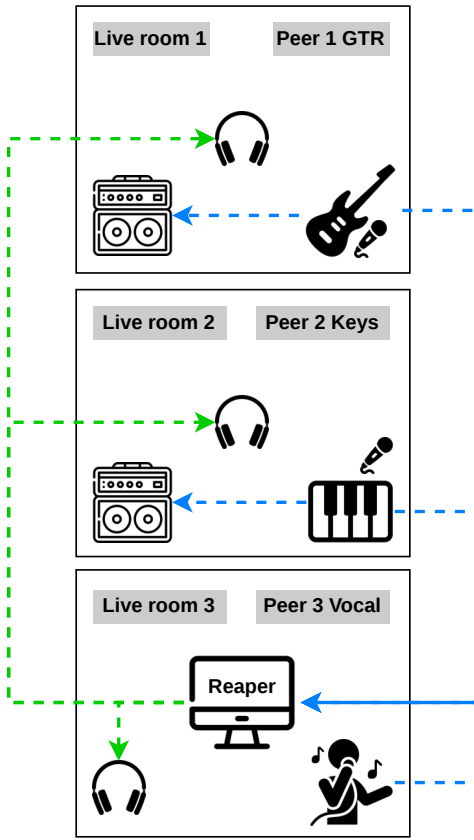


Fig. 2. A diagram illustrates the setup of the experiment, showing simplified audio routing between three participants in isolated live rooms within the music studios at the University of York.

guitarist performed on an electric guitar, the pianist on an 88-key Nord Piano, and the vocalist used a Shure SM7B dynamic microphone. Talkback microphones were provided to the guitarist and pianist to facilitate communication between trials, whereas visual cues were intentionally eliminated as part of the experiment design to prevent visual interaction between participants. Owing to the live room system's base latency of 13 ms, it was not possible to provide participants with direct monitoring, as they would have perceived their own signals with a 13 ms delay. Instead, the guitarist and pianist monitored their instruments via amplifiers, as their instruments do not produce sufficient acoustic output without amplification, unlike the vocalist.

Figure 3 shows the audio routing within the central REAPER system. To elaborate, when three remote peers are connected over the internet, a total of six one-way connections are established as follows:

- Peer 1 sends to Peer 2
- Peer 1 sends to Peer 3
- Peer 2 sends to Peer 1
- Peer 2 sends to Peer 3
- Peer 3 sends to Peer 1
- Peer 3 sends to Peer 2

In order to replicate this within REAPER, six dedicated mono

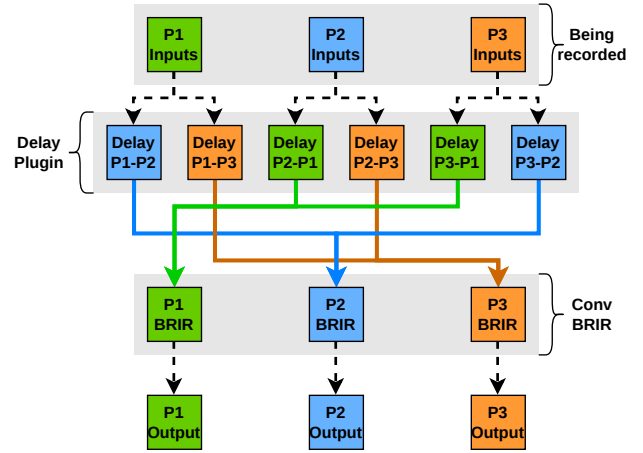


Fig. 3. A diagram illustrates the audio routing within the central REAPER workstation, from receiving input signals and introducing controlled latency conditions to applying BRIR convolution and distributing binauralised output signals to the relevant participants.

audio buses were configured (e.g., Delay_P1–P2 representing Peer 1 sending to Peer 2), with each participant's input routed accordingly. Each audio bus was processed using a bespoke VST3 delay plugin to independently control the send latency of each connection. The delay plugin, as shown in Figure 4, was developed solely for the purpose of this study to facilitate switching between latency conditions across trials and to compensate for the live room system's base latency. The effective latency introduced by the plugin corresponds to the remaining value after subtracting the system latency from the selected target latency (e.g., if the selected latency is 15 ms and the system latency is 13 ms, the plugin introduces 2 ms). Each connection bus was then routed to the corresponding stereo destination bus (e.g., Delay_P1–P2 was routed to P2_BRIR). Here the signals were convolved with a Binaural Room Impulse Response (BRIR) using REAPER's VST ReaVerb convolution plugin before the binauralised outputs were distributed to the corresponding participant. It should be noted that the performances were recorded at the initial stage of the signal chain, as shown in Figure 3, thereby capturing the raw, unprocessed performances prior to the introduction of latency or convolution processing.

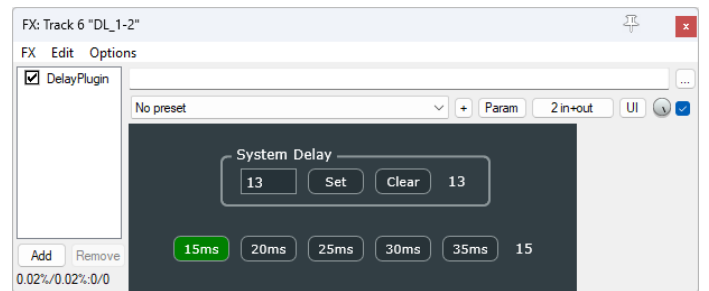


Fig. 4. A bespoke delay VST3 plugin developed using the JUCE framework, with system delay compensation and dedicated toggle buttons for each latency condition.

[illegible]

TABLE II
RANDOMISED EXPERIMENT TRIALS

Order	Trial	Tempo	P1	P2	P3	Order	Trial	Tempo	P1	P2	P3
1	1	Fast	15	15	15	21	39	Slow	Fully	Asym	
2	33	Slow		30		22	8	Fast		20	
3	24	Slow	30	30	30	23	9	Fast	25		
4	35	Slow	35			24	2	Fast	20	20	20
5	14	Fast			30	25	12	Fast	30		
6	6	Fast	20			26	27	Slow		20	
7	34	Slow			30	27	38	Slow	Fully	Asym	
8	31	Slow			25	28	17	Fast		35	
9	13	Fast		30		29	7	Fast		20	
10	16	Fast		35		30	15	Fast	35		
11	11	Fast			25	31	21	Slow	15	15	15
12	23	Slow	25	25	25	32	3	Fast	25	25	25
13	4	Fast	30	30	30	33	28	Slow		20	
14	22	Slow	20	20	20	34	19	Fast	Fully	Asym	
15	26	Slow	20			35	20	Fast	Fully	Asym	
16	32	Slow	30			36	36	Slow		35	
17	18	Fast	Fully	Asym		37	37	Slow		35	
18	25	Slow	35	35	35	38	30	Slow		25	
19	40	Slow	Fully	Asym		39	29	Slow	25		
20	5	Fast	35	35	35	40	10	Fast		25	

E. Data Collection

Alongside the objective data collected from the participants' performance recordings, participants were also asked to rate their comfort level for each trial on a scale from 1 to 5, where:

- 1 represents Very Uncomfortable
- 2 represents Uncomfortable
- 3 represents Neutral
- 4 represents Comfortable
- 5 represents Very Comfortable

Lower comfort ratings may be associated with reduced synchronisation, difficulty coping with latency, and tempo deceleration or acceleration. Participants were also asked to identify the best and worst trials in terms of overall performance experience.

III. RESULTS

A. Subjective Response

Figure 6 shows average comfort ratings given by the three participants across all trials in nominal design order (i.e., not the actual order of the randomised trials performed by the participants), with the Standard Deviation (SD) indicating the degree of variation. Dots highlighted in green represent the top three rated trials, while dots highlighted in red represent the bottom three rated trials. Participants' personal best and worst trials are also indicated using green and red labels, respectively (e.g., P3 B in green indicates Participant 3's personal best).

Figures 7 and 8 show average ratings for four incremental latency mismatch conditions, together with their Corresponding Symmetrical Latency (CSL) trials and the base latency trial at 15 ms, under fast- and slow-tempo conditions, respectively. The first half of each subfigure presents the base latency trial with a horizontal black dashed line to facilitate later comparison, together with the CSL trial with either a green or red dashed line. A green dashed line indicates that its

value is higher than that of the base latency trial, while a red dashed line indicates otherwise. The second half presents three mismatch trials for the same condition, in which each participant was alternately assigned the mismatch latency, followed by the average rating across the three trials, referred to as the Mismatch Average Rating (MAR). Similarly, Figure 9 presents average ratings for two fully asymmetrical conditions, in which all participants were assigned different latencies of 15, 25, and 35 ms.

B. Objective Result

A MIR method was used to extract meaningful datasets from the recorded performances for subsequent analysis. This study employed MATLAB MIRToolbox [14] together with MM-BOP [4], a MATLAB wrapper function for MIRToolbox onset detection algorithm. The wrapper improves the accuracy of the MIRONSETS function in detecting onset points for pitched instruments with sustained notes. It also provides various additional extracted information, including mean tempo, start tempo, tempo slope, onset distribution over time-domain plot, and tempo variation over time plot, some of which were used as part of the analysis in this study. It should be noted that the onset detection method used in MATLAB MIRToolbox was *Emerge* [15], which is described as "an improved version of the *SpectralFlux* method that is able to detect more notes and in the same time ignore the spectral variation produced by vibrato." Moreover, the MM-BOP codebase was modified to adapt the data output format to better suit the analysis and visualisation of the experiment results.

Figure 10 consists of four subfigures. The first three present average performance tempo slopes for three mismatch targets under 20, 25, 30, and 35 ms latency conditions, shown on the x-axis, with all symmetrical latency trials included as reference points and represented by white dots with black outlines. Two tempo conditions are presented at each x-axis position for direct comparison, with F on the left indicating fast tempo and S on the right indicating slow tempo. The fourth presents tempo slopes for the trials associated with the top three and bottom three comfort ratings, as well as the participants' personal best and worst trials. In line with the subjective results presented, Figures 11 and 12 share the same structure, presenting the four mismatch conditions together with the CSL and base latency trials, but with the corresponding average tempo slope data. Figure 13 follows the same approach for the fully asymmetrical conditions.

IV. DISCUSSION

According to Figure 6, the results suggest that the highest average comfort ratings were all associated with fast-tempo trials (i.e., the first half of the plot). The three highest-rated trials consisted of two mismatch latency conditions at 20 ms and one mismatch latency condition at 25 ms. The participants' personal best ratings followed a similar trend, with all three participants selecting fast-tempo trials as their best-performing conditions. Two participants selected fully asymmetrical latency conditions, while one selected a 20 ms

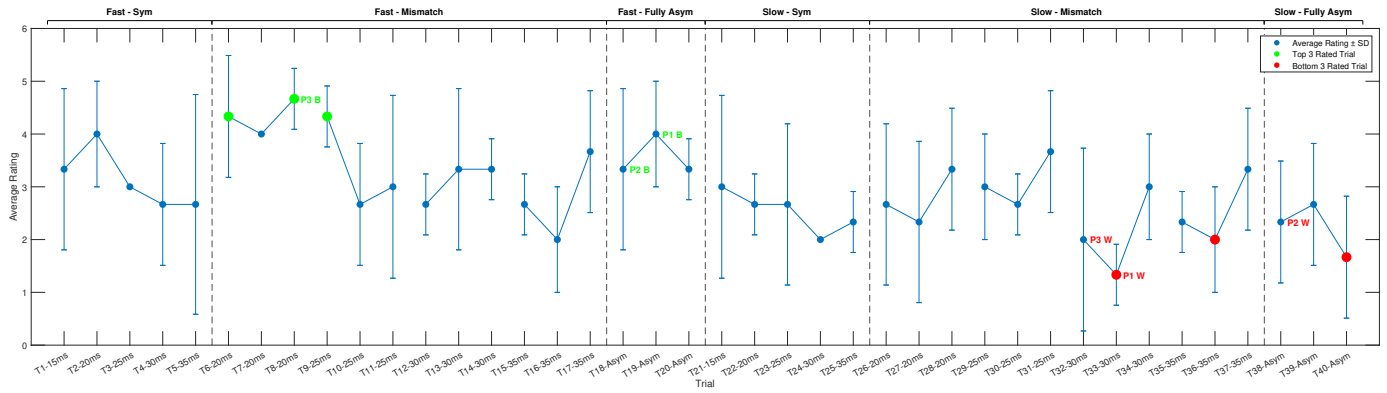


Fig. 6. Average comfort ratings across all trials in nominal design order, with the top three and bottom three rated trials highlighted, along with the performers' personal best and worst ratings labelled.

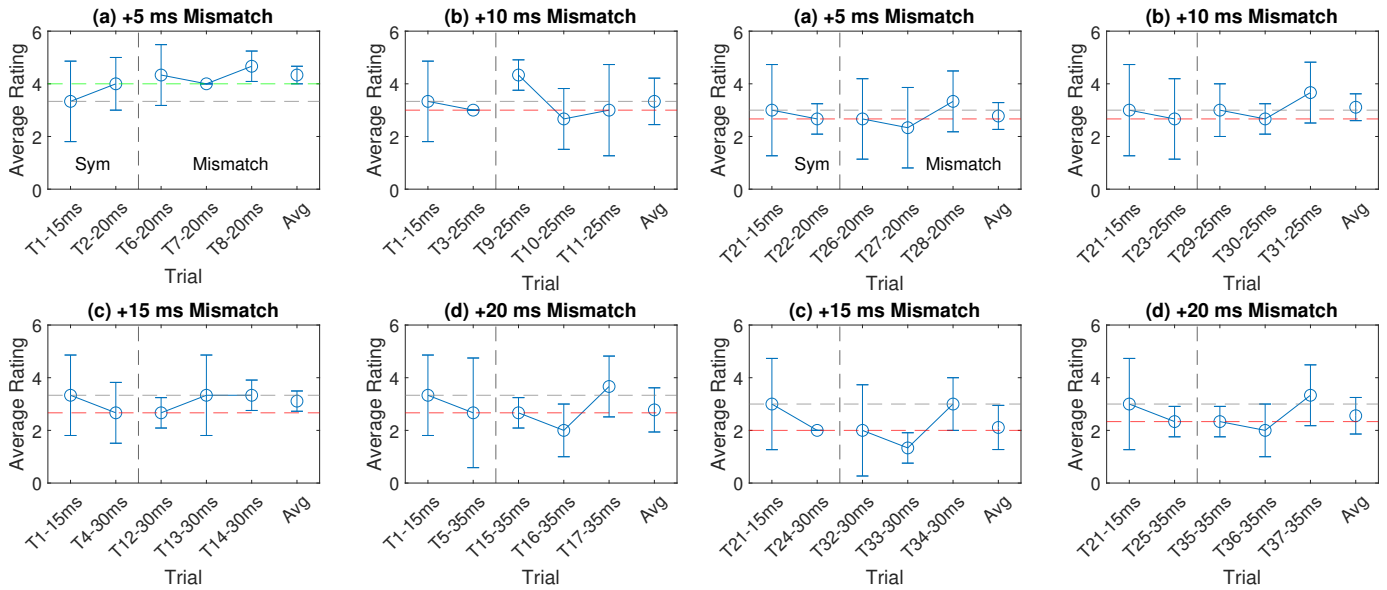


Fig. 7. Average comfort ratings for four mismatch conditions and their CSL trials, with the base latency trial of 15 ms, under fast-tempo: (a) +5 ms, (b) +10 ms, (c) +15 ms, and (d) +20 ms.

mismatch latency condition. The three lowest average comfort ratings were, as expected, all associated with slow-tempo trials. These comprised two mismatch latency conditions at 30 ms and 35 ms, and one fully asymmetrical latency condition. The participants' personal worst ratings showed a similar pattern, with two participants selecting the 30 ms mismatch latency condition and one selecting the fully asymmetrical latency condition. Here, it is interesting that the symmetrical 15 ms and 20 ms latency conditions, across both fast- and slow-tempo trials, did not result in a higher degree of average comfort ratings. This may serve as initial evidence that mismatch latency may influence performers' perceptions in a different manner compared with symmetrical latency conditions. Tempo differences also appear to play an important role in performers' comfort when performing under latency. Faster-tempo conditions generally resulted in higher comfort

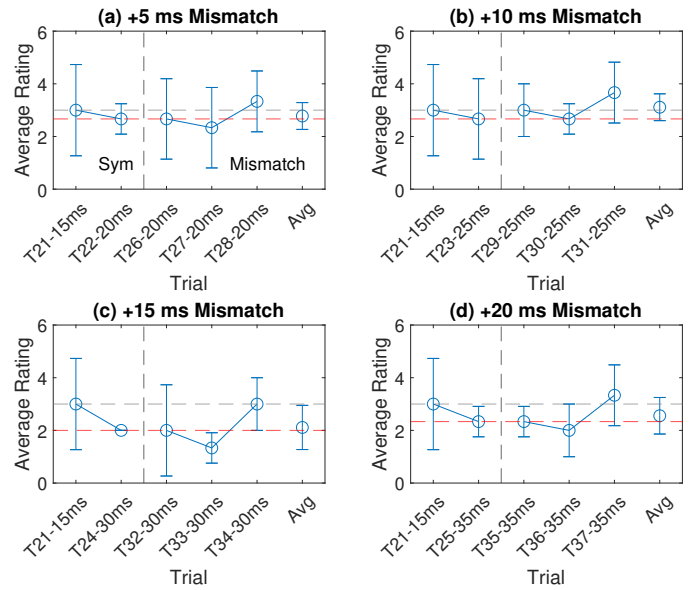


Fig. 8. Average comfort ratings for four mismatch conditions and their CSL trials, with the base latency trial of 15 ms, under slow-tempo: (a) +5 ms, (b) +10 ms, (c) +15 ms, and (d) +20 ms.

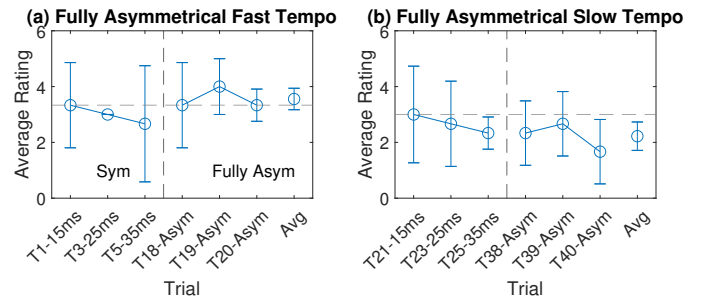


Fig. 9. Average comfort ratings for two fully asymmetrical conditions and their CSL trials, with the base latency trial: (a) fast-tempo, (b) slow-tempo.

ratings compared with slower-tempo conditions, suggesting that increased BPM, which reduces the temporal gap between beats, may result in better adaptation to latency. Conversely, slower-tempo conditions, which have greater temporal separa-

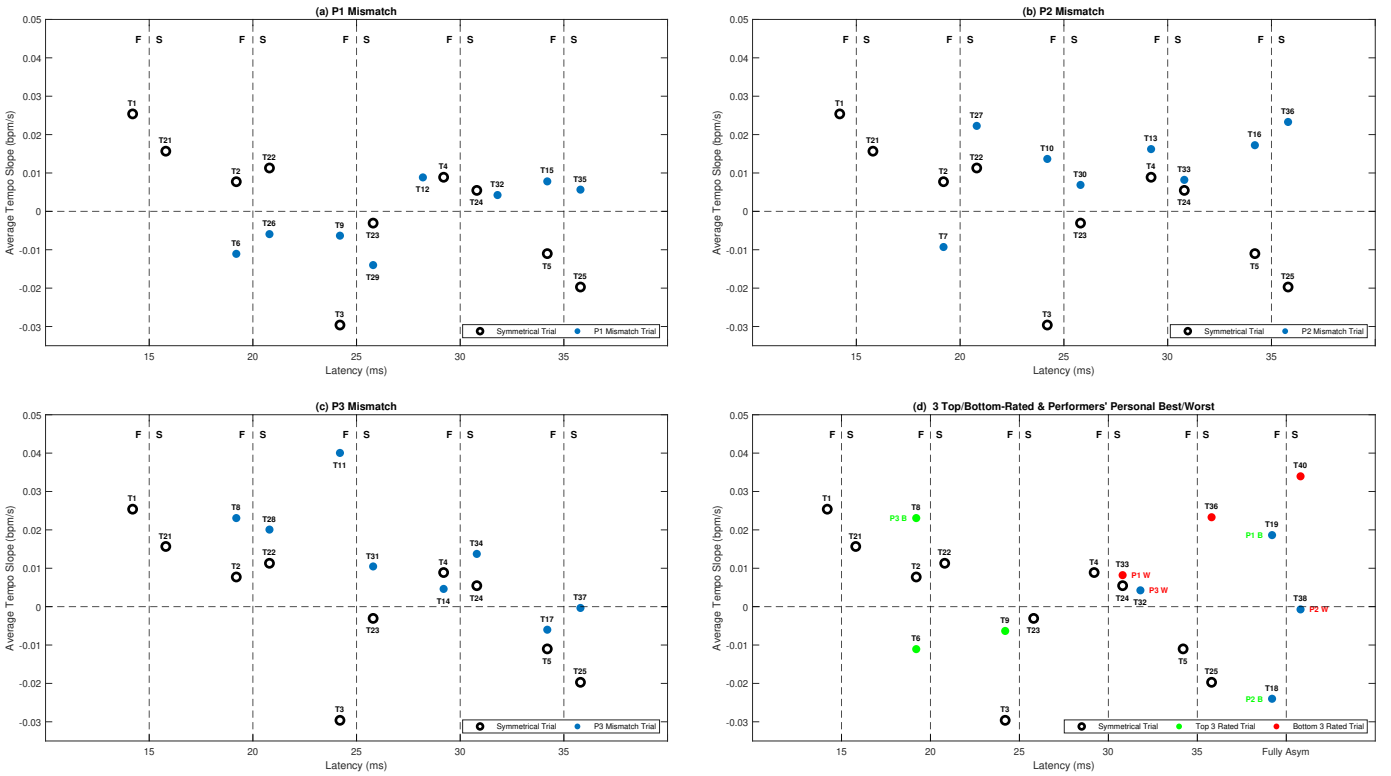


Fig. 10. Average performance tempo slopes for three mismatch targets under 20, 25, 30, and 35 ms latency conditions, alongside all symmetrical latency trials: (a) mismatch target is Participant 1, (b) Participant 2, (c) Participant 3, and (d) the three highest- and three lowest-rated trials by average comfort rating, together with the performers' personal best and worst trials labelled.

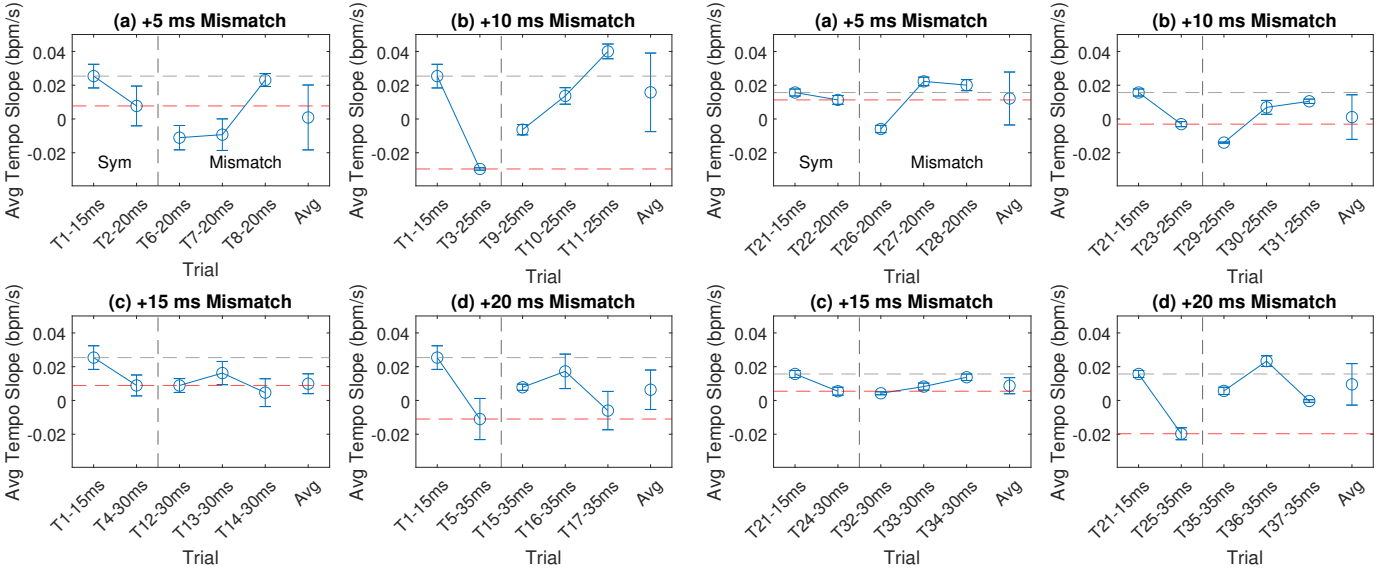


Fig. 11. Average performance tempo slopes for four mismatch conditions and their CSL trials, with the base latency trial, under fast-tempo: (a) +5 ms, (b) +10 ms, (c) +15 ms, and (d) +20 ms.

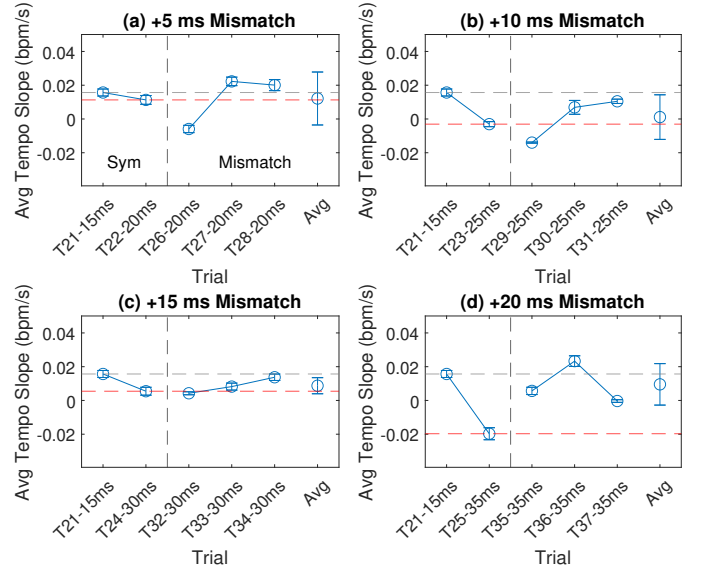


Fig. 12. Average performance tempo slopes for four mismatch conditions and their CSL trials, with the base latency trial, under slow-tempo: (a) +5 ms, (b) +10 ms, (c) +15 ms, and (d) +20 ms.

tion between beats, may provide greater perceptual awareness of timing deviations, making latency-induced asynchronies more noticeable and potentially reducing comfort levels.

Interestingly, this appears to differ from the findings of

Carôt et al. [16]. In their conventional drum-bass interaction experiment, they reported that the Ensemble Delay Acceptance Limit (EDAL) generally decreased with increasing tempo, indicating that musicians tolerated less latency as

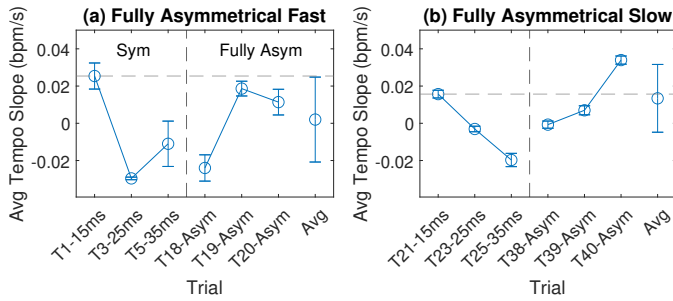


Fig. 13. Average performance tempo slopes for two fully asymmetrical conditions and their CSL trials, with the base latency trial: (a) fast-tempo, (b) slow-tempo.

tempo increased. Their second alternative interaction approach also showed no benefit in performance beyond 100 ms. In contrast, the present study found a more favourable subjective impression for faster rather than slower pieces. While the two studies employed different measures—latency-acceptance limits in Carôt et al. and subjective evaluations in the present study—the differing relationship between tempo and latency tolerance is noteworthy.

The results presented in Figures 7 and 8 are key to understanding the extent to which latency mismatch influences the effect of the underlying base latency and participants' subjective experience. Here, the average rating of the 15 ms base latency, shown as the first dot for each mismatch condition, was approximately 3.33, corresponding to a neutral rating according to the scale described in II-E. This value therefore serves as the baseline for the subjective effect of the base latency. The second dot represents the average rating of the CSL trial for each mismatch condition, providing a comparison of the base latency effect when the latency is symmetrical rather than mismatched. The last dot represents the MAR, indicating the extent to which the mismatch latency influences the subjective effect of the base latency. The closer the MAR is to the base latency rating, the less influence the mismatch latency appears to have on the effect of the base latency, and vice versa.

For the fast-tempo condition shown in Figure 7, although the difference between the base latency and CSL ratings is relatively small, particularly in (b), the +5 ms mismatch condition in (a), which is the only condition where the CSL rating is positive (i.e., higher than the base latency rating), appears to show a strong influence of the mismatch latency, potentially becoming dominant over the effect of the base latency. This is reflected in the MAR being higher than the positive CSL rating. The +20 ms mismatch condition in (d) also suggests a degree of influence on declining comfort, with the MAR being almost equal to the negative CSL rating. For the slow-tempo condition shown in Figure 8, the strongest influence on declining comfort appears in (c), followed by (d). (a) and (b), on the other hand, appear to have minimal to no influence. This suggests that this particular group was influenced to some degree by latency mismatch. In particular, they experienced greater comfort with a +5 ms mismatch than

with the base latency under the fast-tempo condition, and less comfort with the +15 ms mismatch under the fast-tempo condition and the +20 ms mismatch under both fast- and slow-tempo conditions.

Figure 9 shows a rather interesting observation. Under the fully asymmetrical condition with fast tempo in (a), there appears to be little to no influence of the latency mismatch, despite the CSL ratings for the 25 and 35 ms trials showing the expected linear decline in comfort relative to the base latency. Nevertheless, the influence of the fully asymmetrical condition becomes considerably stronger under the slow-tempo condition in (b). This further suggests that, for this particular group, faster-tempo pieces may have been associated with greater comfort when performing under latency conditions.

As for the objective MIR evaluation of the recorded performances, Figure 10(a), (b), and (c), which comprise average tempo slopes from all symmetrical latency and mismatch trials, provides evidence of tempo acceleration in more than half of the trials across the entire experiment. This somewhat contradicts previous studies, which reported that tempo acceleration between pairs of performers occurred at latencies below 10 ms [3]. (a), (b), and (c) in overall also show that at the 20 ms, 25 ms, and 35 ms latency conditions, the average tempo slopes of the mismatch trials do not exhibit the same trend as those of the symmetrical trials across both fast- and slow-tempo conditions. For example, under the 20 ms latency condition, the tempo slope of the symmetrical trial was approximately 0.01 BPM/s, whereas the mismatch trials ranged from -0.011 BPM/s to 0.023 BPM/s. A similar pattern was observed under the 35 ms latency condition, where the tempo slope of the symmetrical trial was approximately -0.015 BPM/s, while the mismatch trials ranged from -0.005 BPM/s to 0.024 BPM/s. Nevertheless, it is difficult to identify any meaningful pattern under the 25 ms latency condition, where the tempo slopes of the two symmetrical trials (fast and slow tempo) do not exhibit a similar range (one being -0.03 BPM/s and the other -0.002 BPM/s). Furthermore, one of the fast-tempo mismatch trials exhibited a tempo slope of 0.04 BPM/s, the highest observed in the entire experiment. This may indicate that the 25 ms latency conditions presented a greater challenge for this particular group of experienced participants than would be expected, as higher latency conditions would generally be anticipated to introduce greater performance challenges; yet, the tempo slopes observed under the 30 ms and 35 ms latency conditions did not exhibit the same level of variation. Interestingly, the 30 ms latency conditions showed a clear trend of minor tempo acceleration across both symmetrical and mismatch trials. Given that the symmetrical trials under the 15 ms latency condition exhibited greater tempo acceleration than most of the 30 ms conditions, it may be suggested that the 30 ms conditions were the least challenging for this group of participants. In addition, it is evident that, in most cases, when P2 (pianist, in (b)) and P3 (vocalist, in (c)) were assigned the mismatch latency, the average tempo slope of the corresponding trial became positive (i.e., acceleration). Conversely, when P1 (guitarist, in (a)) was assigned the mismatch latency, the average tempo slope of the

corresponding trial was more likely to become negative (i.e., deceleration).

Here, it is interesting that the subjective responses show in Figure 10(d), including the three highest and lowest comfort-rated trials and the participants' personal ratings, appear to contradict the tempo slope data from the objective MIR evaluation. The three highest-rated trials would be expected to exhibit tempo slopes closer to zero; however, only one out of five highly rated trials showed a tempo slope close to zero. Conversely, three out of five low-rated trials showed tempo slopes that were much closer to zero. This suggests that performers' perceived comfort may not be solely determined by their ability to maintain a stable tempo, as well as by the degree of tempo slope variation observed, with lower tempo slope variation not necessarily corresponding to a more comfortable performance experience. The results in Figure 11(a) further support this observation, showing a strong influence of mismatch latency on the decline in Mismatch Average Tempo Slope (MATs). This contrasts with the subjective evaluation, which suggests that this group experienced greater comfort under the same mismatch and tempo condition.

Figures 11 and 12 both show a trend of non-linear diminishing MATs compared with the base latency slope, which is expected given the influence of mismatch latency. In some mismatch conditions, this influence is little to none, such as (a) +5 ms and (c) +15 ms under the slow-tempo condition, whereas in others it becomes nearly or fully dominant, such as (a) +5 ms and (c) +15 ms under the fast-tempo condition and (b) +10 ms under the slow-tempo condition. Interestingly, for both (d) +20 ms fast- and slow-tempo conditions, the influence of mismatch latency does not appear to be as strong as expected, especially under the slow-tempo condition, where the MATs barely declines from the base latency slope. It is also interesting that the SDs of the individual tempo slopes for the mismatch trials (not MATs) under the slow-tempo condition are extremely small, indicating very little variation between individual tempo slopes. This could potentially indicate greater synchrony between the performers compared with the fast-tempo condition. From these objective data alone, it must be acknowledged that it is difficult to draw any meaningful conclusions here. However, the results suggest that, overall, this particular group may have been more influenced by mismatch latency under the fast-tempo condition than under the slow-tempo condition. The results for the fully asymmetrical trials further support this observation, as shown in Figure 13. The MATs shows a substantial decline due to the influence of mismatch latency under the fast-tempo condition, whereas the decline is only marginal under the slow-tempo condition.

Taken together, the subjective and objective findings provide some insight into the research questions posed in this study. Although the results should be interpreted cautiously given the use of a single group, the findings allow the influence of latency mismatch to be considered with respect to both the underlying base latency and the performers' subjective experience.

With respect to RQ1, "To what extent does latency mismatch influence the effects of the underlying base latency and performers' subjective experience in a three-performer NMP ensemble?", the findings suggest that latency mismatch influenced both the underlying base latency and the performers' subjective experience, although the extent and nature of this influence varied across mismatch and tempo conditions. In particular, the objective and subjective results did not always exhibit the same pattern, suggesting that the influence of mismatch latency may not be directly reflected in perceived comfort or objective performance measures.

With respect to RQ2, "At what level of latency mismatch does this influence become apparent in the effects of the underlying base latency and performers' subjective experience?", the findings do not indicate a single mismatch level at which its effects consistently become apparent. Instead, effects were observed at different mismatch levels depending on the tempo condition and whether subjective or objective measures were considered. This suggests that the point at which mismatch latency becomes noticeable may depend not only on the magnitude of the mismatch but also on the performance context and the specific measure examined.

V. CONCLUSION

A. Summary of Work and Findings

This study presented a case study investigating the effects of latency mismatch in a trio NMP collaboration. Previous NMP literature involving pairs of musicians was reviewed, and an offline experiment was conducted in live rooms within the music studio at the University of York with one group of three participants who had prior NMP experience. REAPER was used as the main workstation to receive, process, and distribute binauralised output signals, simulating a scenario in which all participants were connected remotely over the internet. The participants completed 40 trials comprising different latency and tempo conditions. Subjective responses and objective performance recordings were collected and analysed using MIR methods.

To summarise the findings, no definitive conclusions can be drawn given the small sample size of this study. However, the subjective and objective evaluations suggest that, for this particular group of experienced participants, mismatch latency conditions generally exhibited effects that differed from those observed under symmetrical latency conditions. Nevertheless, the specific nature and extent of these effects cannot yet be determined, and further data from larger and more diverse participant groups are required for a more comprehensive evaluation. Different tempo conditions also appeared to influence comfort levels and performance experience, with faster-tempo conditions generally associated with greater comfort when performing under different latency conditions, whereas slower-tempo conditions were more frequently associated with reduced comfort.

B. Study Limitations

The major limitation of this study was the explicitly small sample size of only one group of three experienced NMP performers. This substantially limits the extent to which meaningful conclusions can be drawn from the findings, as the observed results may reflect characteristics specific to this particular group rather than effects that are representative of NMP performers more generally. Consequently, the findings should be interpreted as an initial case study rather than as generalisable evidence of the effects of latency mismatch in three-performer NMP.

The study was also limited by the inherent base latency of the live room system, which was 13 ms. As a result, direct monitoring was not provided to the participants. If direct monitoring had been available through their headphones, the results may have differed, as participants may have experienced a different relationship between their own performance feedback and the delayed return signals from other participants, potentially influencing their adaptation to latency, subjective comfort ratings, and performance synchronisation.

C. Further Work

Given the application of adaptive metronomes in NMP systems has already been explored with pairs of musicians [17], [18]. Extending this approach to trio collaborations would provide valuable insight into whether adaptive timing strategies remain effective when additional performers and more complex latency interactions are introduced.

With regards to INMP and the effects of different RTs on perceived clarity and synchronisation, exploring a more extended range of RT IRs would provide a more comprehensive understanding of how room acoustics influence trio collaboration.

ACKNOWLEDGMENT

This study was supported by the Knowledge Transfer Partnership (KTP) programme, funded by Innovate UK, in collaboration with Bonza Music Ltd and AudioLab, University of York, under grant reference KTP 010857 (KTP programme number KTP13962), and by CoSTAR Live Lab, funded by the Arts and Humanities Research Council (AHRC), under grant reference AH/Y001079/1.

The authors would also like to acknowledge the contributions of the three participants from AudioLab, University of York, who participated in this study.

REFERENCES

- [1] N. Schuett, "The effects of latency on ensemble performance," Master's thesis, Stanford University, Stanford, California, USA, May 2002.
- [2] S. Farner, A. Solvang, A. Sæbø, and U. P. Svensson, "Ensemble hand-clapping experiments under the influence of delay and various acoustic environments," *Journal of the Audio Engineering Society*, vol. 57, no. 12, pp. 1028–1041, 2009.
- [3] C. Chafe, M. Gurevich, G. Leslie, and S. Tyan, "Effect of time delay on ensemble accuracy," in *International Symposium on Musical Acoustics*, Nara, Japan, 2004.
- [4] P. Cairns, "Immersive audio and network music performance: Practical applications and user testing," PhD thesis, University of York, School of Physics, Engineering and Technology, Feb. 2025.
- [5] C. Chafe, J.-P. Cáceres, and M. Gurevich, "Effect of temporal separation on synchronization in rhythmic performance," *Perception*, vol. 39, no. 7, pp. 982–992, 2010.
- [6] P. Cairns, H. Daffern, and G. Kearney, "Parametric evaluation of ensemble vocal performance using an immersive network music performance audio system," *Journal of the Audio Engineering Society*, vol. 69, no. 12, pp. 924–933, 2021.
- [7] —, "Immersive network music performance: Evaluation of the effect of auralized reverberation time," *Journal of the Audio Engineering Society*, vol. 73, no. 12, pp. 862–871, 2025.
- [8] A. Carôt, C. Hoene, H. Busse, and C. Kuhr, "Results of the fast-music project—five contributions to the domain of distributed music," *IEEE Access*, vol. 8, pp. 47 925–47 951, 2020.
- [9] K. Tsioutas and G. Xylomenos, "Assessing the effects of delay to nmp via audio analysis," *SN Computer Science*, vol. 4, no. 126, 2023.
- [10] K. Tsioutas, G. Xylomenos, and I. Doumanis, "Impact of audio delay and quality in network music performance," *Future Internet*, vol. 17, no. 8, p. 337, 2025.
- [11] L. Turchet and P. Casari, "On the impact of 5g slicing on an internet of musical things system," *IEEE Internet of Things Journal*, vol. 11, no. 19, pp. 32 079–32 088, 2024.
- [12] R. Hupke, S. Preihs, and J. Peissig, "Immersive networked music performance: Impact of extended reality on the quality of experience," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*, 2024, pp. 1–9.
- [13] M. Tomasetti and L. Turchet, "Playing with others using headphones: Musicians prefer binaural audio with head tracking over stereo," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 3, pp. 501–511, June 2023.
- [14] O. Lartillot and P. Toivainen, "Mir in matlab (ii): A toolbox for musical feature extraction from audio," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, Sep. 2007, pp. 127–130.
- [15] O. Lartillot, D. Cereghetti, K. Eliard, W. J. Trost, M.-A. Rappaz, and D. Grandjean, "Estimating tempo and metrical features by tracking the whole metrical hierarchy," in *Proceedings of the 3rd International Conference on Music & Emotion (ICME3)*, Jyväskylä, Finland, Jun. 2013.
- [16] A. Carôt, C. Werner, and T. Fischinger, "Towards a comprehensive cognitive analysis of delay-influenced rhythmic interaction," in *Proceedings of the International Computer Music Conference (ICMC)*, Montreal, Canada, 2009.
- [17] R. Battello, L. Comanducci, F. Antonacci, A. Sarti, S. D. Monache, G. Cospito, E. Pietrocola, and F. Berbenni, "An adaptive metronome technique for mitigating the impact of latency in networked music performances," in *2020 27th Conference of Open Innovations Association (FRUCT)*, 2020, pp. 10–17.
- [18] R. Battello, L. Comanducci, F. Antonacci, G. Cospito, and A. Sarti, "Experimenting with adaptive metronomes in networked music performances," *Journal of the Audio Engineering Society*, vol. 69, no. 10, pp. 737–747, 2021.