On the Impact of Audio Characteristics to the Quality of Musicians' Experience in Network Music Performance

Konstantinos Tsioutas and George Xylomenos
Mobile Multimedia Laboratory, Department of Informatics
School of Information Sciences and Technology
Athens University of Economics and Business
Patision 76, Athens 10434, Greece
E-mail: {ktsioutas, xgeorge}@aueb.gr

***Abstract*—Audio delay is a crucial factor in the Quality of Musicians' Experience (QoME) in Network Music Performance (NMP). Previous studies have explored the tolerance of musicians to delay and its dependence on the timbre of the instruments used and the tempo of the performance. Although their findings are intriguing, the small size of these studies makes the extraction of concrete conclusions quite hard. In order to shed more light on these issues, we undertook a larger scale NMP study with real musicians, assessing a wide range of subjective QoME variables against delay, and correlated these results with the audio characteristics of the instruments and the performance. Due to the large number of participants, our findings validate and extend previous studies with a wider array of QoME variables and audio characteristics.**

I. Introduction

The *Quality of Musicians' Experience* (QoME) in *Network Music Performance* (NMP), that is, the performance of music over a network, depends on many variables, including audio delay and audio quality, as well as technical, environmental and psycho-acoustic factors. In practical networks, the limiting factor is usually delay; it is commonly accepted that one way delays of more than 25-30 ms prevent synchronization between musicians [1]. However, studies with actual musicians indicate that the impact of delay depends on the audio characteristics of the instruments used and the musical pieces performed [2].

Since the assessment of the impact of delay to NMP is based on either hand-clapping studies or small studies with actual musicians, we believe that it is worth revisiting the issue of how audio characteristics influence delay tolerance in a realistic setting. To this end, we ran a controlled experiment with a large number of musicians performing together over carefully controlled delays, using questionnaires to assess multiple aspects of QoME in a subjective manner. In this paper, in addition to correlating QoME with delay as in our previous work [3], we extract the audio features of the instruments, as in [2], and correlate them with a range of QoME variables and the performance tempo. Therefore, the contributions of this study to the literature include:

- A large set of participants, allowing the extraction of statistically significant results.
- An extended QoME questionnaire plus automated tempo measurement as the dependent variables.
- A wide range of instruments and musical genres with diverse audio characteristics as independent variables.
- An extremely accurate audio delay emulation environment, not subject to any jitter.

The outline of the rest of the paper is as follows. In Section II, we review work on assessing the effects of delay on QoME. In Section III we present the QoME variables that we measured. Section IV describes the setup of our experimental scenarios, while Section V presents the audio characteristics studied for their influence on delay perception. In Section VI we present and analyze our qualitative and quantitative results. We summarize our work in Section VII.

II. Related Work

An extended overview of past studies in NMP can be found in [4]. The simplest way to assess synchronization is to use hand claps, since they allow non-musicians to participate and their simple audio signature simplifies the analysis of tempo variations. Schuett [1] investigated the effect of delay in tempo and proposed the *Ensemble Performance Threshold (*EPT) as the delay below which clapping performers can synchronize. Two performers participated in five scenarios with different starting tempos and delays. The authors found that as the delay grew beyond 30 ms, the tempo began to slow down.

Chafe [5] used 17 pairs (34 performers) in clapping sessions under variable time delays. For delays below 11.5 ms, 74% of the performances sped up. At delays over 14 ms, 85% slowed down. No correlation with the starting tempo was found. Even though a metronome was used at the start, the initial tempos varied noticeably.

Farner [6] asked 11 musically experienced pairs (22 subjects) to clap together for at least seven measures of a simple complementary rhythmic pattern. The tempo was found to decrease more rapidly for higher delays, and the relation was approximately linear. In addition, the tempo tended to increase for the shortest delays, but the actual value of this zero crossing, around 15 ms, had a large uncertainty and was sensitive to the way the tempo change was calculated.

Chafe [7] examined performances by 24 pairs (48 performers) clapping, with the delays between pairs set electronically in the range of 3-78 ms. With extremely short delays, clappers had a tendency to accelerate from anticipation. Synchronization lagged at longer delays, producing an increasingly severe deceleration and then deterioration of the performed rhythms.

Based on hand clapping tests then, the EPT seems to be in the 25-30 ms range, with higher delays resulting in the tempo slowing down. However, a real assessment of QoME requires musicians using real instruments and performing real musical pieces. Musical instruments, as opposed to hand claps, have diverse audio characteristics, while actual musical performances can employ different musical styles, with the performers taking different roles.

Barbosa [8] investigated how the *Perceptual Attack Time* (PAT) of different instruments affected tempo. He used two musicians performing cello and violin and analyzed the recordings at delays between 0 and 90 ms, with a starting metronome set to 80 BPM (Beats Per Minute). He conducted one experiment with slow attacks and one with sharp attacks. The tempo was found to be higher with sharp attacks, while it decreased with delay in both cases.

Chew [9] asked two pianists to perform Pulenc's sonata for two pianos. This sonata has three movements played at 46, 132 and 160 BPM. The audio delay used was in the range 0–150 ms. The musicians had visual contact, but only heard each other via headphones. The authors reported that in the fast tempo movement, ensemble issues appear at delays of 50 ms. In the slow tempo movement, issues arose at delays of 75 ms. Finally, in the very fast movement, musicians had issues even at delays of 10 ms.

Carôt [10] asked 5 professional drummers to perform with the same professional bass player. The one way delay was in the range of 0–70 ms. The experiments were performed at speeds of 60, 100, 120 and 160 BPM and the delay between the two players was increased in steps of 5 ms until one of the players felt uncomfortable or slowing down occurred. The author reported that acceptable delays were in the 5–65 ms range, depending on the musicians, also noting that at higher BPM the musicians found it harder to synchronize, something also noted in [9].

Olmos [11] experimented with six singers, one conductor and one pianist. The authors reported that for the most part, the singers managed to cope with the delays. An important observation is that a big factor was familiarity with the technology. As the experiments proceeded, the singers were less distracted by the mediated experience and showed a greater sense of emotional connection, regardless of latency. Interestingly, the authors reported that the average tempo actually increased as latency was increased. This is likely due to the mediating role of the conductor, who, in effect, established the tempo.

Rottondi [2] employed 8 musicians with at least 8 years of musical experience in the experiments. The authors reported that the interaction quality was strongly correlated to perceived delay. The perceived delay was strongly affected by the timbral and rhythmic characteristics of the combination of instruments and musical parts. Among the timbral features, instruments with a higher spectral entropy and spectral flatness (such as guitars and drums) led to a larger tempo slowdown in case of higher network delays. In addition, they amplified the negative impacts of delay. Large delays (above 75 ms) did not prevent musical interaction, but they limited the selection of the instrument/part combinations for which interaction was feasible.

Delle Monache [12] asked 10 musicians (5 duets) to participate. Each duet performed under six different delays in the 28-134 ms range, using screens for visual interaction. They were asked to fill a small 5-item questionnaire after each repetition, and a larger 27-item questionnaire at the end of each session. Further comments were collected at the end of the test. Their answers show a negative effect of delay to the musicians involvement. Interestingly, the visual quality did not seem to distract from the performance, while the audio quality did. The authors believe that this is due to the poor quality of visual immersion, which made it difficult for the performers to rely on vision. Indeed, all the pairs reported that the frontal placement of the screens resulted in a less natural interaction, since they normally use peripheral vision to monitor the other performers.

III. Evaluation Variables

Considering the diverse results reported in previous work, rather than asking the participants to rate a session with a single *Mean Opinion Score* (MOS), we created a questionnaire covering different aspects of perception. We applied our original questionnaire in a pilot study with 4 pairs of participants [13]. Based on these results, we refined and extended the questionnaire for the larger scale study reported in this paper. The questionnaire was designed so that it could be easily answered after each individual NMP session. Each musician simply had to choose a score by touching a "button" on a tablet. We elaborate upon the questions relevant to this paper in the rest of this section.

Evaluate your satisfaction. The *Perception of Satisfaction* (PoSat) is the MOS metric used to evaluate QoME in most studies. It is evaluated in a scale from 1 (not satisfied at all) to 5 (very satisfied). As satisfaction is a very complex phenomenon, in this study we complement it with other subjective variables.

Evaluate the degree of delay you perceived. Even though in our experiments we controlled the delay ourselves, it is important to understand how musicians perceive delay during their performance. The *Perception of Audio Delay* (PoAD) variable shows how the participants grade it in a scale from 1 (no delay) to 5 (too much delay).

Evaluate the degree of synchronization. Achieving synchronization is a critical issue in NMP, with past work indicating its strong dependence on delay. The *Perception of Synchronization Degree* (PoSD) was graded in a scale from 1 (cannot synchronize at all) to 5 (can fully synchronize).

To what degree did you follow your partner? The *I was Trying to Follow my Partner* (TTF) variable examines if a musician tries to follow her partners' tempo. Previous work indicates that as musicians find it harder to synchronize, they try to follow their partner, establishing a master-slave relationship [14]. This variable is assessed in a scale from 1 (not at all) to 5 (I followed a lot).

Did you focus on audio or visual contact? As mentioned in [2] the use of visual contact is an aspect that needs to be examined in NMP, as musicians often use visual cues for synchronization. We used an ultra-low delay camera/monitor setup and asked the musicians whether they mostly focused on audio or video contact.

Did you feel anxiety? and *Did you feel irritation?* As many musicians are not keen with technology, there is a possibility that anxiety and irritation may emerge during NMP sessions. Anxiety may be a result of unfamiliarity with the equipment used for NMP, while high delay or poor quality may irritate the musicians. We investigate the existence of these phenomena using again a 5 point scale.

Tempo variation. In addition to the subjective evaluation, we calculated the average tempo of each performance, by applying the MIRtoolbox to the audio recordings of each session (see also Section V), in order to assess how it varies with different delay settings and audio characteristics.

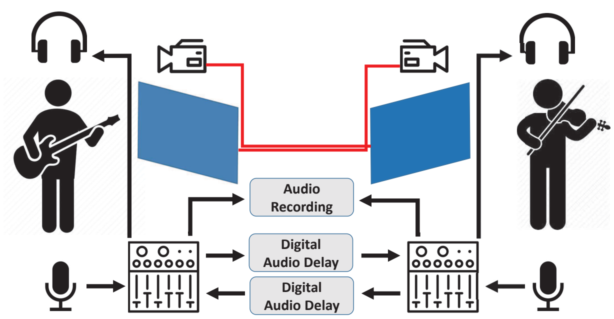

Fig. 1. Experimental topology.

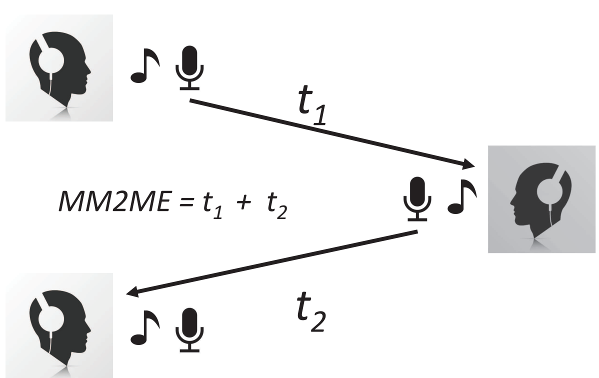

Fig. 2. My Mouth to My Ear delay.

IV. Experimental Setup

As shown in Figure 1, the two musicians were located in two visually and aurally isolated rooms in the same floor of our University. For audio, we used a mixing console with an auxiliary output, a condenser microphone and closed type headphones at each endpoint. The mixing consoles were connected to each other via the existing network cabling; the cables were patched to allow analog audio to pass through, without an intervening switch. Delay was manipulated via AD-340 digital delays by Audio Research in each direction of the audio path. We did *not* use computers to capture and playback the audio signal at each endpoint, so as to avoid the unpredictable delays introduced by sampling and packetization. Hence, the physical delay over the direct cable connection was fixed to a sub-ms level.

For video, we used a camera in each room that sent analog video (composite) via the existing cabling (again, directly patched) to a monitor in the other room. We did *not* add delays to the video and experimentally measured the one way camera to monitor delay to be around 15 ms.

Unlike most NMP studies which use *Mouth to Ear* (M2E) delay, which is the end-to-end delay between the microphone at one end and the speaker at the other end, in our work we use the *My Mouth to My Ear* (MM2ME) delay. As shown in Figure 2, MM2ME is the two-way counterpart to M2E, over which it has two advantages. From a perceptual viewpoint, when musicians play together, each musician plays one note and expects to listen to the other musicians' note to play his next one, and so on. From a technical viewpoint, measuring MM2ME delay accurately is much easier than measuring the M2E delay, as it can be done at one endpoint, by simply reflecting the transmitted sound at the other endpoint; in contrast, M2E needs to be measured at both endpoints, thus requiring perfectly synchronized clocks [15].

In our experimental scenario, each duet performed a 60 sec musical part of their choice, repeating it ten (10) times, with a different delay for each repetition. After each repetition, each musician was asked to answer the electronic questionnaire. The MM2ME delays used were in the range of 0-120 ms (equivalent to 0-60 ms M2E), in the order 10, 25, 35, 30, 20, 0, 40, 60, 80 and 120 ms. The delays were set in this random order and the participants were not informed about the purpose of the experiment, so as to prevent any bias in the results. No metronome was used, letting the musicians synchronize by themselves.

We kept the musical parts short, to avoid tiring the musicians, as they had to repeat their performance; we wanted to prevent boredom from affecting their QoME evaluation. We found that even with this short duration, musicians would slow down their tempo as delay increased, or even stop the performance when the delay was too high, so we have a measure of confidence that the short duration of the performances does not hide the effects of delay.

We conducted experimental sessions with 22 individual musicians (11 duets). The musicians performed with a variety of instruments, including piano, acoustic guitar, electric guitar, electric bass, violin and flute, as well as traditional instruments including the lute, toumberleki, santouri and oud, in a variety of musical genres. Table I shows the musical genre and instruments for each duet.

V. Audio Characteristics

Our study attempts to assess the effects of delay on musical performances, in conjunction with a set of audio and musical features. Audio feature analysis is used in many studies [2], [16], [17], [18], [19], where the spectral characteristics of each instrument are checked against the musicians' tolerance to delay. As described in [2], the *Spectral Centroid* (SC) captures

TABLE I
MUSICAL GENRES AND INSTRUMENTS PLAYED BY EACH DUET AND THEIR AUDIO FEATURES.

	Duet 1	Duet 2	Duet 3	Duet 4	Duet 5	Duet 6	Duet 7	Duet 8	Duet 9	Duet 10	Duet 11
Genre	Folk	Folk	Rock	Rock	Funk	Funk	Rock	Rock	Classic	Folk	Folk
Instr1	Piano	Piano	El Gtr	El bass	Organ	El bass	El bass	El Gtr	Flute	Ac Gtr	Lute
SC	5642	6082	7469	1669	1606	6041	3459	5598	6971	3569	6249
SSp	6106	6452	7252	4442.3	2676	6952	6122	4120	6518	3884	4566
SSk	1.26	1.15	0.6	3.24	3.76	0.92	1.72	0.3	0.8	1.03	1.32
SK	3.44	3.01	1.89	12.89	20.3	2.29	4.58	1.83	2.32	2.77	4.31
SF	0.47	0.44	0.61	0.13	0.10	0.48	0.29	0.12	0.56	0.05	0.46
SE	0.93	0.92	0.94	0.74	0.83	0.9	0.81	0.93	0.94	0.89	0.94
Instr2	Santuri	Oud	El Gtr	El Gtr	El Gtr	Perc	Ac Gtr	Violin	Violin	Bouz	Violin
SC	2384	1382	1010	1346	1456	1959	2807	2762	2185.9	2750	3386
SSp	2505	2587	2868	3215	2357	2901	4410	2782	2322	3394	4771
SSk	2.28	3.88	4.69	3.98	5.65	3.05	2.18	2.73	3.5	2.19	1.82
SK	11.07	21.03	25.9	19.6	40.82	14.34	7.35	13.3	21.05	8.68	5.56
SF	0.09	0.09	0.08	0.11	0.06	0.13	0.26	0.11	0.08	0.17	0.31
SE	0.85	0.8	0.75	0.79	0.82	0.86	0.86	0.85	0.81	0.87	0.75

TABLE II
CLASSIFICATION RANGES FOR THE AUDIO FEATURES.

Audio Feature	S Entropy	S Flatness	S Kurtosis	S Skewness	S Centroid	S Spread
Low Range (1)	0.740-0.810	0.05-0.24	1.80-14.9	0.30-2.09	1010-3163	2322-3993
Middle Range (2)	0.811-0.879	0.25-0.42	15.0-28.0	2.10-3.87	3164-5317	3994-5665
High Range (3)	0.880-0.950	0.43-0.61	28.1-41.0	3.88-5.66	5318-7470	5666-7338

the brightness of the sound, while the *Spectral Spread* (SSp) measures the noisiness of a sound source. The *Spectral Skewness* (SSk) captures the symmetry of its frequency distribution and the *Spectral Kurtosis* (SK) describes the size of the tails of that distribution. The *Spectral Entropy* (SE) is a measure of the flatness of the spectrum, while the *Spectral Flatness* (SF) estimates the similarity of the source to a flat shape; these also reflect the noisiness of an audio source.

We used the MIRToolbox [20] to extract these features for each instrument, based on the NMP session recordings. Since each session was repeated 10 times, we calculated the average value for each metric across them. As shown in Table I, even these averaged values can vary depending on the performance; for example, there were two sessions with electric piano, each showing slightly different audio features. The same phenomenon occurred with the multiple electric guitar sessions. In practice, the way of playing each instruments influences its timbral metrics: arpeggios, solos and rhythm playing lead to different scores.

In order to assess the correlation between the spectral characteristics and the responses to the QoME questionnaires we used the methodology proposed in [2], which we explain below. First, after calculating the average metrics for each of the 22 performances, we found the minimum and maximum value across all performances. The range for each metric was then divided in three equal parts, which we labeled as low, middle and high. For example, SE ranged from 0.740 to 0.950, so we divided this range to 0.740–0.810 (low), 0.811–0.879 (middle) and 0.880–0.950 (high). Then, we assigned each performance to one of these ranges and correlated each range and delay value with the evaluation variables. Table II shows the ranges for each metric.[1]

In addition to the timbral features of the instruments, we also took into account two musical features. First, we considered whether each musician performed the rhythm or the solo part of the duet, since previous work indicates that rhythmic parts are more sensitive to delay variations. Second, we considered the musical genre of each performance which reflects the rhythmic structure of each musical piece.

TABLE III
ANOVA ANALYSIS: DELAY VS. QOME VARIABLES.

Dependent Var	PoSD	PoAD	PoSat	TTF
Independent Var	delay	delay	delay	delay
p value	0.001	0.013	0.819	0.002

VI. EVALUATION RESULTS

Throughout the experiments, the participants did not report anxiety or irritation: almost all gave an answer of 1 (no anxiety / no irritation) to the corresponding questions, regardless of delay. This indicates that the participants felt comfortable and did not find the experience frustrating. Furthermore, the musicians indicated a strong preference to audio contact, compared to video, with answers ranging from 4 (mostly audio) to 5 (only audio) to the relevant question, despite the fact that video delay was only 15 ms, while audio delay increased up to 120 ms. This indicates that musicians mostly used aural cues for synchronization.

For the remaining subjective QoME variables, in order to determine whether there is a statistically significant correlation

[1]The entire dataset, that is, the QoME scores, the audio feature metrics and the measured tempo for each performance, is available at `https://github.com/mmlab-aueb/nmp/` .

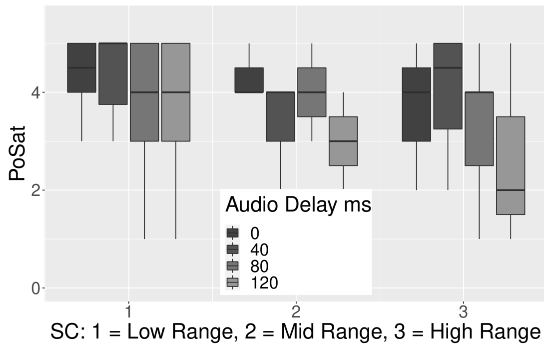

Fig. 3. PoSat against delay and Spectral Centroid (SC).

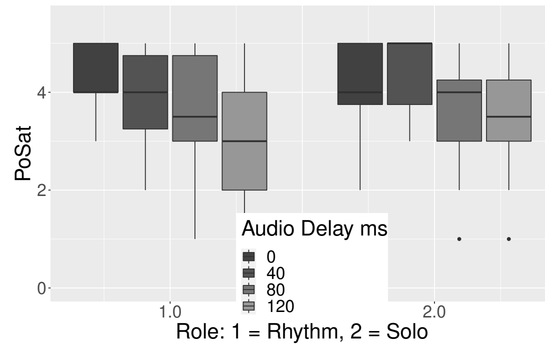

Fig. 6. PoSat against delay and Rhythm or Solo.

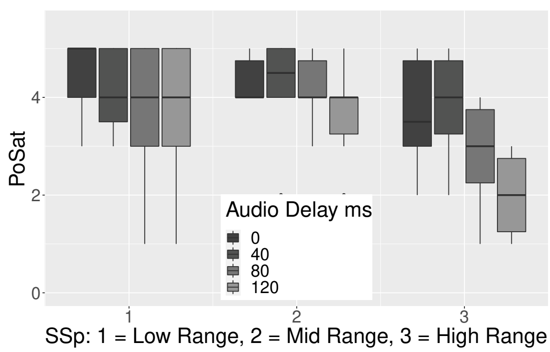

Fig. 4. PoSat against delay and Spectral Spread (SSp).

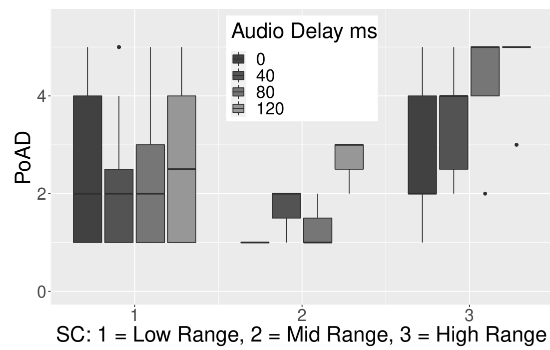

Fig. 7. PoAD against delay and Spectral Centroid (SC).

with delay, we performed an ANOVA analysis for repeated measures with delay as the independent variable and *Perception of Satisfaction* (PoSat), *Perception of Synchronization Degree* (PoSD), *Perception of Delay* (PoAD) and *I was Trying to Follow my partner* (TTF) as the dependent variable. Table III shows the results for the entire set of 22 participants. Most of the p values are lower than 0.05, indicating a strong probability of correlation with delay. The only exception is PoSat; this may be due to the fact that PoSat was quite high, even for the highest delay tested. A more detailed analysis of the QoME variables against delay can be found in our previous paper [3].

In this paper, we expand upon this analysis by looking at how each subjective variable (PoSat, PoAD, PoSD and TTF) was influenced not only by delay, but also by the audio features of the instruments and the musical features of the performances; we perform the same analysis for an objective variable, the observed tempo. To this end, we plotted one boxplot per characteristic, showing one set of boxes for each range of that characteristic (low, middle and high) and one box for each of the main MM2ME delays (0, 40, 80 and 120 ms). The boxes include a black horizontal line corresponding to the median value, with the entire box corresponding to 50% of the values (from the 25th to the 75th percentile). The whiskers (vertical lines above and below the box) show the minimum and maximum values excluding outliers: if we define IQR, the inter-quartile range, to be the difference between the 75th and the 25th percentile (that is, the height of the box), any value more than 1.5 times the IQR away from the box edges, is an outlier, shown as a dot in the plot.

Although the PoSat variable did not have a statistically significant correlation to delay, interesting observations arise when we look at the audio characteristics. PoSat was mostly affected by delay for instruments in the middle and high range of the Spectral Centroid (Figure 3), meaning instruments with brighter sounds, and in the high range of the Spectral Spread (Figure 4) and in the middle and high range of the Spectral Flatness (Figure 5), meaning noisier instruments. In addition, PoSat was more influenced by delay for rhythm rather than for solo performances (Figure 6).

Regarding PoAD, its dependence on the actual delay was higher for instruments in the middle and high ranges of the Spectral Centroid (Figure 7), while the Spectral Skewness was a factor for instruments in the low and high ranges (Figure 8) and Spectral Spread was a factor for instruments in the high range only (Figure 9). Finally, PoAD (like PoSat)

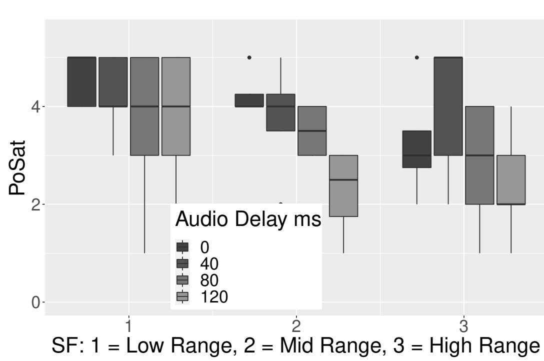

Fig. 5. PoSat against delay and Spectral Flatness (SF).

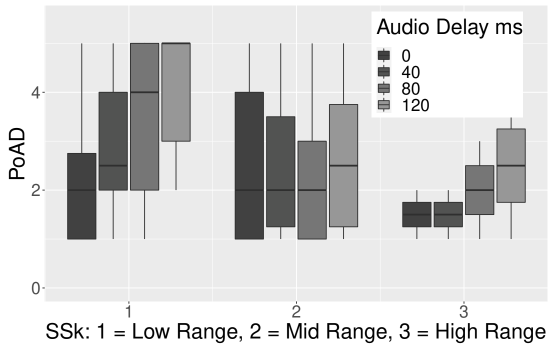

Fig. 8. PoAD against delay and Spectral Skewness (SSk).

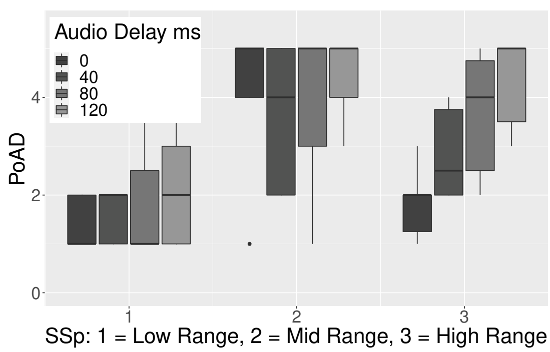

Fig. 9. PoAD against delay and Spectral Spread (SSp).

was more influenced by delay for rhythm rather than for solo performances (Figure 10).

PoSD was one of the most crucial metrics in our experiments. It was mostly influenced by delay for instruments in the high range of the Spectral Centroid (Figure 11), with very similar results for the Spectral Spread and Spectral Flatness (not shown). It was also influenced by instruments in the middle and high ranges of Spectral Entropy (Figure 12). Again, it was more influenced by delay for rhythm rather than for solo parts (not shown).

The TTF metric was influenced by nearly all the audio and musical features. Regarding audio features, it was mostly affected by delay for instruments in the low and high ranges of the Spectral Centroid (Figure 13) and Spectral Spread (not shown), in the high range of Spectral Skewness (Figure 14)

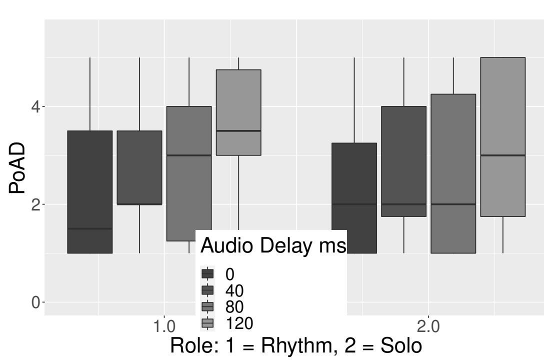

Fig. 10. PoAD against delay and Rhythm or Solo.

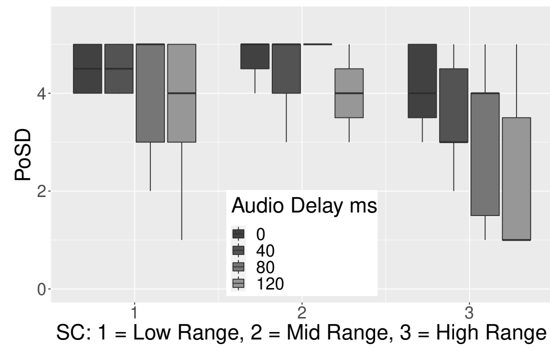

Fig. 11. PoSD against delay and Spectral Centroid (SC).

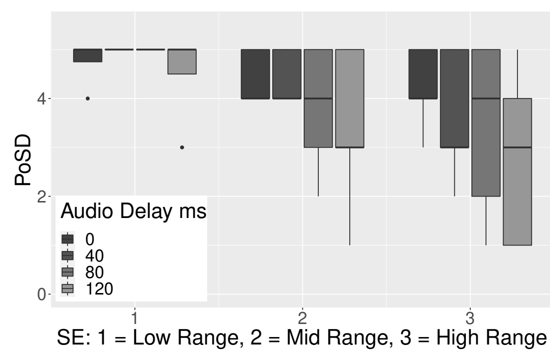

Fig. 12. PoSD against delay and Spectral Entropy (SE).

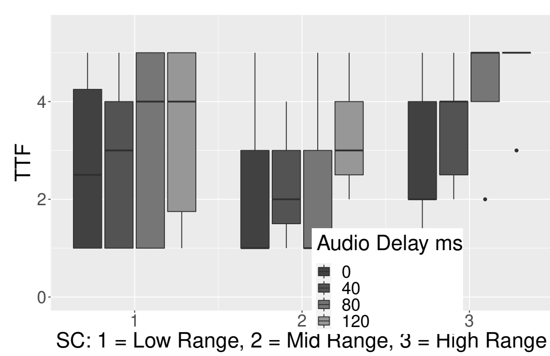

Fig. 13. TTF against delay and Spectral Centroid (SC).

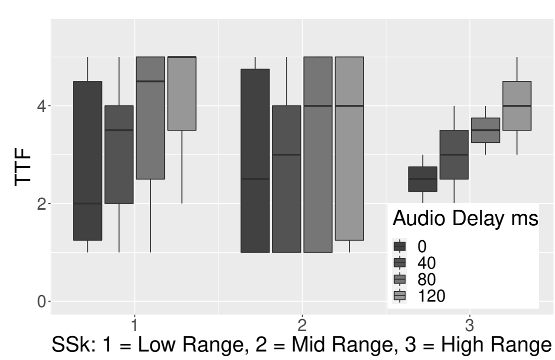

Fig. 14. TTF against delay and Spectral Skewness (SSk).

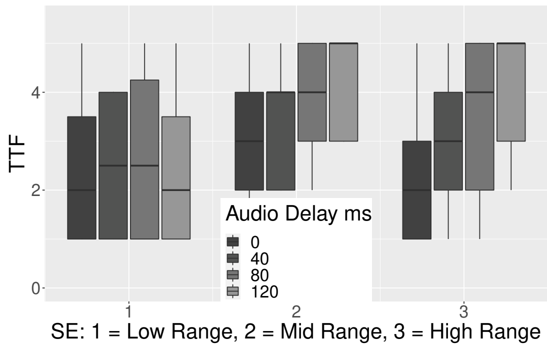

Fig. 15. TTF against delay and Spectral Entropy (SE).

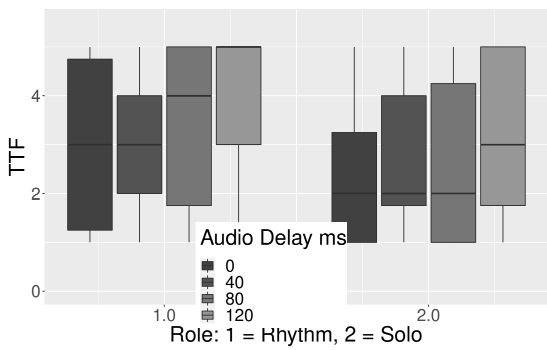

Fig. 16. TTF against delay and Rhythm or Solo.

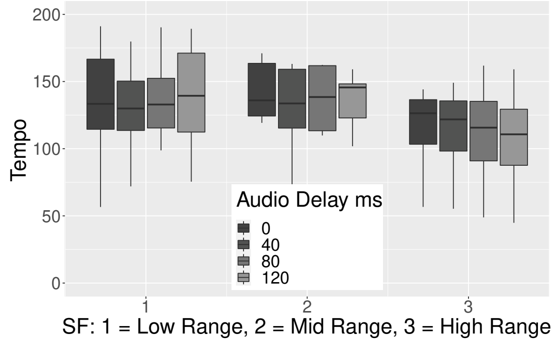

Fig. 18. Tempo against delay and Spectral Flatness (SF).

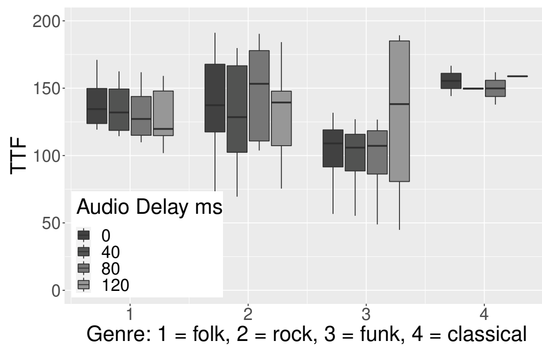

Fig. 19. Tempo against delay and Music Genre.

and in the middle and high ranges of Spectral Entropy (Figure 15) and Spectral Flatness (not shown). It was again more influenced by instruments playing rhythm parts (Figure 16), as was PoSat, PoAD and PoSD. Finally, it was more affected by delay with folk and rock performances, and less by funk and classical (Figure 17); it should be noted that the folk pieces performed were all intended for dancing, hence highly rhythmic.

Finally, the only objective variable assessed, tempo, was found to be mostly influenced by delay for instruments in the high range of Spectral Flatness (Figure 18), while amongst the four musical genres performed, only folk music had an influence on the tempo (Figure 19), with the other genres being either insensitive (classical) or showing no clear correlation pattern (rock and funk).

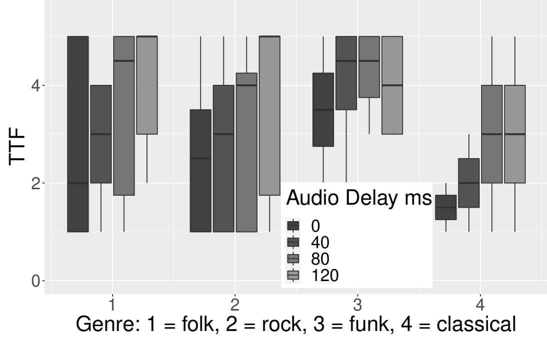

Fig. 17. TTF against delay and Music Genre.

To summarize, all the QoME variables (PoSat, PoAD, PSD and TTF) were more affected by delay with brighter and noisier instruments, performers that had a rhythm role and musical pieces with a more rhythmic structure. On the other hand, the effects of the audio and musical features on the performance tempo are not that clear. While our results are in line with previous studies of real musical performances (for example, [2] and [12]), our study shows that delay is more detrimental to rhythmic performances *in general*, not just on performances with faster initial tempos.

VII. Summary

We conducted a set of NMP experiments, where the delay between 11 pairs of musicians was varied in a controlled manner; the musicians answered an extensive QoME assessment questionnaire at the end of each session. After the sessions ended, we analyzed the audio features of the instruments performed, the musical features of the performance and the tempo variations observed. The sessions included a wide range of instruments and musical genres.

We studied the correlation between delay and four variables characterizing the perception of an NMP session (perception of satisfaction, audio delay, synchronization and following the partner) plus the performance tempo. We also studied whether the effects of delay to those variables were dependent on the audio features of the instruments, the role of the musician and the genre of the piece performed. We found a statistically significant correlation of most subjective variables with delay, and also clear indications that delay influences perception to a higher degree for brighter and noisier instruments, rhythm rather than solo parts and highly rhythmic musical pieces.

Our results lend additional credibility to previous results due to the large sample size of our experiments, as well as a clearer picture of QoME due to the consideration of multiple variables and the use of both audio and musical features of the instruments used and the pieces performed.

Acknowledgments

We would like to thank the participating musicians for their patience, the students who helped with the experiments, and Dr. Ioannis Doumanis (University of Central Lancashire) for his help with the experimental design.

References

[1] N. Schuett, C. Chafe, *The Effects of Latency on Ensemble Performance*, Master's thesis, CCRMA Department of Music, Stanford University (2002 May).

[2] C. Rottondi, M. Buccoli, M. Zanoni, D. Garao, G. Verticale, A. Sarti, "Feature-Based Analysis of the Effects of Packet Delay on Networked Musical Interactions," *Journal of the AES*, vol. 63, pp. 864–875 (2015 Nov), [Online]. Available: 10.17743/jaes.2015.0074.

[3] K. Tsioutas, G. Xylomenos, I. Doumanis, "An Empirical Evaluation of QoME for NMP," presented at the *IFIP International Conference on New Technologies, Mobility and Security (NTMS)* (2021 Apr), [Online]. Available: 10.1109/NTMS49979.2021.9432657.

[4] C. Rottondi, C. Chafe, C. Allocchio, A. Sarti, "An Overview on Networked Music Performance Technologies," *IEEE Access*, vol. 4, pp. 8823–8843 (2016 Dec), [Online]. Available: 10.1109/ACCESS.2016.2628440.

[5] C. Chafe, M. Gurevich, "Network Time Delay and Ensemble Accuracy: Effects of Latency, Asymmetry," presented at the *AES Convention 117* (2004 Oct).

[6] S. Farner, A. Solvang, A. Sæbø, U. P. Svensson, "Ensemble Hand-Clapping Experiments under the Influence of Delay and Various Acoustic Environments," *Journal of the AES*, vol. 57, pp. 1028–1041 (2009 Dec).

[7] C. Chafe, J.-P. Cáceres, M. Gurevich, "Effect of Temporal Separation on Synchronization in Rhythmic Performance," *Perception*, vol. 39, pp. 982–92 (2010 Jan), [Online]. Available: 10.1068/p6465.

[8] Á. Barbosa, J. Cordeiro, "The Influence of Perceptual Attack Times in Networked Music Performance," presented at the *AES Conference* (2011 Nov).

[9] E. Chew, A. Sawchuk, C. Tanoue, R. Zimmermann, "Segmental Tempo Analysis of Performances in User-centered Experiments in the Distributed Immersive Performance Project," presented at the *Sound Music Computer Conference* (2005 Nov).

[10] A. Carôt, C. Werner, T. Fischinger, "Towards a Comprehensive Cognitive Analysis of Delay-Influenced Rhythmical Interaction," presented at the *International Computer Music Conference* (2009 Aug).

[11] A. Olmos, M. Brulé, N. Bouillot, M. Benovoy, J. Blum, H. Sun, N. W. Lund, J. R. Cooperstock, "Exploring the Role of Latency and Orchestra Placement on the Networked Performance of a Distributed Opera," presented at the *Annual International Workshop on Presence* (2009).

[12] S. Delle Monache, M. Buccoli, L. Comanducci, A. Sarti, G. Cospito, E. Pietrocola, F. Berbenni, "Time is not on my Side: Network Latency, Presence and Performance in Remote Music Interaction," presented at the *Colloquium on Music Informatics-Machine Sounds, Sound Machines* (2018 Nov).

[13] K. Tsioutas, G. Xylomenos, I. Doumanis, C. Angelou, "Quality of Musicians' Experience in Network Music Performance: A Subjective Evaluation," presented at the *AES Convention 148* (2020 May).

[14] A. Carôt, C. Werner, "Network Music Performance-Problems, Approaches and Perspectives," presented at the *Music in the Global Village Conference* (2007 Sep).

[15] A. Carôt, C. Hoene, H. Busse, C. Kuhr, "Results of The Fast-Music Project: Five Contributions to The Domain of Distributed Music," *IEEE Access* (2020 Mar), [Online]. Available: 10.1109/ACCESS.2020.2979362.

[16] G. Tzanetakis, P. Cook, "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, pp. 293–302 (2002 Nov), [Online]. Available: 10.1109/TSA.2002.800560.

[17] K. Jensen, T. H. Andersen, "Real-Time Beat Estimation Using Feature Extraction," presented at the *International Symposium on Computer Music Modeling and Retrieval*, pp. 13–22 (2003), [Online]. Available: 10.1007/978-3-540-39900-1_2.

[18] H. Misra, S. Ikbal, H. Bourlard, H. Hermansky, "Spectral Entropy Based Feature for Robust ASR," presented at the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (2004), [Online]. Available: 10.1109/ICASSP.2004.1325955.

[19] K. Drossos, R. Kotsakis, G. Kalliris, A. Floros, "Sound Events and Emotions: Investigating the Relation of Rhythmic Characteristics and Arousal," presented at the *International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 1–6 (2013), [Online]. Available: 10.1109/IISA.2013.6623709.

[20] O. Lartillot, P. Toiviainen, "MIR in Matlab (II): A Toolbox for Musical Feature Extraction from Audio," presented at the *International Conference on Music Information Retrieval (ISMIR)*, pp. 287–288 (2007).