An empirical evaluation of QoME for NMP

Konstantinos Tsioutas*, George Xylomenos* and Ioannis Doumanis†
*Athens University of Economics and Business, Department of Informatics, Greece
†University of Central Lancashire, School of Physical Science and Computing, United Kingdom

***Abstract*—For Network Music Performance (NMP), low end-to-end delay is probably the most critical factor affecting the Quality of Musicians' Experience (QoME), as inflated delays inhibit the synchronization between musicians. Even though customized NMP tools can greatly reduce application-layer delays, generally available Internet connections impose considerable network-layer delays to NMP sessions in wide-area settings. In order to empirically assess the delay tolerance of NMP, multiple pairs of musicians were asked to perform in a controlled network environment, while we varied the end-to-end delay. The musicians were then asked to evaluate a number of metrics. The statistical analysis of the results indicates that the tolerance of actual musicians to delay is higher than previously thought, thus expanding the applicability of NMP to longer distances.**

***Index Terms*—QoME, NMP, delay.**

I. Introduction

The *Quality of Musicians' Experience* (QoME) in *Network Music Performance* (NMP), that is, the performance of music when musicians are connected over a network, is a complex function which depends on many variables, including audio delay and audio quality, as well as technical, environmental and psycho-acoustic factors [1]. As in other applications that involve live communication, NMP has strict delay requirements. Indeed, it is commonly accepted that for NMP delays of more than 30 ms prevent synchronization between musicians [2]; in contrast, video conferencing can tolerate delays of up to 100 ms. Such delays are hard to achieve with commonly available residential Internet connections.

Although this would seem to make NMP an academic curiosity, many musicians have found that when using low delay NMP tools, they can perform satisfactorily over moderate distances. For this reason, we believe it is important to revisit the issue of how much delay is acceptable for NMP under realistic circumstances. To this end, we designed a controlled experiment with eleven pairs of actual musicians performing over carefully controlled delays, using questionnaires to assess QoME in a subjective manner. Each pair of musicians performed a different musical piece, making our study the largest and most diverse to date. After verifying the statistical validity of the results, we found that the delay tolerance of actual musicians in NMP is closer to 40 ms.

The outline of the rest of the paper is as follows. In Section II, we briefly present related work on assessing the effects of delay on musicians' experience. In Section III we present the variables affecting QoME that we measured. Section IV describes the setup of our experimental scenarios, while in Section V we present and analyze the questionnaire results. We summarize our work in Section VI.

II. Related Work

A large amount of research touches upon QoME evaluation for NMP, looking at it from different perspectives. Many studies employ subjective strategies where musicians respond to surveys evaluating their experience while audio delay is manipulated. Schuett [2] investigated the effects of delay to synchronization, proposing the term *Ensemble Performance Threshold* (EPT) for the one way delay below which synchronization is possible, reporting that it lies between 20 and 30 ms. Chafe [3] reached the same conclusion, experimenting with musicians who clapped their hands, indicating that when the latency was below 11 ms, the tempo was actually accelerating. Driessen [4] experimented with two musicians who performed a clapping session without any external tempo reference. The authors reported that the tempo of two musicians slowed down as the delay is increased. Farner [5] asked eleven pairs of musically experienced subjects to clap together, reporting that the tempo was found to decrease more rapidly for higher delays, and the relation was approximately linear. Gurevich [6] used seventeen pairs of subjects in clapping sessions under varied time delays. Authors reported that for delays shorter than 11.5 ms, 74% of the performances sped up. At delays of 14 ms and above, 85% slowed down.

Barbosa [7] asked four musicians to play bass, percussion, piano and guitar. The authors found that regardless of the instrumental skills or the music instrument, all musicians were able to tolerate more delay for slower tempos. Barbosa [8], investigated how the attack period of notes affects the tempo, using two musicians performing cello and violin. The analysis of the recordings showed that the tempo was generally higher with fast attack times than with slow attack times. In both cases, it decreased with increasing latency.

Olmos [9] worked with six singers and a conductor over a network. For the most part, the singers that participated in the experiment managed to cope with the various delays. The singers mentioned that to a certain extent, they were able to establish emotional connections with each other.

Bartlette [10] asked two pairs of musicians to performed two Mozart duets. Although the musicians chose different strategies to handle the latency, both duets were strongly affected by latency at and above 100 ms. At these levels, the musicians rated the performances as neither musical nor interactive, and they reported that they played as individuals and listened less and less to one another.

Rottondi [11] asked seven pairs of musicians to participate in the experiments. Each repetition was characterized by different settings in terms of reference tempo, network latency, and jitter. The authors reported that the perceived delay was strongly affected by the timbral and rhythmic characteristics of the combination of instruments and parts. They reported that the noisiness of the instrument has an impact on the perceived delay. They concluded that the possibility of enjoying an NMP session is not only a function of delay, but also of the role and the timbral characteristics of the involved musical instruments, as well as the rhythmic complexity of the performance.

Carot [12], asked five professional drummers to perform with five professional bass players, reporting that the overall delay thresholds ranges between a minimal delay of 5 ms and a maximal delay of 65 ms. He also noted that the players did not exhibit a common latency acceptance value.

Finally, Monache [13], asked ten volunteers to participate in pairs, performing mandolin, accordion, guitar, percussion, harp, flute, alto sax. Delay had a negative effect to the musicians involvement. The author also reported that a general distress was caused by latency and a willingness to find ways to cope with it emerged from the answers.

To summarize, the hand clapping studies indicate that synchronization is hard when delay exceeds 30 ms. However, studies with real musical performances show quite diverse results, indicating that in real NMP sessions musicians adapt to higher delays, often by slowing down their tempo, depending on the type of music performed and the instruments used.

III. Subjective Evaluation

In this study, we focus on the human perception of delay in an NMP setting. 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 four pairs of participants [14]. Based on these results, we refined and extended the questionnaire for the larger study reported in this paper, which involves eleven pairs of participants. The questionnaire was designed so that it could be easily answered after each individual NMP session. Each musician simply had to choose a score in a Likert scale for each question by touching a “button” on a tablet. We elaborate upon the questions in the rest of this section.

Evaluate the audio quality. Although in this study we do *not* vary the audio quality across experiments, in our pilot study (where we *did* test different sampling rates), we found that participants often confused higher delay with lower audio quality [1]. For the *Perception of Audio Quality* (PoAQ) variable, musicians were asked to evaluate in a 1 to 5 Likert scale the perceived audio quality (higher is better).

Evaluate the degree of synchronization. Achieving synchronization is a critical issue in NMP, with past work indicating its strong dependent on delay. For the *Perception of Synchronization Degree* (PoSD) variable, musicians provided responses in a Likert Scale from 1 (cannot synchronize at all) to 5 (can fully synchronize).

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 asks the participants to grade the delay in a Likert scale from 1 (no delay) to 5 (too much delay).

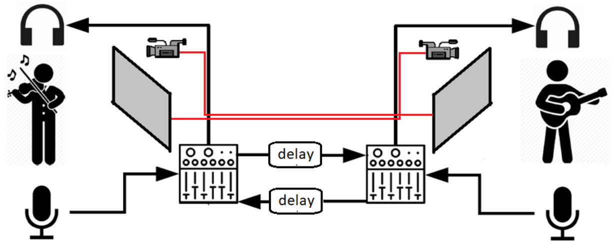

Fig. 1. Experimental setup.

Evaluate your level of musical and emotional expression. Musicians’ experience in NMP is strongly correlated to their musical and emotional expression during the performance. Thus we propose *Perception of Musical and Emotional Expression* (PoMEE), a variable evaluated with a 5 point Likert Scale (higher is better).

Evaluate your satisfaction. The *Perception of Satisfaction* (PoSat) is basically the MOS metric which is widely used to evaluate Quality of Experience in similar studies. As satisfaction is a very complex phenomenon, in this study we complement this metric with many other subjective variables.

Evaluate how did you perform and *Evaluate how your partner performed.* As proposed in [1], the QoME of each musician is strongly correlated to the QoME of the other. The *Perception of My Performance* (PoMP) and *Perception of My Partner’s Performance* (PoMPP) variables evaluate in a scale from -3 (very bad) to +3 (very good) how each musician assesses his own and his partner’s performance.

To what degree did you follow your partner? With the *I was Trying to Follow Partner* (TTF) variable, we examine whether a musician tries to follow her partners tempo or not. Our previous work has indicated that as musicians find it harder to synchronize, they try to follow their partner. This variable is assessed in a scale from 1 (not at all) to 5 (I followed a lot).

Did you focus on the audio or the visual contact? As mentioned in [11] 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 audio delay or poor audio quality may irritate the musicians. We investigate the existence of these phenomena using a 5 point Likert scale.

IV. Experimental Setup

As shown in Figure 1, the topology used is a peer to peer architecture. The endpoints were two visually and aurally isolated rooms in the same floor of the main AUEB building. For audio, we used an eight channel mixing console with

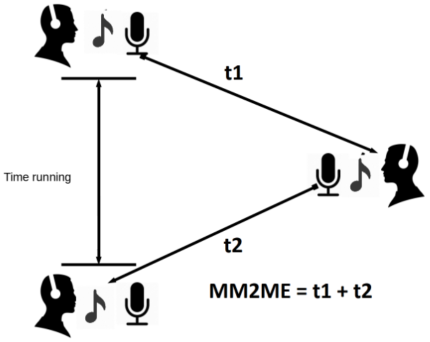

Fig. 2. My Mouth to My Ear delay.

an auxiliary output, a condenser microphone and closed type headphones for each of the two musicians participating. The two mixing consoles were connected to each other via the existing network cabling in the building; the cables were patched together to allow analog audio to pass through. 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 and video 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 HD camera in each room that sent *uncompressed video* via the existing cabling (without any intervening switches) to a HD monitor in the other room. We did *not* add any delays, but we estimated the one way video delay to be around 10 ms, due to the in-camera capture and processing delay.

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 proposed in [15]. 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 unconsciously 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 [16].

In our experimental scenario, each pair of musicians played a one minute musical part of their choice, repeating it ten (10) times, with a different delay setting for each repetition. After each repetition, each musician was asked to answer an electronic questionnaire using a tablet. As shown in Table I, the delay values were set in a random order for each repetition. In addition, the participants were not informed about the delay variations, or about the purpose of the experiment. The goal was to conduct an experiment that would allow us to evaluate the effect of delay on the various QoME variables without bias or noise in the answers. No metronome was used, leaving the musicians to synchronize by themselves.

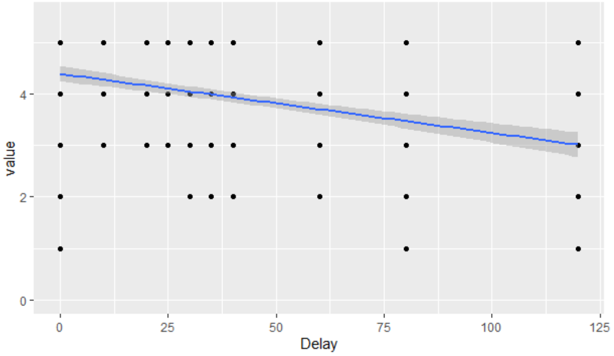

Fig. 3. Perception of Synchronization Degree (N = 22).

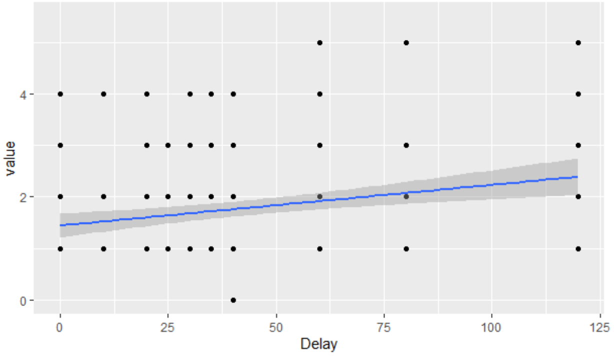

Fig. 4. Perception of Audio Delay (N=22).

V. Evaluation Results

We conducted experimental sessions with 22 musicians (11 pairs). 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, shown in Table II.

We present below graphs with the results for some of our questions, with points representing individual answers for each MM2ME delay value, and a line showing a linear model computed by the R statistic package; all graphs were created using RStudio. Unless otherwise indicated, the graphs reflect results from the 22 participants mentioned above.

Figure 3 shows the results for the *Perception of Synchronization Degree* (PoSD) depending on the MM2ME delay. As shown in the graph, the fitted line has a clear negative rate as delay increases, starting at 4.5 and ending at 3; it seems that musicians perceived that could not synchronize with delays of more than 80 ms, rather than the 60 ms reported in the literature (twice the EPT of 30 ms).

Figure 4 shows the results for the *Perception of Audio Delay* (PoAD) variable. A slightly increasing slope is observed, with scores ranging from 1.5 to 2.5. However, when we focus only on the sessions where piano was used (3 pairs), the responses from the pianists and their partners (6 participants), shown in Figure 5, indicate a higher rate of growth, from 1.5 to 3.5. This observation is evidence that the characteristics of the instruments used affect the perception of delay; specifically, piano performances seem to be more sensitive to delay.

Repetition	1	2	3	4	5	6	7	8	9	10
MM2ME delay (ms)	10	25	35	30	20	0	40	60	80	120

TABLE I
MM2ME DELAYS FOR EACH REPETITION.

Duet 1	Duet 2	Duet 3	Duet 4	Duet 5	Duet 6	Duet 7	Duet 8	Duet 9	Duet 10	Duet 11
Piano	Piano	Elec Gtr	Elec bass	Piano	Elec bass	Elec bass	Elec Gtr	Flute	Bouzouki	Lute
Santouri	Oud	Elec Gtr	Elec Gtr	Elec Gtr	Percussion	Acous Gtr	Violin	Violin	Bouzouki	Violin

TABLE II
INSTRUMENTS PLAYED BY THE MUSICIANS.

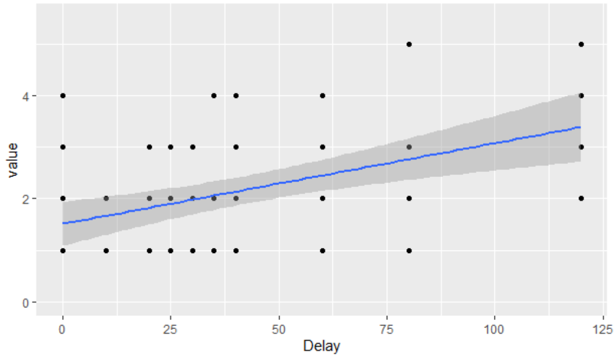

Fig. 5. Perception of Audio Delay (Pianists and partners, N=6).

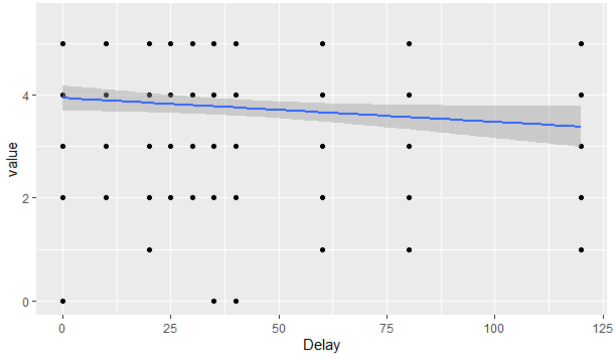

Fig. 6. Perception of Satisfaction (N=22).

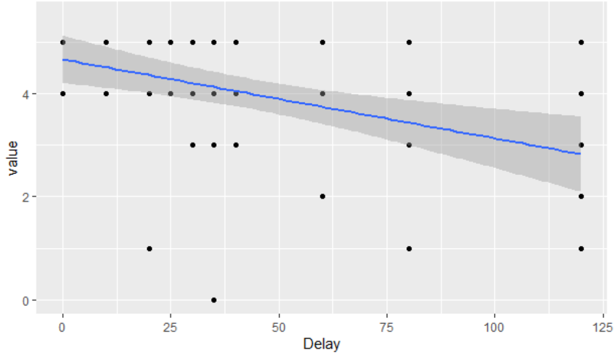

Fig. 7. Perception of Satisfaction (Pianists and Partners, N=6).

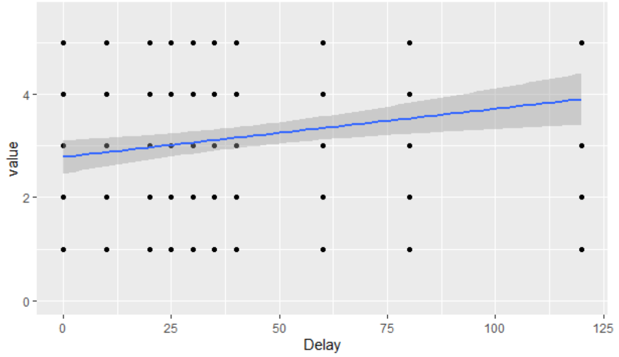

Fig. 8. I was Trying to Follow my partner (N=22).

One of the most critical variables in our experiment is the *Perception of Satisfaction* (PoSat), which is basically a MOS. As shown in Figure 6, the fitted line has a small negative slope. When we focus again on pianists and their partners, Figure 7 shows that they were more influenced by increasing delay, as the slope of the line is steeper. Note also that while with all musicians the answers were widely spread for all delay values, pianists and their partners gave a more narrow range of answers at lower delay values, which grows as delay increases.

In Figure 8 we can see the results for the *I was Trying to Follow my partner* (TTF) question. The fitted line has the same positive rate as the perceived audio delay in Figure 4. For delays of up to 60 ms we have very similar (and wide) result ranges, indicating that the musicians are split between leading and following, as would be expected in a non NMP scenario. With the highest delay of 120 ms though, the fitted line tends to reach 4, showing that synchronization is harder and each musician tries to follow the other.

Anxiety and irritation (not shown) were not observed: almost all participants gave an answer of 1 (no anxiety / no irritation), regardless of delay. This indicates that the participants felt comfortable and did not find the NMP experience frustrating. This allows us to place more trust on the other QoME variables, as our previous work has indicated that when participants are uncomfortable with their NMP experience, they tend to provide lower scores to unrelated questions.

Regarding the emphasis on audio or visual contact, the results (not shown) indicated a strong preference to audio contact, with answers ranging from 4 (mostly audio) to 5 (only audio). This was despite the fact that video delay was fixed to a low 10 ms, while the audio delay increased up to 120 ms.

Dependent Variable	PoSD	PoAD	PoSat	TTF
Independent Variable	delay	delay	delay	delay
Sample Size	22	22	22	22
p = 0.05	0.001	0.013	0.819	0.002

TABLE III
ANOVA ANALYSIS: DELAY VS. QOME VARIABLES.

This indicates that musicians are mostly based on aural and not visual cues for synchronization.

To test whether our results have statistical significance, we performed ANOVA for repeated measures with delay as the independent variable and *Perception of Synchronization Degree* (PoSD), *Perception of Delay* (PoAD), *Perception of Satisfaction* (PoSat) and *I was Trying to Follow my partner* (TTF) as the dependent variable; we did not test the *Did you focus on the audio or the visual contact*, *Did you feel anxiety* and *Did you feel Irritation* variables, as it was clear that they were not influenced by delay. 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 the PoSat variable, which may be due to the fact that the Perception of Satisfaction was quite high even for the highest delay tested (see Figure 6).

VI. SUMMARY

We conducted a set of NMP experiments, where the delay between a pair of musicians was varied in a controlled manner for each session, with the musicians answering a QoME assessment questionnaire at the end of each session. The results indicate that even though increasing delay does have an effect on the QoME of the participating musicians, the range of acceptable delays is larger than previously reported based on hand clapping experiments. The musicians participating in our study considered the performances to be synchronized and satisfactory with one way (M2E) delays of up to 40 ms (or two way, MM2ME, delays of up to 80 ms), with many duos satisfied even with M2E delays of up to 60 ms (MM2ME of up to 120 ms). This means that the acceptable EPT is closer to 40 ms over a wide range of instruments and musical pieces, rather than the 20-30 ms widely cited in the literature.

ACKNOWLEDGMENT

We would like to thank all the participating musicians for their patience during the experiments, as well as the fellows who helped with setting up and carrying out the experiments. This research was supported by the Research Center of the Athens University of Economics and Business.

REFERENCES

[1] K. Tsioutas, I. Doumanis, and G. Xylomenos, "A framework for understanding and defining quality of musicians' experience in network music performance environments," in *Audio Engineering Society Convention 146*, March 2019.

[2] N. Schuett, "The effects of latency on ensemble performance," Bachelor Thesis, CCRMA Department of Music, Stanford University, 2002.

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

[4] P. F. Driessen, T. E. Darcie, and B. Pillay, "The effects of network delay on tempo in musical performance," *Computer Music Journal*, vol. 35, no. 1, pp. 76–89, March 2011.

[5] S. Farner, A. Solvang, A. Sæbø, and U. P. Svensson, "Ensemble handclapping experiments under the influence of delay and various acoustic environments," *Joirnal of the Audio Engineering Society*, vol. 57, no. 12, pp. 1028–1041, 2009.

[6] M. Gurevich, C. Chafe, G. Leslie, and S. Tyan, "Simulation of networked ensemble performance with varying time delays: Characterization of ensemble accuracy," in *International Computer Music Conference*, November 2004.

[7] A. Barbosa, J. Cardoso, and G. Geiger, "Network latency adaptive tempo in the public sound objects system," in *International Conference on New Interfaces for Musical Expression*, January 2005, pp. 184–187.

[8] A. Barbosa and J. Cordeiro, "The influence of perceptual attack times in networked music performance," *44th AES International Conference*, November 2011.

[9] A. Olmos, M. Brulé, N. Bouillot, M. Benovoy, J. Blum, H. Sun, N. W. Lund, and J. R. Cooperstock, "Exploring the role of latency and orchestra placement on the networked performance of a distributed opera," in *12th International Workshop on Presence*, 2009, pp. 1–9.

[10] 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.

[11] C. Rottondi, M. Buccoli, M. Zanoni, D. Garao, G. Verticale, and A. Sarti, "Feature-based analysis of the effects of packet delay on networked musical interactions," *Journal of the Audio Engineering Society*, vol. 63, pp. 864–875, November 2015.

[12] A. Carôt, C. Werner, and T. Fischinger, "Towards a comprehensive cognitive analysis of delay-influenced rhythmical interaction," in *International Computer Music Conference*, 2009.

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

[14] K. Tsioutas, G. Xylomenos, I. Doumanis, and C. Angelou, "Quality of musicians' experience in network music performance: A subjective evaluation," in *Audio Engineering Society Convention 148*, May 2020.

[15] K. Tsioutas, G. Xylomenos, and I. Doumanis, "Aretousa: A competitive audio streaming software for network music performance," in *Audio Engineering Society Convention 146*, March 2019.

[16] 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, March 2020.