Visual Contact in Network Music Performance

Konstantinos Tsioutas*, George Xylomenos* and Chrisoula Alexandraki†
* Department of Informatics, Athens University of Economics and Business, Greece
†Department of Music Technology and Acoustics, Hellenic Mediterranean University, Greece

***Abstract*—In Network Music Performance (NMP), where musicians perform together over the Internet, visual communication is often absent, due to bandwidth limitations and concerns about latency. In this study, we investigate the impact of visual contact to NMP by experimenting with 18 musicians who performed as music duos under variable audio delays and with or without visual contact. After each performance the musicians rated their experience using a questionnaire; the audio recordings of the performances were also subjected to tempo analysis. The results indicate that visual contact exposes disrupting effects in musical synchronization, based on both the subjective and objective analyses, having a negative impact on the overall experience.**

***Index Terms*—Network Music Performance, Flow Short Scale, Visual Contact.**

I. Introduction

Visual Contact (VC) is a core component of video conferencing applications. To accommodate bandwidth constraints across diverse networking environments, commercial video conferencing platforms heavily compress video, leading to considerable coding/decoding latencies. While such latencies are acceptable for most communication scenarios, in *Network Music Performance* (NMP), where musicians perform together over the Internet, high latency prevents synchronization. This fact motivates investigations concerning the importance of visual contact during NMP sessions.

While certain research initiatives show that visual contact is beneficial for performers in the same room [1], other works indicate that musicians rely primarily on auditory information [2], [3] and musical expectation [4], rather than on visual cues. Our own NMP experiments which combined audio and video connectivity also indicated that the participants primarily relied on audio for synchronization [5].

In this study we investigate the effects of visual contact to NMP by observing NMP sessions with real performers, using the same audio setup with and without visual contact. We combine subjective analysis, by adapting the *Flow Short Scale* (FSS) [6] questionnaire, with objective analysis, by studying the tempo of the audio recordings. Our results indicate that visual contact in our setup detracts from musical experience, disrupting musician co-ordination.

The rest of this paper is structured as follows. In Section II we describe our experimental setup. In Section III we present the questionnaire used in the study. In Section IV we analyze the results from the subjective ratings and the tempo analysis. We conclude and discuss future work in Section V.

II. Experimental Design

A. Experimental Scenarios

Musicians were grouped into nine duos, each completing a three-stage experiment. In the first stage, all duos performed a music piece of their own choice in the same room, with the musicians facing each other; we refer to this setting as *Real Presence* (RP) to differentiate it from NMP. In the next two stages, the musicians of each duo were located in two separate rooms, communicating via an NMP testbed, with or without *Visual Contact* (VC). The first five duos were requested to perform firstly without visual contact and then with visual contact; this was the *No VC/VC* group. The remaining four duos performed first with visual contact and then without visual contact; this was the *VC/No VC* group. This was an intentional design choice to investigate whether the order of using VC had any effect on the performances.

In each of the NMP stages, the musicians performed the same piece seven (7) times, under varying audio delays. After each repetition, they had to answer a questionnaire (see Section III) on a laptop placed next to them. All repetitions of each scenario were recorded for post-experiment analysis.

B. Experimental Testbed

We set up our testbed using two rooms of the Department of Music Technology and Acoustics of the Hellenic Mediterranean University. Each room was equipped with Ubuntu laptops, mixing consoles, dynamic microphones, closed type headphones, a CCTV camera and a 32” monitor. The two rooms were connected via Ethernet cabling and a Gigabit Ethernet switch; all machines were in the same LAN.

For the audio channel we used Sonobus[1] in conjunction with the JackTrip audio server[2] to achieve the lowest possible end-to-end latency. At each endpoint the audio buffer size was set to 128 Bytes, and the audio format used was PCM 44.1 kHz, 16 bits per sample. We measured the end-to-end *Mouth to Ear* M2E latency to be 12 ms, using the reflected pulse method described in [7]; this was the lowest latency we could achieve in LAN conditions.

To manipulate audio delay for the purposes of the experiments, we used the `tc` Linux tool with the `netem` delay emulator. We applied seven (7) different values of additional delay, from 0 to 60 ms in 10 ms increments, ending up with M2E latencies of 12 to 72 ms. We used a seemingly

[1]https://www.sonobus.net
[2]https://ccrma.stanford.edu/software/jacktrip/

TABLE I: NMP Questionnaire and Original FSS Correspondence

Modified NMP FSS	**FSS English Version Items**
Q1. I lost the sense of time during this repetition	Q3. I didn't notice time passing.
Q2. I felt myself becoming completely immersed in the musical performance.	Q6. I am totally absorbed in what I am doing.
Q3. I felt a strong sense of connection and synchrony with my colleague.	New NMP-specific item.
Q4. I felt challenged to make an even greater effort in this repetition.	Q1. I feel just the right amount of challenge.
Q5. I felt a sense of effortless control over my musical contribution.	Q2. My thoughts/activities run fluidly and smoothly.
Q6. I felt a deep sense of satisfaction and musical fulfillment in this repetition.	New NMP-specific item
Q7. I felt energized and motivated while playing with my colleague.	New NMP-specific item
Q8. I felt that I had control over my musical expression and contribution.	Q9. I feel that I have everything under control.
Q9. I felt a sense of flow, where my actions and awareness merged seamlessly.	Q7. The right thoughts/movements occur of their own accord.
Q10. Overall, the repetition was enjoyable and deeply absorbing.	Q10. I am completely lost in thought.
Unused FSS item.	Q4. I have no difficulty concentrating.
Unused FSS item.	Q5. My mind is completely clear.
Unused FSS item.	Q8. I know what I have to do each step of the way.

random order of delays in each set of experiments, rather than increasing or decreasing delays, to avoid bias. Note that in our setup there was no jitter, i.e., latency fluctuations, as commonly observed for Internet connections.

For the NMP experiments with visual contact, we used a composite video connection between the CCTV cameras and the TV monitors, to avoid digital coding delays. We ran the signal over direct Ethernet cables using RCA to Ethernet adapters, without any intermediate switches. We measured the *Glass to Glass* (G2G) delay between the CCTV camera and the monitor to be 32 ms, using the reflected timer method described in [7]. This was fixed for all the NMP + VC experiments.

C. Experimental Procedure

Eighteen (18) musicians (9 duos), participated in the experiments. They were all males, aged from 20 to 30 years old. They played various instruments (including traditional instruments of the Cretan music) and chose their own musical piece to perform. For statistical analysis, we divided the 18 musicians in two groups. The five odd-numbered duos were the No VC/VC group, while the four even-numbered duos where the VC/No VC group.

Each duo was first in the same room. They were informed about the procedure and after feeling comfortable, they performed their chosen musical piece once, facing each other (RP mode). After the first repetition in the same room, they completed the questionnaire. Then, they were located in separate rooms (NMP mode), where they performed seven repetitions of the same piece without VC and another seven with VC under different audio delays (No VC/VC group), or first with VC and then without it (VC/No VC group), completing the questionnaire after each repetition.

III. The Questionnaire

In studies related to NMP, the experiences of the participating musicians are usually assessed in a subjective manner via a questionnaire. In our past work on NMP we used custom-made questionnaires focusing on the effects of audio delay or audio quality [5]. In another NMP study, the authors combined items from three different questionnaires to evaluate presence in NMP sessions [8].

In NMP experiments, musicians usually perform the same piece multiple times under varying conditions, repeatedly completing the same questionnaire. For our study, considering the fifteen repetitions per duo, we set an upper limit to ten questionnaire items. We based our questionnaire on the *Flow Short Scale* (FSS) questionnaire, a validated questionnaire that evaluates absorption, effortless performance, the sense of control, and intrinsic enjoyment during an activity.

The original FSS questionnaire is in German [6]; we used the English version [9] to create a Greek questionnaire, which combined FSS questions with NMP-specific questions. We first selected seven out of the ten core flow-related items of the FSS, which could be reused or adapted to NMP; we omitted Q4, Q5 and Q8 of the FSS, as they were less relevant to NMP, to make space for three NMP-specific items. We then adapted each of the seven remaining FSS items to make them more specific to NMP, for example, indicating that the response referred to the last performance repetition, or that control referred to musical expression.

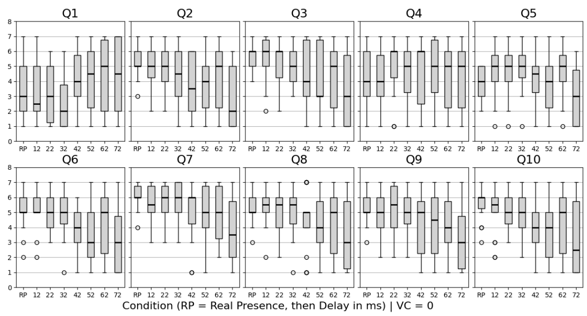
Finally, we added three new items to evaluate musical experience under varying conditions. First, we asked for connection and synchrony, two core concepts of collaborative musical performance (Q3). Second, we asked about satisfaction and musical fulfillment, an overall satisfaction measure not considered by the FSS (Q6). Third, we asked about energy and motivation during the performance (Q7).

The final questionnaire is shown in Table I, where the first column shows the actual items used (translated to English from Greek) and the right side the English versions of the corresponding FSS items; we point out the new NMP-specific items and the unused FSS items. The questionnaire was answered via a Google form and the participants rated each question using a 7 point Likert Scale (1 = totally disagree, 7 = totally agree).

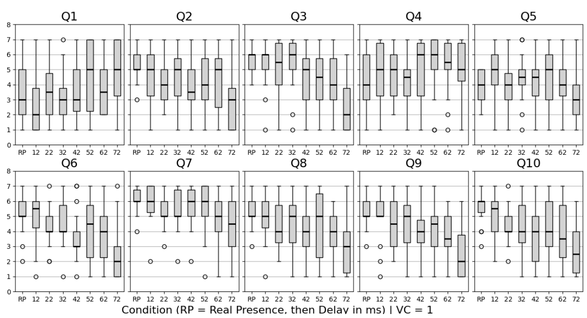
To validate the final questionnaire, we calculated Cronbach's Alpha coefficient for the entire questionnaire to be 0.849; Table II shows the Alpha coefficients for the individual items. The table indicates that Q1 and Q4 may be problematic; this is confirmed by the analysis in Section IV.

TABLE II: Validation of the Questionnaire

Item	Cronbach's Alpha	Item	Cronbach's Alpha
Q1	-0.1466	Q6	0.8172
Q2	0.8077	Q7	0.6514
Q3	0.7418	Q8	0.7529
Q4	0.0124	Q9	0.7958
Q5	0.5701	Q10	0.7997



(a) Responses for RP and NMP + No VC.



(b) Responses for RP and NMP + VC.

Fig. 1: Comparison of responses without and with VC.

IV. Subjective and Objective Analysis

We begin with the results of the subjective questionnaires. Figure 1(a) summarizes with boxplots the responses from all 9 duos to each item in the questionnaire; for each item, the x-axis starts with *Real Presence* (RP) as a baseline, and continues with increasing delays for NMP without Visual Contact as audio delays are increased. Figure 1(b) shows the same results but for the performances with NMP with Visual Contact (again, starting with RP as the baseline). For most items, we can see a negative trend with increasing delay, with the exception of Q1 and Q4, which were flagged as possibly problematic during the validation analysis.

For Q1, the goal was to evaluate whether participants experienced a loss of time perception during immersion. The median score for the RP condition is relatively low, equal to 3, and under NMP conditions it increases with delay. This indicates that musicians most likely interpreted it as an indication of delay. For Q4 there is no such clear trend, with the scores seemingly independent of delay.

We next performed a series of linear mixed-effects models separately for each item to examine the effects of delay (treated as a continuous variable in 10 ms increments), separately for the NMP experiments with no VC and the NMP experiments with VC. The results, in the form of delay-induced increments and p-values are shown in Table III.

TABLE III: Correlation of per-Item Responses to Delay

Item	NMP + No VC	NMP + VC
Q1	$0.19, p = 0.051$	$0.34, p = 0.3$
Q2	$-0.19, p = 0.018$	$-0.09, p = 0.39$
Q3	$-0.15, p = 0.10$	$-0.22, p = 0.58$
Q4	$0.014, p = 0.87$	$0.03, p = 0.89$
Q5	$-0.10, p = 0.19$	$-0.02, p = 0.46$
Q6	$-0.24, p = 0.004$	$-0.18, p = 0.6$
Q7	$-0.15, p = 0.02$	$-0.10, p = 0.59$
Q8	$-0.19, p = 0.01$	$-0.22, p = 0.77$
Q9	$-0.19, p = 0.01$	$-0.27, p = 0.48$
Q10	$-0.179, p = 0.03$	$-0.21, p = 0.73$

We can see again that Q1 was possibly misinterpreted, with higher delay having a positive effect, and Q4 being inconclusive with very small increments. For most other items though, there is a clear negative effect of delay on the scores, indicating that network delay broadly disrupts core aspects of musical performance and experience.

Comparing the results with and without visual contact, we do not see a general trend across items; in some items delay has a more pronounced effect with visual contact, while in others delay has a more pronounced effect without visual contact. However, the p-values are generally lower in the experiments without visual contact.

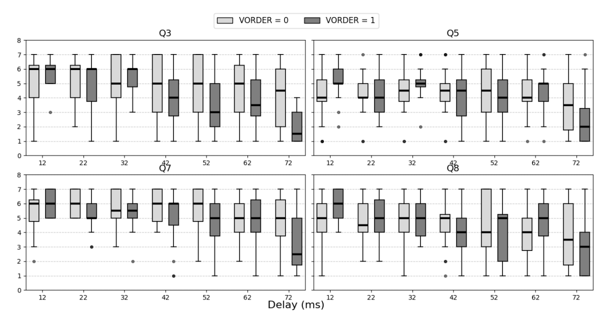
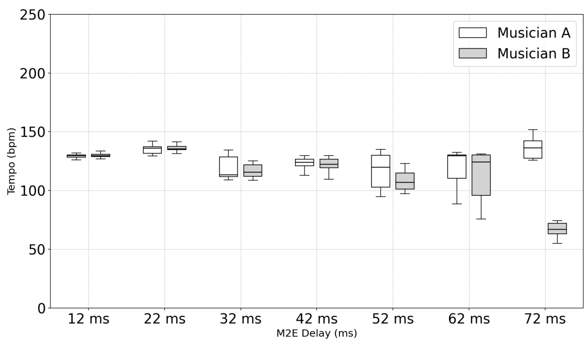


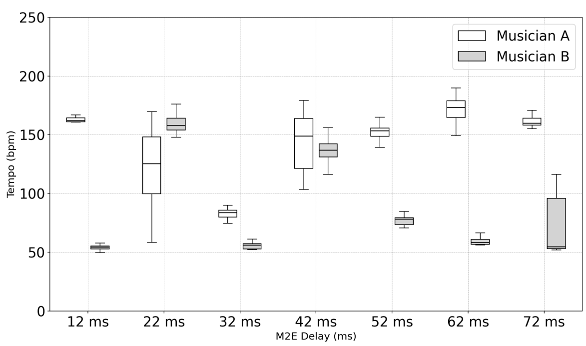
Fig. 2: Interactions between delay and order of visual contact to the ratings of items Q3, Q5, Q7 and Q8 ($p < 0.05$ for all four questions).

We then examine whether the order of running the NMP experiments affected the results. In Figure 2 we show side-by-side boxplots of the responses for the No VC/VC duos (light gray) and the VC/No VC duos (dark-gray for each delay value for items Q3 (connection and synchrony), Q5 (control), Q7 (energy and motivation) and Q8 (expressive control). These findings indicate that musicians who first performed with visual contact (dark gray) were slightly more sensitive to delay overall on these items. In contrast, no such interaction was observed for items Q2 (immersion), Q6 (satisfaction), Q9 (flow) and Q10 (enjoyment), suggesting that these aspects of experience are comparatively stable and less influenced by adaptation effects.

We also conducted a tempo analysis of the recordings of all duos. The tempo computation was performed using the MIRToolBox for Matlab [10]. In previous work, a similar analysis revealed that musicians in a duo tend to decrease their tempo to maintain synchronization as delay increases; beyond a delay threshold though, synchronization is lost and the tempo varies wildly between them [5]. In this study, we created boxplots of the tempo variation of each musician for each delay setting, using one figure for the performances with visual contact and one for the performances without it.



(a) Tempo variation against delay for duo 9 without visual contact.



(b) Tempo variation against delay for duo 9 with visual contact.

Fig. 3: Tempo variation for duo 9 without and with VC.

In Figure 3 we show the tempo variations for duo 9, without and with visual contact, respectively; this was a duo in the No VC/VC group. The x axis presents the M2E latency of each experiment. The white boxplots correspond to the tempo variations of musician A while the gray ones show the tempo variations of Musician B. Without visual contact (Figure 3(a)) the tempo variations between the two musicians are very close for low delay values, and then start to differ for higher delays of 52, 62 and 72 ms. On the other hand, with visual contact (Figure 3(b)), the tempo variations are clearly spread vertically. This strongly suggests that visual contact caused a deterioration in synchronization.

However, that was not always the case; in other duos the tempo was affected neither by visual contact nor by latency, suggesting that the participating musicians were more experienced and could compensate for higher delays. We did not perform any statistical analysis related to the tempo variations, which we plan to do as future work.

V. Conclusions and Future Work

We conducted a series of musical experiments with 18 musicians playing in duos, under real presence or networked conditions, with varying audio latencies and with or without visual contact, using a customized version of the FSS questionnaire and tempo analysis to evaluate the impact of visual contact in NMP.

The questionnaire responses indicated that higher audio latency negatively affected flow, immersion, fulfillment, motivation, control and satisfaction, verifying previous results. In addition, though, our questionnaires indicated that visual contact negatively affected musical fulfillment, motivation, expression, flow and satisfaction. The tempo analysis of the recordings also revealed a clear disruption of synchronization when visual contact was provided, for many duos.

Based on these subjective and objective analyses, we conclude that visual contact in our NMP setup had a negative effect on the overall musical experience. However, we must point out that the video channel had a non-negligible fixed latency (32 ms), which was not synchronized with audio latency. This mismatch between the two modalities may explain the negative impact of the video. We must also point out that the number of participants was quite low to draw general conclusions. In the future, we plan to perform tests with synchronized audio and video delays, to further assess whether visual contact has an effect on NMP.

References

[1] S. D'Amario, H. Daffern, and F. Bailes, "Synchronization in singing duo performances: The roles of visual contact and leadership instruction," *Frontiers in Psychology*, vol. Volume 9 - 2018, 2018.

[2] L. Bishop and W. Goebl, "When they listen and when they watch: Pianists' use of nonverbal audio and visual cues during duet performance," *Musicae Scientiae*, vol. 19, no. 1, pp. 84–110, 2015.

[3] S. Kawase, "Gazing behavior and coordination during piano duo performance," *Attention, perception & psychophysics*, vol. 76, 10 2013.

[4] P. Keller and M. Appel, "Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles," *Music Perception*, vol. 28, p. 27–46, 09 2010.

[5] K. Tsioutas, G. Xylomenos, and I. Doumanis, "Impact of audio delay and quality in network music performance," *Future Internet*, vol. 17, no. 8, 2025.

[6] F. Rheinberg, R. Vollmeyer, and S. Engeser, "Die Erfassung des Flow-Erlebens," in *Diagnostik von Selbstkonzept, Lernmotivation und Selbstregulation*, J. Stiensmeier-Pelster and F. Rheinberg, Eds. Hogrefe, 2003.

[7] K. Tsioutas, Y. Thomas, F. Bistas, I. Barous, G. Xylomenos, and G. C. Polyzos, "Network music performance beyond 4G," in *Proceedings of the International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2025.

[8] S. Delle Monache, L. Comanducci, M. Buccoli, M. Zanoni, A. Sarti, E. Pietrocola, F. Berbenni, and G. Cospito, "A presence- and performance-driven framework to investigate interactive networked music learning scenarios," *Wireless Communications and Mobile Computing*, vol. 2019, no. 1, p. 4593853, 2019.

[9] F. Rheinberg, R. Vollmeyer, S. Engster, and R. R. Sreeramoju, "FSS - flow short scale (english version)," 09 2023.

[10] O. Lartillot, P. Toiviainen, and T. Eerola, "A Matlab toolbox for music information retrieval," in *Data Analysis, Machine Learning and Applications*, C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker, Eds. Springer, 2008, pp. 261–268.