An Adaptive Metronome Technique for Mitigating the Impact of Latency in Networked Music Performances

Riccardo Battello*, Luca Comanducci †, Fabio Antonacci †, Augusto Sarti†
Politecnico di Milano
Milan, Italy
* riccardo.battello@mail.polimi.it,
†<name.surname@polimi.it>

Stefano Delle Monache, Giovanni Cospito, Enrico Pietrocola, Filippo Berbenni
Conservatory of Music G. Verdi of Milan
Milan, Italy,
<name.surname>@consmilano.it

***Abstract*—The improvement of digital communication networks increased the interest towards Networked Music Performance both for entertainment and education. One of the main issues that affects the effectiveness of remote performances is the inherent network latency in the communication chain. In this paper we propose a technique that, to the best of our knowledge, is the first work that explores the use of adaptive metronomes in order to track the tempo of the musicians and help them in coping with the latency. We test our proposal with preliminary experiments in the context of chamber music.**

I. Introduction

The increasing developments in network communication speeds have made the fruition and sharing of content on the internet a daily part of most people lives. These factors have led to the development of Networked Music Performance (NMP) as the field that gathers all the techniques that enable the process where geographically displaced musicians perform together as if they were in the same environment [1]. NMP finds its application both in entertainment scenarios such as geographically distributed concerts [2] and in educational ones such as blended and distance learning [3].

The EU funded INTERmusic(Interactive Environment for Music Learning and Practicing, 2017 - 2020) project aims at integrating the solutions developed in NMP in distance learning and education. The main aim of the project is to develop a series of best practices and guidelines that will enable the implementation of remote environments for music interaction and tuition, also in higher education institutions. This involves the development of three online pilot courses in music theory and composition, chamber music practice, vocal training. Certain parts of these courses can be based on the implementation of Massive Open Online Courses (MOOCs), where the interaction between teacher and students is one-to-many and usually asynchronous, posing no strict synchronization requirements [4]. However, the part of the courses that needs to access NMP tools such as distance music practice in the context of master-student lessons or rehearsals, exhibits stricter synchronicity requirements posed by the one-to-one (or few-to-few) communication. In this article we consider the NMP-based paradigm and we propose a technique aimed at helping the musicians in coping with the latency introduced by the network, which severely hinders the ability of the musicians to synchronize themselves.

Regarding this aspect, the research output of the INTERmusic project has already investigated the impact of latency by planning and performing a pilot experiment [5], where musicians had to play chamber music pieces remotely, while different levels of artificial latency were introduced in the communication chain that connected them. This experiment allowed us to obtain some insights about the effects observed and the techniques used by the musicians to cope with the latency. The observations drawn from the observed results enabled us to define a framework [6], [7] that models the main entities present in a NMP context, centered on the chamber music pedagogical scenario, but easily extendable to others. Following this framework we designed experiments that investigate NMP in pedagogical scenarios, such as in the case of the work presented in this paper.

A consistent body of literature already considers the impact of latency in NMP scenarios, finding as a general rule that when the latency values range between $20\ \mathrm{ms}$ and $60\ \mathrm{ms}$, the performance of the musicians degrades with increasing decelerations of the tempo kept by the performers [8]. Experiments in rhythmic hand-clapping have been performed in order to assess the impact of latency in the performance of two musicians playing over a network and positioned in two different rooms [9]–[13]. It was found that the best results can be achieved when the end-to-end delay between the musicians approximate a temporal separation corresponding to a real-life scenario. Findings also show that unnaturally low latencies (i.e. less than $10\ \mathrm{ms}$) could cause an acceleration in the performance of the musicians.

While several NMP dedicated softwares such as JackTrip [14], LOLA [15] and UltraGrid [16] have been developed, each of them does not solve all the issues posed by the INTERmusic project, as it aims at bringing low-latency solutions available also for setups where no high-end hardware is present.

A metronome is a device that produces a tick at a regular interval and is used during music rehearsal in order to help the musician in keeping the correct tempo while practicing. Even though chamber music performers usually avoid the use of a metronome during practice, it may be a helpful instrument in helping the musician keeping the tempo when facing network latency.

Also a metronome is a concept that could be very well inserted in the paradigm defined by the Internet of Musical Things [17]. For example we can imagine smart instruments [18] embedding a metronome in a way to provide an efficient way to face the network latency.

In [19] an approach that combines a user-adjustable distributed metronome is tested for different music genres and for end-to-end delays of several seconds. In this case the authors considered Delayed Feedback Compensation approach [20], which means that also an additional delay, regulated by the users, was inserted in the audio chain in order to contrast the latency.

Using a metronome, whose pulse was regulated by the Round Trip Time (RTT) of the underneath network connection, the authors of [21] organized a NMP between Stanford University, USA and Peking University, China.

A global metronome mechanism was proposed in [22]. This approach uses the Global Positioning System (GPS) to generate a globally synchronized audio signal that can be used as a means of a global metronome in order to regulate the tempo of the musicians [23]. Following the path of the aforementioned rhythmic hand-clapping experiments, the global metronome approach was tested in [24], which showed a stabilization of the tempo acceleration in the performance of the musicians.

In this paper we propose a latency-compensation method for NMP, which, to the best of our knowledge, is the first one that is based on adaptive metronomes that track and follow the tempo of the musicians by means of a beat tracker. Specifically, a musician hears the metronome that tracks the tempo of a second musician and helps him in synchronizing with him, following a master-slave [20] strategy. We test the method in the chamber music context, being this a preliminary experiment, we consider a simplified scenario where the latency is symmetrical and fixed, thus avoiding to consider the impact of jitter. The results, although limited in scope due to the number of participants, show that this method could be a viable technique for contrasting network latency in NMP scenarios.

II. Hypothesis and Proposed Methods

In this section we introduce the latency compensation technique based on the adoption of an adaptive metronome, whose tempo is regulated by a beat tracker. Specifically, we use the multipath beat tracker presented in [25], which has a low computational cost and can be easily inserted in an adaptive scenario. The method here presented is inspired by the master-slave latency compensation technique, where the musicians take two precise distinct roles, the *leader*, which dictates the tempo of the performance, and the *follower*, which tries to synchronize himself as much as possible with the former.

To combine the use of adaptive metronomes with the master-slave strategy, we devised a technique that performs the beat tracking on the leader and lets the follower hear the metronome adaptively based on the tempo of the former. Ideally, during the performance, the follower should try his best to comply with the tempo of the leader.

A scheme representing the architecture of the method can be seen in Fig. 1. The system can be seen as a double input-output structure, where the two main entities are the leader and the follower musicians. The inputs consist of the time-varying signals acquired from the musicians and are denoted $s_L^{in}(t)$ and $s_F^{in}(t)$, where t stands for the discrete time, L denotes the leader musician, while F the follower one. We denote the audio signals heard by the leader and follower musicians as $s_L^{out}(t)$ and $s_F^{out}(t)$, respectively. The inherent latency due to the network connection between the two musicians is modeled as an additive term d, which causes the signals to be delayed. The audio signal received by the leader

$$s_L^{out}(t) = s_F^{in}(t-d), \tag{1}$$

is defined as the delayed input of the follower.

On the side of the follower, $s_L^{in}(t)$ is used both for the auditory feedback and as the input of the adaptive metronome, which creates the signal $s_M(t)$. The output for the follower is defined as

$$s_F^{out}(t) = s_L^{in}(t-d) + s_M(t), \tag{2}$$

where $s_M(t)$ models the metronome signal, and can be defined as a pulse train, where each pulse corresponds to one beat. The time differences Δt_M between the pulses depends on the tempo value τ_M, expressed in BPM, of the adaptive metronome

$$\Delta t_M = \frac{60}{\tau_M}. \tag{3}$$

The adaptive metronome architecture, depicted in Fig. 2, consists of two main processes. The first one is the actual metronome, which generates the signal $s_M(t)$, depending on the chosen BPM. The second process is the beat tracking of the input signal, which modifies the BPM of the metronome, depending on the performance of the leader musician.

The beat tracker detects a new tempo τ_d every β beats, where β is a user-specified parameter. Each time a detection occurs, τ_d is compared to the target tempo τ_t in order to modify the adaptive metronome accordingly. Since abrupt tempo changes will result in an uncomfortable indication to the musicians [26], care should be used in specifying the β parameter, while too low values will not give the beat tracker sufficient time to detect the right tempo, values that are too high would sacrifice the adaptive characteristics of the metronome. The initial tempo of the metronome τ_d is

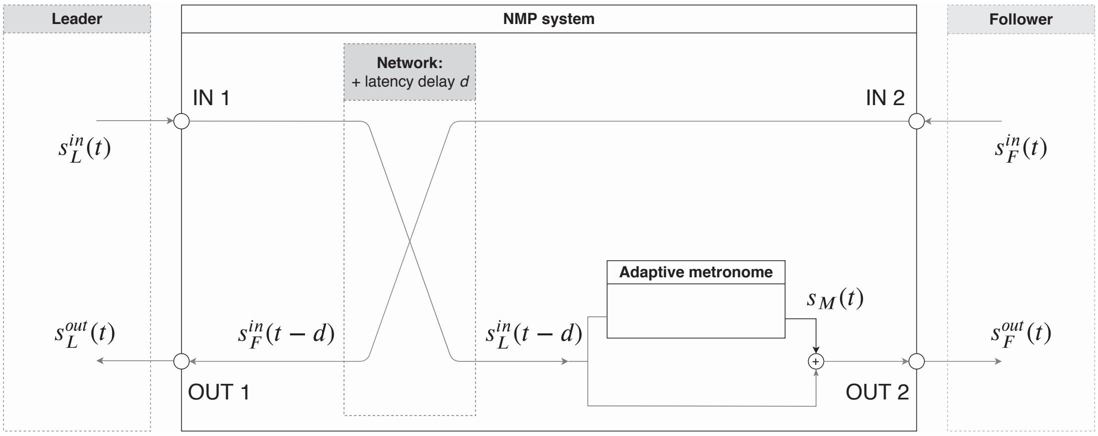


Fig. 1. Schematic representation of the adaptive metronome technique

initialized at the target one τ_t and is updated every β beats following the rule

$$\tau_M(\eta\beta) = \begin{cases} \tau_d & if \ |\tau_d - \tau_t| < \varepsilon\tau_t \\ \tau_M((\eta-1)\beta), & \text{otherwise} \end{cases}, \quad (4)$$

where ε defines the maximum valid difference between the detected tempo and the target tempo as a percentage value of τ_t, and η is a positive natural number. It is important to define this boundary not as a fixed value but dependent on the specific tempo, because a certain variation in BPM can be more or less perceivable at different values of reference. As stated in Eq. (4), if the difference between the target tempo and the detected one $|\tau_d - \tau_t|$ is smaller than the selected threshold $\varepsilon\tau_t$, then the metronome tempo is modified accordingly. Instead, following the idea of avoiding to present changes that are too abrupt to be useful for the musicians, if the detected tempo is too far from the target value, which could happen for an error of the beat tracking or the leader, the tempo of the metronome is not changed and keeps the value corresponding to the previous detection. After τ_d is updated, the pulse train is updated with the new frequency and the signal $s_M(t)$ is given as the output of the system.

III. Experimental Study

This section presents experimental results regarding the adoption of the adaptive metronome technique. The experiment was conducted during December 2019 at the Conservatorio di Musica “Giuseppe Verdi” in Milan. The aim of this test was to examine how a master/slave approach, where the slave musician is guided by an adaptive metronome, could improve the performances of the musicians when facing network latency. The results obtained during the musical interactions can provide a basic assessment of the method and a guide for potential improvements, but cannot be interpreted with a significant statistical value, because of the relatively small number of participants.

In Sec. III-A we present the setup used to perform the experiment, while in Sec. III-B the participants and the task that they performed.

A. Setup

The setup used for the experiment is depicted in Fig. 3. The musicians of each couple, were located in two acoustically isolated rooms, and interacted by means of microphones and loudspeakers for the auditory feedback and cameras and screens for the visual feedback. In order to improve the eye contact, which has shown to be beneficial in NMP interactions [27], the musicians were positioned in a frontal position, as shown in Fig. 4a and Fig. 4b. The equipment consisted of two Audio Technica ATM350 cardioid condenser clip-on microphones, two low latency Ximea MQ13CG-E2 USB 3.1 Gen 1 cameras, two Dynaudio BM5 mk3 7” studio monitor loudspeakers, two 27” 144 Hz Asus ROG video monitors. The acquisition and transmission devices were connected to two Windows computers, powered with Intel/Nvidia components and equipped with PCIe audio cards and Gigabit ethernet connection to a common LAN. The communication between the workstations was performed through LOLA [15], which is optimized both for low-latency audio and video transmission. A server was inserted in the communication chain and acted as a Network Emulator in order to add an arbitrary latency. Using the Linux command “tc” we inserted a two-way latency, while subtracting the processing latency in order to precisely obtain the desired delay. The processing latency was empirically measured and set at $22.5\mathrm{ms}$. Being this a preliminary experiment, focusing on understanding if musicians are able to use the metronome as a tool to contrast the latency, we reproduced a simplified version of a real NMP scenario. The experiment considered a fixed and symmetric latency, with no jitter. The signals produced by the two microphones were recorded for the analysis by a Digital Audio Workstation, from the perspective of room 1, where the follower musician was playing. As shown in Fig. 3 and Fig. 4b, the signal of the musician located in room 2, defined as the leader during the performance, was also processed by an additional computer in room 1, in order to perform the beat tracking operation. The generated adaptive metronome was heard by the musician in room 1 by means of open headphones, in order to allow him

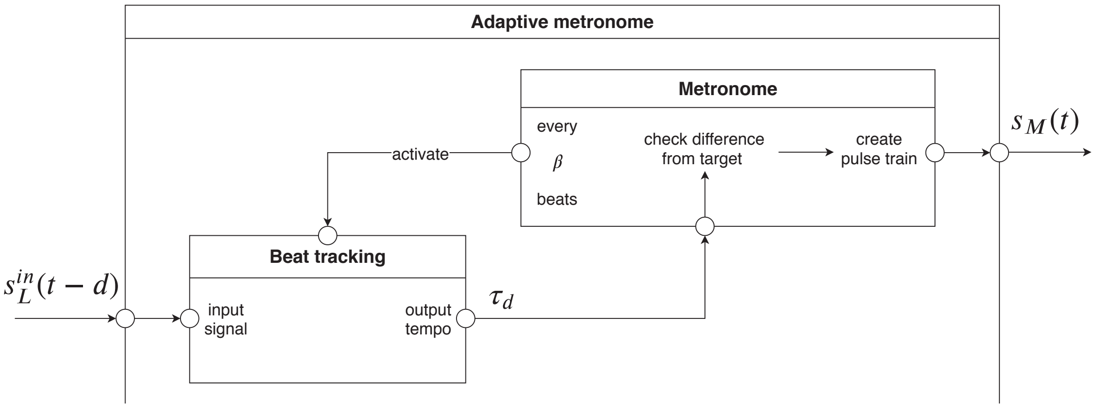


Fig. 2. Schemat

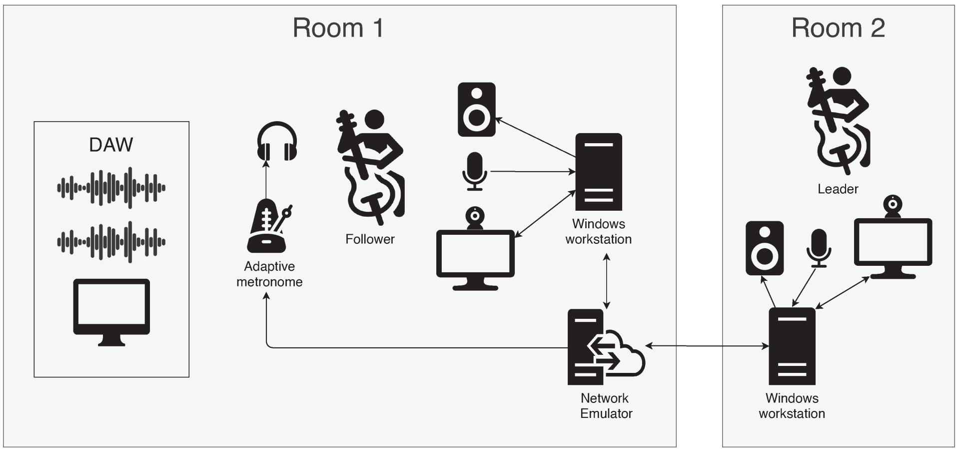


Fig. 3. Setup of the experiment

to also hear clearly the sound of the instruments.

B. Participants and tasks

Four couples of musicians have been involved in the experiment. All the participants were conservatory students, aged between 17 and 22 years, with at least 7 years of musical experience. In each couple, summarized in Table I, both members were playing the same instrument. Each couple played four repetitions of the same musical score at a target tempo $\tau_t = 104$ BPM. The stimuli proposed to the musicians consisted in a score composed by one of the authors, whose aim was to consider basic structures of synchronicity and interaction in a chamber music context. For a further analysis of the scores, we refer the reader to [7].

TABLE I. PARTECIPANTS TO THE EXPERIMENT

Couple A	Couple B	Couple C	Couple D
Saxophone	Guitar	Oboe	Oboe

TABLE II. EXPERIMENT CONDITIONS

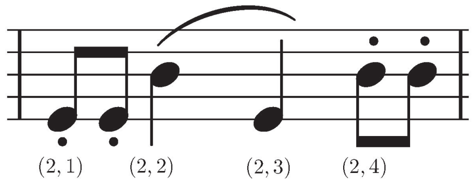
Condition label	C1.0	C1.1	C2.0	C2.1
One-way latency [ms]	25	25	60	60
Adaptive metronome	No	Yes	No	Yes

We devised four different conditions, divided depending on the use of the adaptive metronome and on the additive latency introduced between the musicians, which are summarized in Table II. We considered one-way latency values of 25 ms and 60 ms in order to analyze a case of low and high delay between the performances of the musicians. The parameters of the metronome were set as $\beta = 8$ and $\varepsilon = 0.1$, therefore the metronome performed the beat tracking on the leader and updated its tempo every 8 beats, while the maximum allowed difference from the target tempo was 10.4 BPM.

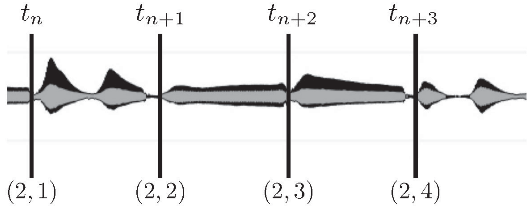
At the beginning of the experiment, the two musicians were introduced to the procedure and could read and practice together the musical score in room 1. Then, after defining the two roles, they were separated and the audio and video communication was tested without latency in the two different rooms. The repetitions were played in random order, different for each couple, and the participants did not have any informa-

(a) Session without adaptive metronome (b) Session with adaptive metronome

Fig. 4. Participants to the experiment



(a) Score



(b) Annotated waveform

Fig. 5. Example of annotation procedure, labels t_n correspond to time values relative to beat onsets, tuples correspond to measure/beat pairs (m_n, b_n)

tion about the latency condition. After each repetition, the musicians answered independently to a questionnaire regarding the difficulty of the interaction and, when present, the effect of the metronome. Finally, the musicians were reunited in room 1 and a brief interview was conducted about the sensations and the preferences of the participants.

IV. Results and Discussion

In this section we present and discuss the results related to the experiment described in Sec. III. Given the limited number of participants, the analysis of the results does not provide statistically significant conclusions, yet the results provide a meaningful insight on the potential of the use of adaptive metronomes in NMP contexts.

In Sec. IV-A we describe the objective evaluation metrics used to analyze the performances of the musicians, whereby the results obtained are then discussed in Sec. IV-B. Finally, in Sec. IV-C we describe the subjective results obtained by means of a short questionnaire administered to the musicians after each repetition.

A. Objective Evaluation Metrics

In the literature, several metrics have been proposed and used in order to analyze the rhythmic trend of the performances [28]. The first step needed to compute these metrics is to manually annotate the audio recordings in order to produce a one-to-one correspondence between the waveform and the score, an example of this procedure is shown in Fig. 5. This procedure consists in annotating the instants when an onset occurs on the beat, denoted as $t_n,\ n = 1, \dots, N$, where N is the number of onsets contained in the recording. Each t_n onset was annotated using as a tuple (m_n, b_n), where m_n denotes the musical measure, while b_n represents the number of beats since the beginning of the measure.

1) Tempo Slope: In order to provide some objective information regarding the ability of the musicians in keeping the tempo during the performance, we first have to convert the set of annotations into tempo samples. For each annotated onset of the performance of a musician, the tempo is measured with respect to the previous onset, therefore is valid in the time interval between two consecutive annotations. The estimation of the tempo has to take into account also the number of beats between the two annotations, which is denoted as Δb_n for the note n with respect to its predecessor $n-1$. The musical score can establish pauses between the notes, or the duration of a note can be longer than one beat, so Δb_n indicates the intervals between the annotations. The tempo between the notes $n-1$ and n is computed as

$$\overline{\delta}(n) = \Delta b_n \frac{60}{t_n - t_{n-1}}, \tag{5}$$

the tempo slope κ is then estimated by computing the linear regression of the points $\overline{\delta}(n), n = 1, \dots, N$. A positive slope $\kappa > 0$ indicates a tendency towards acceleration, while a negative slope $\kappa < 0$ suggests a deceleration. An ideal tempo slope $\kappa = 0$ would suggest that the musician has kept the target tempo τ_t for the whole duration of the performance.

2) Asymmetry: The asymmetry measures the amount of time that one of the two performers lags before the other, it is therefore an indicator of the misalignment between the two musicians during the remote performance. In the scope of this paper, we consider the asymmetry as the mean time that the follower musician F lags behind the leader L. Let us consider the set $\mathcal{N} = \{n | m_n^{(F)} = m_n^{(L)}, b_n^{(F)} = b_n^{(L)}\}$, which contains the indices of the annotated onsets corresponding to notes on the scores that should be played simultaneously by the two musicians. The resulting asymmetry between the musicians can be written as

$$\alpha^{(FL)} = \frac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} (t_n^{(F)} - t_n^{(L)}). \tag{6}$$

B. Objective Results

Absolute value asymmetry results averaged over all the considered couples are presented in Table III. Regarding

TABLE III. ASYMMETRY RESULTS

Condition label	C1.0	C1.1	C2.0	C2.1
$\lvert\alpha^{(FL)}\rvert$[ms]	25	29	42	37

condition C1, the metronome seemed to slightly worsen the musicians' synchronicity, which could be due to the fact that since the latency considered in this condition is small enough to enable natural synchronization between the musicians, the metronome was perceived as a distraction. Instead, as expected, when considering condition C2, where the one-way latency inserted is substantially higher, the synchronicity of the musicians sensibly worsens. In this case, however the use of the adaptive metronome seems to allow a small decrease in the asymmetry and thus an higher synchronicity. Although these results need to be evaluated with further experiments, they seem to suggest that the adaptive metronome could help the musicians in keeping the synchronicity when they face high levels of latency, such as in C2. When the latency is small, as in C1, the musicians are able to synchronize themselves, up to a certain extent, without any external help.

Tempo trend results related to some of the couples and conditions are shown in Fig. 6. Due to the few couples that participated into the experiment and the inherent subjective differences in the musician performances, we decided to avoid showing average results related to the tempo trend and instead we opted to analyze some specific cases. However, we believe that these results suggest that the adaptive metronome could be a viable technique for contrasting the effects due to network latency in NMP contexts.

In Fig. 6a and Fig. 6b we compare tempo trend results related to couple B and to conditions C1.0 and C1.1, respectively. This comparison is interesting, since, contrary to what suggested by the asymmetry results, the adaptive metronome seems to be able to help the musicians in keeping the tempo even when the introduced latency is not excessively high, such as during condition C1. While without any help from the adaptive metronome, Fig. 6a, the musicians progressively decelerated from a tempo of 100 BPM to a tempo of almost 90 BPM. When using the adaptive metronome, Fig. 6a, they seemed to be able to maintain a more stable tempo around 100 BPM, thus remaining closer to the target tempo of $\tau_t = 104$ BPM

In Fig. 6c and Fig. 6d we compare tempo trend results always related to couple B, but to conditions C2.0 and C2.1, respectively, thus considering the setting where the latency level was more challenging. The behavior observed in this case is similar with the one obtained for C1 by the same couple, even though the musicians struggle more at keeping the target tempo. The metronome aids the musicians at keeping a more stable tempo, closer to the target one.

In Fig. 6e and Fig. 6f we compare tempo trend results related to conditions C2.0 and C2.1, respectively, computed regarding the performances of couple C. In this case we can see that the adaptive metronome is not able to avoid the deceleration of the musicians, but it helps in keeping them in a tempo closer to the target one.

TABLE IV. QUESTIONS AND MEAN ANSWERS PER CONDITION TO THE QUESTIONNAIRE REFERRING TO A 5-POINT LIKERT SCALE

Question	Mean			
	C1.0	*C1.1*	*C2.0*	*C2.1*
1) During the performance, how much have you been stressed/dissatisfied?	1.375	1.5	2.75	2.75
2) Has the performance been difficult?	1.875	2.25	2.875	4
3) Has the metronome increased the confidence during the musical interaction? (follower only)	-	3.25	-	2.75
4) Has it been possible to use both the metronome and the auditory feedback of the other musician as a guide during the performance? (follower only)	-	2.75	-	3

An interesting case is also the one of couple D, whereas both the musicians commented positively the usefulness of the metronome, the comparison of their performance in C2.0, shown in Fig. 6g, and C2.1, shown Fig. 6h, seems to suggest that not only the adaptive metronome did not help them, but also prompted them to slightly accelerate.

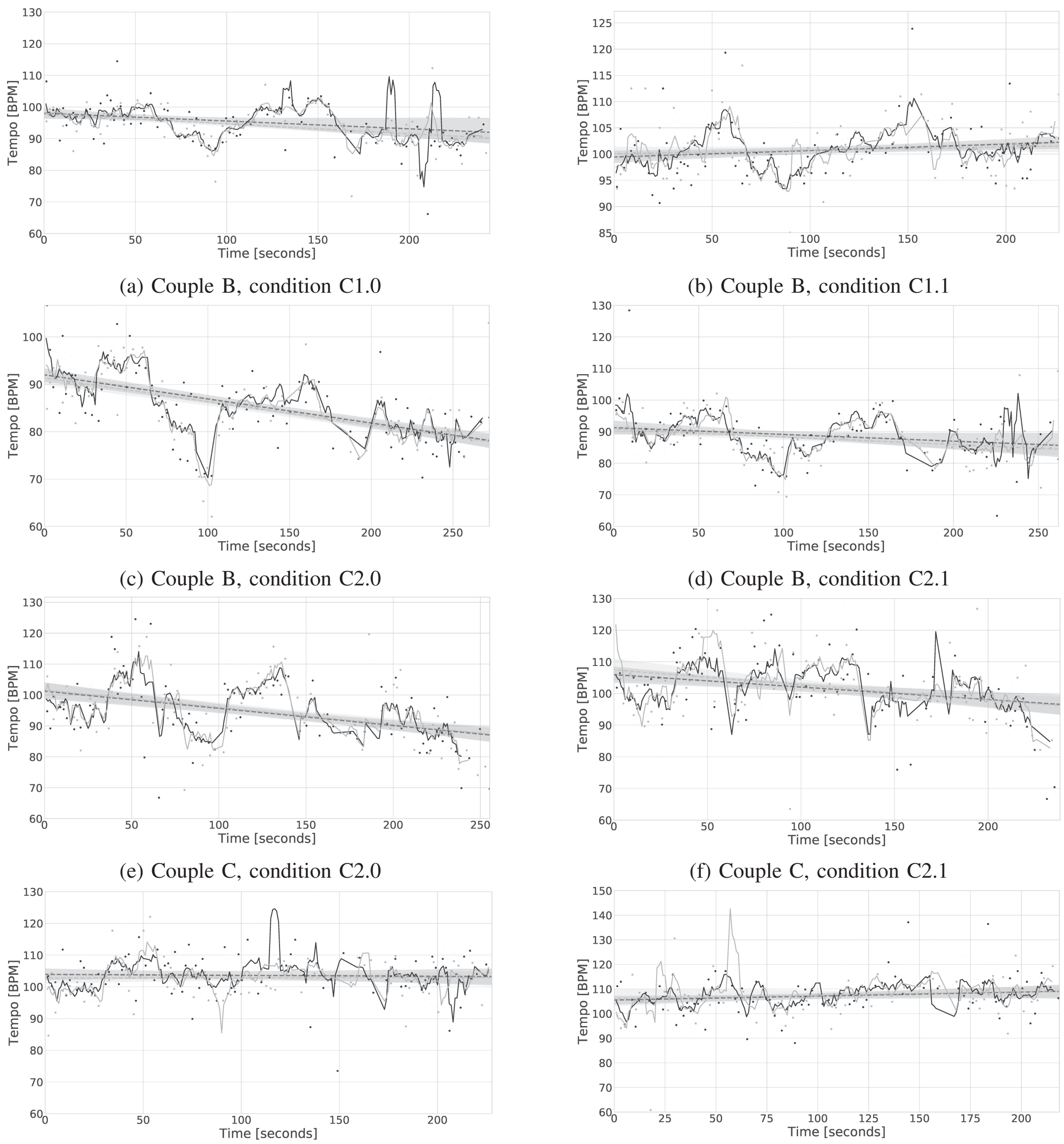
C. Subjective results

Table IV shows the questions and the mean answers of the musicians (in a scale from 1 to 5) over the four conditions of the performances. The first two questions were addressed to all the participants, while the last two questions refer to the perception of the adaptive metronome and thus were presented only to the follower musicians. The mean values of question 1 suggests that the sense of frustration during the performance is more related to the latency rather than the presence of the metronome, in fact the discomfort grows in correspondence to an increase of the latency and does not seem to be influenced by the metronome. The difficulty of the interaction could be instead linked to both the factors, as shown by the answers to question 2, which may suggest that the followers probably found difficulties when trying to balance the auditory feedback of the metronome and the one of the leader.

V. CONCLUSIONS

The increased interest towards the use of network-based remote musical interactions, both for performance and tuition, poses us the need to confront the most important factors that enable a satisfactory use of this technology. The most important among these factors is the latency caused by the network connection, which severely hinders the inherently time-based musical interaction.

In this paper we propose a technique that aims at contrasting the latency, which, to the best of our knowledge, is the first one that adopts the use of adaptive metronomes in NMP contexts. Preliminary experiments, although limited in the number of



(a) Couple B, condition C1.0
(b) Couple B, condition C1.1
(c) Couple B, condition C2.0
(d) Couple B, condition C2.1
(e) Couple C, condition C2.0
(f) Couple C, condition C2.1
(g) Couple D, condition C2.0
(h) Couple D, condition C2.1

Fig. 6. Tempo trend analysis for various couples and conditions, black plots refer to the follower musician, while gray ones to the leader, • is the BPM trend $\overline{\delta(n)}$ - - - - - is the BPM tempo slope κ, ——— is the BPM smoothed trend

participants, show that there is room for improvement in the direction proposed by this paper.

We can devise a few ways in which the presented work could be extended, it would be interesting to apply the adaptive metronome to both musicians, in order to create some sort of virtual conductor that, as done in real ensembles, helps the musicians in synchronizing themselves.

Acknowledgment

This study was conducted for the InterMUSIC project, which received the financial support of the Erasmus+ National Agency under the KA203 Strategic Partnership action under grant number: 2017-1-IT02-KA203-036770.

The authors would like to thank the participants of the tests for their effort and their availability to be contacted for further experiments.

References

[1] J. Lazzaro and J. Wawrzynek, “A case for network musical performance,” in *Proceedings of the 11th international workshop on Network and operating systems support for digital audio and video*, 2001, pp. 157–166.

[2] Á. Barbosa, “Displaced soundscapes: A survey of network systems for music and sonic art creation,” *Leonardo Music Journal*, pp. 53–59, 2003.

[3] M. Iorwerth, D. Moore, and D. Knox, “Challenges of using networked music performance in education,” in *Audio Engineering Society Conference: UK 26th Conference: Audio Education*, Aug 2015. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17857

[4] A. Margaryan, M. Bianco, and A. Littlejohn, “Instructional quality of massive open online courses (moocs),” *Computers & Education*, vol. 80, pp. 77–83, 2015.

[5] 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), Udine*, 2018, pp. 20–23.

[6] L. Comanducci, M. Buccoli, M. Zanoni, A. Sarti, S. Delle Monache, G. Cospito, E. Pietrocola, and F. Berbenni, “Investigating networked music performances in pedagogical scenarios for the intermusic project,” in *2018 23rd Conference of Open Innovations Association (FRUCT)*. IEEE, 2018, pp. 119–127.

[7] 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, 2019.

[8] 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, no. 11, pp. 864–875, 2015.

[9] M. Gurevish, C. Chafe, G. Leslie, and S. Tyan, “Simulation of networked ensemble performance with varying time delays: Characterization of ensemble accuracy.” in *ICMC*, 2004.

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

[11] C. Chafe, M. Gurevich, G. Leslie, and S. Tyan, “Effect of time delay on ensemble accuracy,” in *Proceedings of the International Symposium on Musical Acoustics*, vol. 31, 2004, p. 46.

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

[13] S. Farner, A. Solvang, A. Sæbo, 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.

[14] J.-P. Cáceres and C. Chafe, “Jacktrip: Under the hood of an engine for network audio,” *Journal of New Music Research*, vol. 39, no. 3, pp. 183–187, 2010.

[22] R. Hupke, L. Beyer, M. Nophut, S. Preihs, and J. Peissig, “A rhythmic synchronization service for music performances over distributed networks,” *Fortschritte der Akustik: DAGA*, vol. 45, 2019.

[15] C. Drioli, C. Allocchio, and N. Buso, “Networked performances and natural interaction via lola: Low latency high quality a/v streaming system,” in *International Conference on Information Technologies for Performing Arts, Media Access, and Entertainment*. Springer, 2013, pp. 240–250.

[16] P. Holub, L. Matyska, M. Liška, L. Hejtmánek, J. Denemark, T. Rebok, A. Hutanu, R. Paruchuri, J. Radil, and E. Hladká, “High-definition multimedia for multiparty low-latency interactive communication,” *Future Generation Computer Systems*, vol. 22, no. 8, pp. 856–861, 2006.

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

[18] L. Turchet, “Smart musical instruments: vision, design principles, and future directions,” *IEEE Access*, vol. 7, pp. 8944–8963, 2018.

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

[20] A. Carôt and C. Werner, “Network music performance-problems, approaches and perspectives,” in *Proceedings of the “Music in the Global Village”-Conference, Budapest, Hungary*, vol. 162, 2007, pp. 23–10.

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

[23] R. Oda and R. Fiebrink, “The global metronome: Absolute tempo sync for networked musical performance,” in *Proc. 13th Int. Conf. New Interface Musical Exp., 2013,*, 2016, pp. 1–4.

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

[25] B. D. Giorgi, M. Zanoni, S. Böck, and A. Sarti, “Multipath beat tracking,” *Journal of the Audio Engineering Society*, vol. 64, no. 7/8, pp. 493–502, 2016.

[26] R. D. R. Z. Bos, Pieter and A. Nijholt, “Interacting with a virtual conductor,” in *International Conference on Entertainment Computing*. Springer, 2006, pp. 25–30.

[27] S. Duffy and P. Healey, “A new medium for remote music tuition,” *Journal of Music, Technology and Education*, vol. 10, no. 1, pp. 5–29, 2017.

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