Experimenting With Adaptive Metronomes in Networked Music Performances

Article · October 2021

DOI: 10.17743/jaes.2021.0034

Luca Comanducci
Politecnico di Milano
27 PUBLICATIONS **341** CITATIONS
SEE PROFILE

Giovanni Cospito
Conservatorio di Musica "Giuseppe Verdi"
9 PUBLICATIONS **116** CITATIONS
SEE PROFILE

Augusto Sarti
Politecnico di Milano
390 PUBLICATIONS **5,379** CITATIONS
SEE PROFILE

Experimenting with Adaptive Metronomes in Networked Music Performances*

RICCARDO BATTELLO,[1] **LUCA COMANDUCCI**,[1]
(riccardo.battello@mail.polimi.it) (luca.comanducci@polimi.it)
FABIO ANTONACCI,[1] **GIOVANNI COSPITO**,[2] **AUGUSTO SARTI**,[1]
(fabio.antonacci@polimi.it) (giovanni.cospito@consmilano.it) (augusto.sarti@polimi.it)

[1] *Politecnico di Milano, Milan, Italy,*
[2] *Conservatorio di Milano, Milan, Italy*

The increasing attention towards interactions at a distance and the improvement of digital communications networks have steadily increased the interest towards Networked Music Performance both regarding entertainment and education. Unfortunately the unavoidable network latency remains one of the main issues that prevents a satisfiable remote performance. In this work we propose three different techniques that try to contrast this issue by relying on adaptive metronomes, i.e. metronomes that are able to track the tempo of the musicians through a beat tracking technique. We present a series of preliminary experiments with both professional an amateur musicians that demonstrate that these techniques could be a promising approach as an additional tool for contrasting the impact of latency.

0 INTRODUCTION

The interest towards Networked Music Performance (NMP) has increased steadily over the years, both due to the increased quality of network communication chains and to the constantly growing need for a fruitful distant interaction between people. NMP may be defined as the research field that comprises all the techniques that aim to enable remote musical interaction between musicians as if they were in the same environment [1].

The work presented in this manuscript was developed under the scope of the EU funded INTERmusic (Interactive Environment for Music Learning and Practicing, 2017 - 2020) project [1]. The aim of the project was that of developing a series of best practices in the context of remote environments for music interaction and tuition, especially regarding higher education institutions. Some of the courses courses considered for the project rely on the Massive Open Online Course (MOOC) paradigm, e.g. musical theory courses, and thus do not pose significant synchronicity requirements [2]. Teaching of subjects where, instead, a timed interaction between musicians and tutors is needed in a one-to-one (of few-to-few) fashion, e.g. music practice courses, poses stricter latency requirements.

The impact of latency on one-to-one interactions in the context of chamber music, specifically, was already extensively studied by a series of pilot experiments [3], where musicians had to play through a network connection while facing different artificially introduced latency levels. These experiments were inspired by what already done in the literature through the analysis of rhythmic hand clapping-based performances by musicians positioned in different rooms [4, 5, 6, 7, 8]. The results of the experiments enabled us to define a framework [9] that allows us to devise and evaluate results related to NMP in pedagogical contexts, such as the one presented in this manuscript. As already mentioned, the impact of latency in NMP has been extensively studied in the literature, finding that when the latency begins to fall in the range between 20 ms and 60 ms the musicians generally tend to decelerate w.r.t. the musical tempo [10]. The most natural setting can be devised when the temporal separation between remote musicians is comparable to the one that would have been obtained in a real-life situation.

The issue of latency was partially solved by the introduction of NMP-dedicated softwares such as Jacktrip [11], LOLA [12] or UltraGrid [13]. However, such tools either require high-end hardware, an assumption often unavailable in educational contexts, or do not satisfy the very strict latency requirements needed in chamber music practice.

The aforementioned issues and analyses motivated the development of techniques that can contrast the la-

---

*To whom correspondence should be addressed e-mail: luca.comanducci@polimi.it

[1] http://intermusicproject.eu/

tency without dealing with transmission protocol, but instead by processing the audio signal received during the performance. This poses several advantages, due to the lightweight computational performance of a metronome, which could enable us to envision them as embedded in paradigms like the Internet of Musical Things (IoMusT)[14] and devise smart instruments adapted to NMP by embedding the adaptive metronome procedure [15]. This feature could also make these techniques viable candidates in haptic systems [16] designed to synchronize geographically-displaced visually-impaired musicians [17].

The application of metronomes in NMP was extensively studied in the literature. An user-adjustable metronome was proposed in [18], considering a latency up to several seconds. In [19] the authors organized a NMP between Stanford University and Peking University using a metronome whose tempo was determined by the estimation of the Round Trip Time (RTT) of the underlying network connection.

A global metronome synchronized through the Global Positioning System (GPS) was first presented in [20]. The impact of a global metronome on the performance of the musicians was first studied in [21] through rhythmic handclapping experiments and later also through the use of percussions in [22].

The results obtained in the literature show how metronomes can help the musicians in mantaining the synchronization even when high latency values are present. This is probably due to the fact that they provide a steady and reliable guidance to the musicians during the performance, similarly to what a director does with orchestras.

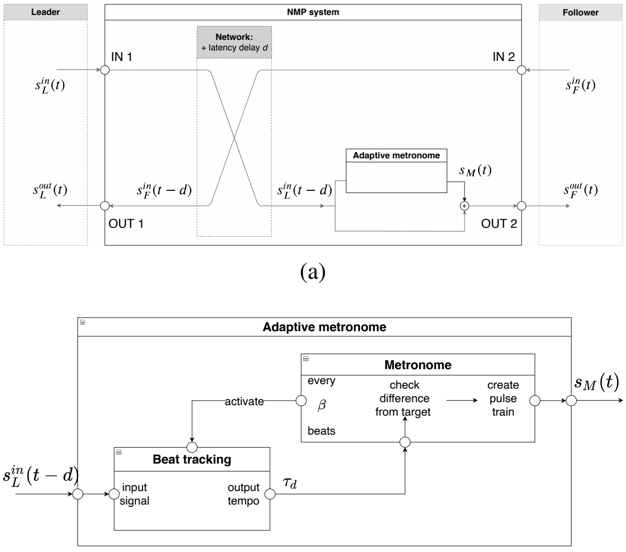
The work presented in this manuscript concerns the use of adaptive metronomes and extends what already presented in [23]. An adaptive metronome is a metronome that is able to change its tempo depending on the tempo of the musicians. This behavior is implemented through a beat tracker. In [23] the adaptive-metronome was implemented in order to emulate the master-slave [24] latency compensation strategy, where a musician, defined as the *follower*, adapts its tempo to that of the partner, i.e. the *leader*. This was implemented by letting the follower hear an adaptive metronome that tracked the tempo of the leader.

In this work we reprise and extend this concept by considering two beat tracking algorithms running on the two interacting musicians. This choice poses more complex design problems, related to how to merge the information related to the tempo of the two musicians. This led to the development of the Crossed Beat Tracking (CBT) technique, where the two musicians simply hear the metronome that tracks the tempo of the partner, and of the Unique Metronome with Virtual Conductor (UMVC) technique, where both musicians are tracked in order to provide a common reference, ideally emulating the concept of a virtual conductor [25]. We also partially report the results already presented in [23] - referred as Single Beat Tracking (SBT) - to ease a comparison. We present objective results. While the SBT technique was tested in a conservatory with professional musicians, the CBT and UMVC were tested with amateur musicians. While more extensive experimental campaigns had been planned, these were not possible due to restrictions due the COVID-19 pandemic. For these reasons, the results presented in this paper should be considered as anecdotal and do not claim statistical meaningfulness.

Nonetheless these results show that the metronome generally helps the musicians in keeping a steady tempo, especially at higher latency values, and is not perceived as intrusive during the performance. We believe that these analysis suggest that adaptive metronomes could be beneficial in NMP applications and motivate us to further experimentation.

The rest of the manuscript is organized as follows. Sec. 1 presents the details of the technique based on the use of a single adaptive metronome, while Sec. 2 presents the details of the two techniques based on the adoption of multiple adaptive metronomes. Sec. 3 presents results related to the experimental application of the aforementioned techniques. Finally Sec. 4 draws some conclusions.

1 SINGLE BEAT TRACKING WITH MASTER/SLAVE APPROACH



(a)

(b)

Fig. 1. Schematic representation of the SBT technique. (a) shows the technique procedure, while (b) the architecture of the adaptive metronome.

In this section we present the latency-compensation SBT technique already introduced in [23]. This technique works under the assumption that the musician behave following the *master-slave* latency compensation strategy [24]. In this context, the musicians take two distinct roles: the leader, who determines the tempo of the performance, and the follower, who tries to synchronize as much as possible with the tempo of the leader. The metronome, then, is applied in order to track the tempo of the leader and the corresponding beat signal is sent to the follower in order to help him to keep a steady performance. A schematic representation of the technique is shown in Fig. 1(a).

We can think of SBT as a multiple-input multiple-output system. The inputs consists of the audio signals of the real-time recorded performances of the musicians $s_L^{in}(t)$ and $s_F^{in}(t)$, where t denotes the discrete time index and L and F the leader and follower musician, respectively. The output of the system, instead, consists of the two signals $s_L^{out}(t)$ and $s_F^{out}(t)$. If we denote with d the latency due to the network connection, we can define the signal received by the leader as

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

which is simply the delayed signal of the follower's performance. The signal received by the follower can instead be defined as

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

which consists of the sum of the delayed signal of the leader plus the corresponding metronome signal $s_M(t)$ consisting in a train of pulses, where each pulse corresponds to one beat. The time difference Δ_{t_M} between adjacent pulses depends on the desired Beats Per Minute (BPM) tempo τ_M of the performance and can be computed as $\Delta_{t_M} = 60/\tau_M$.

The internal working of the adaptive metronome is shown in Fig. 1b, it can be subdivided into two main processes: the beat tracking of the follower's signal, which modifies the current BPM estimation and the metronome, which generates the corresponding signal $s_M(t)$ and modifies the tempo accordingly to the BPM estimate. The beat tracker detects a new tempo τ_d every β beats, where β is a user-specified parameter selected in order to avoid abrupt changes in the tempo that would be harmful for the performance [25]. Each time a new tempo value τ_d is detected, it is compared to the target tempo τ_t and the tempo of the metronome τ_M is updated every $n\beta$ beats, where $n \in \mathbb{N}$, according to

$$\tau_M(n\beta) = \begin{cases} \tau_d & if\ |\tau_d - \tau_t| < \varepsilon\tau_t \\ \tau_M((n-1)\beta), & \text{otherwise} \end{cases}, \tag{3}$$

where ε defines the maximum valid difference between target and detected tempo as a percentage of the target tempo τ_t. If the absolute difference $|\tau_d - \tau_t|$ between detected and target tempo is smaller than the threshold $\varepsilon\tau_t$ then τ_d is selected as the current BPM value for the metronome, otherwise it is left unmodified, thus avoiding extremely abrupt tempo changes.

We initialize the detected tempo τ_d to the target tempo τ_t, then, when a new value for τ_d is detected, the period of the pulse train $s_M(t)$ is modified according to (3).

The multipath beat tracker [26] is used for the estimation of the tempo. Its main advantage is the low computational cost and the ease of adjustment in an adaptive scenario.

2 Double Beat Tracking

In this section we present two latency compensation techniques based on the adoption of adaptive metronomes that remove the master/slave assumption. Specifically we employ two beat trackers that simultaneously track the two musicians. This enhances the possibilities of the system, but also poses questions regarding how to merge the tempo information related to the two musicians.

We follow an approach that gradually enhances the complexity of the method. While the SBT uses only one metronome tracking the performance of the leader, the CBT technique simultaneously tracks both performers and provides to each of them the metronome adapted to the tempo of the other musician. The UMVC technique does an additional step, by merging the information acquired from the tracking of both musicians in order to provide a common reference tempo. Both the techniques that will be presented in the following apply the multipath beat tracker presented in [26] for tempo estimation.

2.1 Crossed Beat Tracking

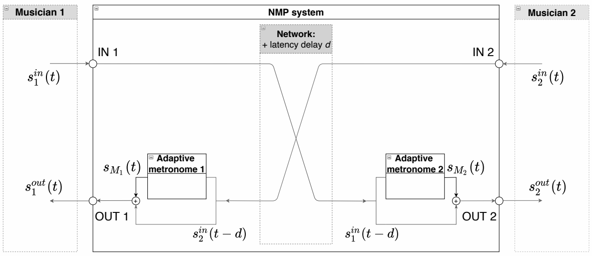


Fig. 2. Schematic representation of the CBT technique.

In CBT two simultaneous and independent metronomes are used, each tracking separately one of the two musicians.

The hypothesis behind this approach is of following a sort of reciprocal master/slave approach. By being able to hear the metronome of the other performer, the musicians could be able to adapt to each other, trying to contrast the impact of the latency.

A schematic representation of the CBT technique is depicted in Fig. 2. The two entities that communicate through the NMP system are *musician 1* and *musician 2*. The signals acquired from the musicians are respectively $s_1^{in}(t)$ and $s_2^{in}(t)$, while the signals presented to the musicians are respectively $s_1^{out}(t)$ and $s_2^{out}(t)$. In the CBT scheme

$$s_1^{out}(t) = s_2^{in}(t-d) + s_{M_1}(t), \tag{4}$$

where $s_{M_1}(t)$ is the pulse train produced by the adaptive metronome on the side of *musician 1* and d is the latency of the network. The tempo τ_{M_1} of this metronome is initialized at the target tempo and set at the beat $n\beta$, where $n \in \mathbb{N}$, as

$$\tau_{M_1}(n\beta) = \begin{cases} \tau_{d_2} & |\tau_{d_2} - \tau_t| < \varepsilon\tau_t \\ \tau_{M_1}((n-1)\beta) & \text{otherwise} \end{cases}, \tag{5}$$

where β, ε and τ_t are defined as in Section 1 and τ_{d_2} is the detected tempo of *musician 2*. The signal reproduced for *musician 2* is likewise defined as

$$s_2^{out}(t) = s_1^{in}(t-d) + s_{M_2}(t), \tag{6}$$

where $s_{M_2}(t)$ is the output of the second adaptive metronome. Its tempo τ_{M_2} is computed at each beat $n\beta$ as

$$\tau_{M_2}(n\beta) = \begin{cases} \tau_{d_1} & |\tau_{d_1} - \tau_t| < \varepsilon\tau_t \\ \tau_{M_2}((n-1)\beta) & \text{otherwise} \end{cases}, \qquad (7)$$

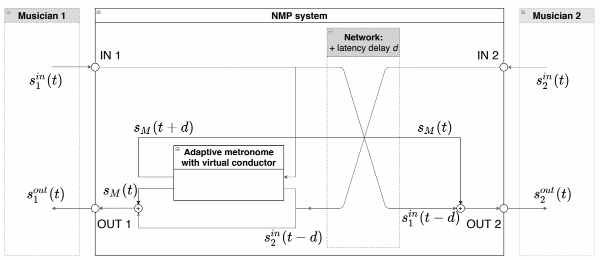
where τ_{d_1} is the detected tempo of *musician 1.*

2.2 Unique Metronome with Virtual Conductor

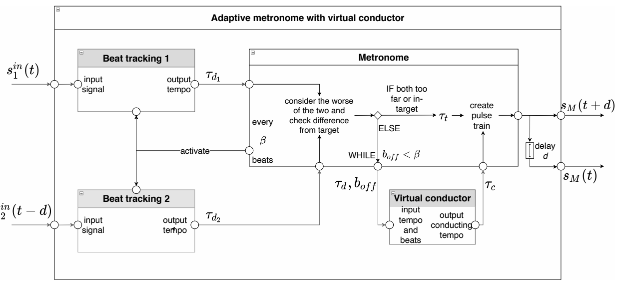
In this section we present the UMVC technique, which, as in the CBT case, employs two beat trackers in order to detect the tempo of the two musicians. However, differently from the CBT technique, UMVC derives a common metronome. This technique poses more challenging issues w.r.t. the previous cases. The metronome has to adapt to the performance considering two different tempos and at the same time has to create a guide that can help both musicians, while taking into account the inherent latency of the system.

The derivation of a signal that can act as a common guide to the musicians can be obtained by extending the framework of the Virtual Conductor (VC) [25]. The VC is a system designed in order to automatically emulate the behavior of a human conductor in local performances. The idea behind the VC is to provide a tool that can follow the tempo of the musicians and guide them in order to maintain the target tempo of the performance, by balancing through the metronome the acceleration and decelerations of the musicians.

Differently from the previoulsy proposed technique, in the UMVC method a proper synchronization of the two metronomes is required in order to have a proper behavior of the system. In this work we consider the latency to be fixed and known a-priori, however synchronization can be achieved using a global metronome [20, 27] or by performing an estimation of the network latency [28].



(a)



(b)

Fig. 3. Schematic representation of the UMVC technique. (a) shows the technique procedure, while (b) the architecture of the adaptive metronome.

A schematic representation of the method is shown in Fig. 3. The signals of the performances of the musicians 1 and 2 are defined as $s_1^{in}(t)$ and $s_2^{out}(t)$, respectively. The audio signal reproduced for musician 1 is defined as

$$s_1^{out}(t) = s_2^{in}(t-d) + s_M(t), \qquad (8)$$

while for musician 2 it is

$$s_2^{out}(t) = s_1^{in}(t-d) + s_M(t+d), \qquad (9)$$

where $s_M(t+d)$ compensates the network latency to assure synchronization between the musicians. Being $s_M(t)$ a synthetically generated pulse, a temporal anticipation of the signal is not an issue.

A representation of the architecture of the UMVC adaptive metronome is depicted in Fig. 3b.

The adaptive metronome receives as input the two signals coming from the musicians, $s_1^{in}(t)$ and $s_2^{in}(t-d)$ and the tempo τ_{d_1} and τ_{d_2} of musicians 1 and 2, respectively. A new value is detected every β beats.

Since the musicians need a common reference value, the metronome selects only one of the detected tempos of the two musicians τ_{d_1} and τ_{d_2}, according to

$$\tau_d(n\beta) = \begin{cases} \tau_{d_1} & if\ |\tau_t - \tau_{d_1}| > |\tau_t - \tau_{d_2}| \\ \tau_{d_2} & if\ |\tau_t - \tau_{d_2}| > |\tau_t - \tau_{d_1}| \end{cases}, \qquad (10)$$

i.e. the common reference is the tempo more distant from the target one τ_t . However, always with the objective of avoiding tempo changes that are too abrupt, if $\tau_d > \varepsilon_{max}\tau_t$, i.e. the new detected tempo is too far from the target, the tempo of the other musician is selected.

The period of the train of pulses is

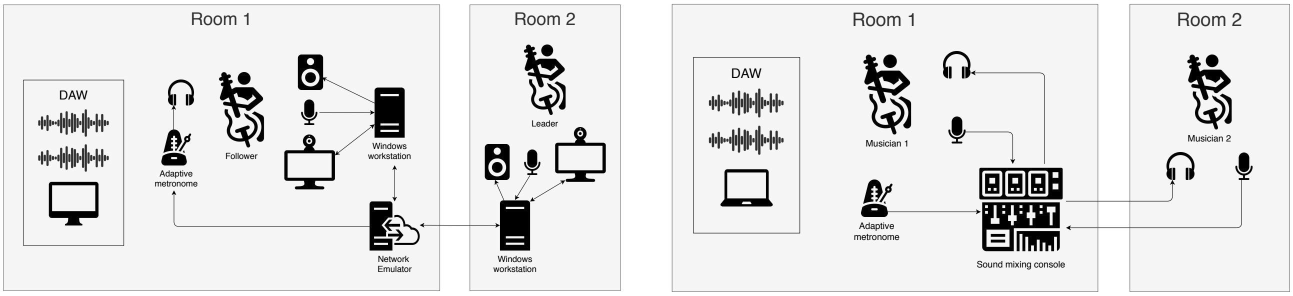
$$\tau_M(n\beta) = \begin{cases} \tau_c & \varepsilon_{min}\tau_t < |\tau_d - \tau_t| < \varepsilon_{max}\tau_t \\ \tau_t & \text{otherwise} \end{cases}, \qquad (11)$$

where τ_c is the conducting tempo computed by the virtual conductor.

For example, if the metronome receives two detected tempos that are both smaller than the target and in the range defined by ε_{min} and ε_{max}, it sets τ_d to the tempo which is farther from the target. During the successive β beats, the tempo of the metronome gradually accelerates starting from a value close to τ_d, in the attempt of guiding the musicians towards the target. Then, the beat tracking processes detect that both musicians play at the target tempo, so the tempo of the metronome is fixed to the target for the successive β beats.

3 EXPERIMENTAL STUDIES

In this section we present two experimental studies that aim at a preliminary demonstration of the effectiveness of the techniques presented in this manuscript. The first study was already presented in [23] and will be partially re-proposed here in order to provide results related to the



(a) (b)

Fig. 4. Setups of experiment 1 (a) and of experiment 2 (b).

SBT technique. The second study, instead, deals with the application of the techniques based on the adoption of multiple beat trackers, CBT and UMVC. We will first review the adopted evaluation metrics and then present in detail the aforementioned experiments.

3.1 Evaluation Metrics

The task of evaluating the rhythmic trend of a musical performance has been extensively studied in the literature and led to the development of a series of well-known metrics [29].

In order to be able to analyze the ability of the musicians in executing the performance all the considered metrics require a one-to-one mapping between the score and the raw audio signals. Since an automatic procedure would be inevitably prone to detection errors, this task is usually performed by manually annotating the time instants when onsets occur on the beat, denoted as $t_n, n = 1, \dots, N$, where N corresponds to the total number of onsets in the performance. In the case of this work, each t_n was annotated through a tuple (m_n, b_n), where m_n denotes the index of the musical measure, while b_n the number of beats occurred since the beginning of the corresponding measure. For a thorough explanation of the annotation procedure, we refer the interested reader to [30].

To quantify the ability of the musicians in maintaining the tempo, we first have to convert the annotations into tempo values. Let us define Δb_n as the number of beats occurring between a note n and the previous one. This number varies, due both to the presence of rests in the score and to the fact that notes can last for either more or less than one beat. We can define the tempo between notes $n-1$ and n as

$$\overline{\delta}(n) = \Delta b_n \frac{60}{t_n - t_{n-1}}, \tag{12}$$

The *tempo trend* κ can then be defined as the discrete time derivative of the sequence $(n, \overline{\delta}(n)), n = 1, \dots, N$. This measure allows us to analyze the overall trend of the performance, i.e. $\kappa < 0$ suggests a tendency over a deceleration, while $\kappa > 0$ a tendency towards acceleration. In an ideal performance $\kappa = 0$, i.e. the musician keeps always the same tempo.

The other objective metric that will be considered in this work is the *asymmetry*, which measures the amount of time that one performer lags w.r.t. the other. The asymmetry of the musician 1 w.r.t. musician 2 can be defined as

$$\alpha^{12} = \frac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} (t_n^1 - t_n^2), \tag{13}$$

where $|\cdot|$ denotes the cardinality of a set.

3.2 Experimental study 1

Table 1. Participants to experiment 1

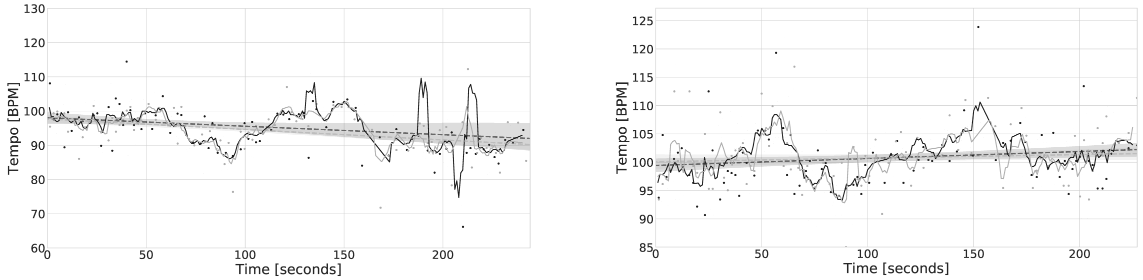
Couple A	Couple B	Couple C	Couple D
Saxophone	Guitar	Oboe	Oboe

Table 2. Experiment 1 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

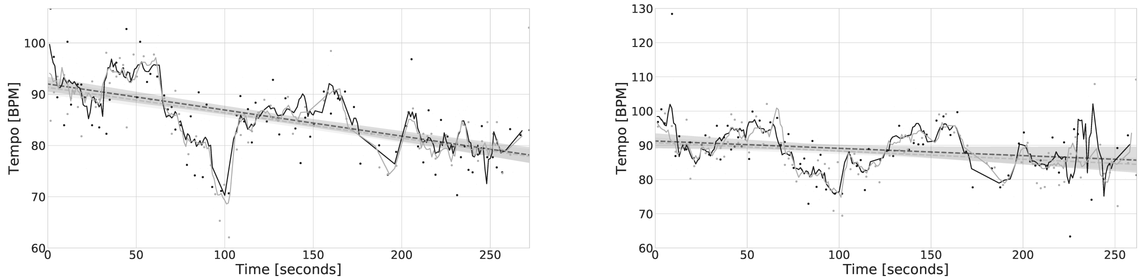
The first experiment was conducted at the Conservatorio di Musica ”Giuseppe Verdi” in Milan, Italy. The aim of this test was to evaluate how the SBT approach could improve the musicians’ performances when dealing with a fixed network latency. The objective of this test was also to provide a first feedback regarding the application of adaptive metronomes to networked music performances before testing more sophisticated techniques such as CBT and UMVC.

The setup of the experiment is depicted in Fig. 4(a). We considered two musicians located in two acoustically isolated rooms, connected through loudspeakers and microphones in order to provide auditory feedback, and cameras and screens in order to provide visual feedback. Further details regarding the instrumentation setup are contained in [23]. The latency was introduced in the system through a network emulator, using the linux command *tc* to insert a two-way latency. Four pairs of musicians took part in the experiment, as summarized in Table 1, all of them were conservatory students aged between 17 and 22 years, with at least 7 years of musical experience. Players in all the pairs were playing the same instruments. Each couple



(a) Couple B, condition C1.0

(b) Couple B, condition C1.1



(c) Couple B, condition C2.0

(d) Couple B, condition C2.1

Fig. 5. Experiment 1. Tempo trend analysis for couple B, considering all the proposed conditions. Black and gray colors are used for the follower and leader musicians, ● is the BPM trend $\overline{\delta(n)}$, - - - - - is the BPM tempo slope κ, ——— is the BPM smoothed trend. The bands around the BPM tempo slope represent the corresponding confidence intervals.

Table 3. Asymmetry results - experiment 1

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

played four repetitions of the same musical score according to the conditions reported in Table 2. We compared the tempo of the musicians when SBT was active and inactive for two latency values, namely 25 ms and 60 ms. The latencies used in this experiment and the labels associated to the different conditions are summarized in Table 2. The parameters of the metronome were set to $\beta = 8$ and $\varepsilon = 0.1$, thus the metronome updated itself every 8 beats allowing a maximum deviation of 10.4 BPM from the target tempo $\tau_t = 104$ BPM. The scores considered in the experiment were inspired by Bartok's *Mikrokosmos* piano pieces [31], due to their focus towards rhythmic and melodic expression relationships, as detailed in [30]. We report in Table 3 the absolute asymmetry values averaged over all the pairs. Results suggest that in the condition C1 the metronome slightly worsens the performance w.r.t. condition C0. Instead, if we consider condition C2, the performance improves. This could be due to the fact that when the latency is relatively small, the musicians are able to adapt themselves without the need of external help. When facing higher levels of latency instead, the metronome may be helpful in allowing the musicians to maintain the synchronicity between them.

While a systematic or statistical analysis of the tempo trend of the musicians is not possible, due to the small number of participants, we show here a few examples regarding pair B and the considered conditions, reported in Fig. 5. In Fig. 5(a) and 5(b) we show results related to the tempo trend obtained by the musicians during the conditions C1.0 and C1.1, respectively. In this case the SBT technique seems to actually help the musicians, even in a condition where the amount of latency is not excessive and contrarily to what indicated by the asymmetry results. In Fig. 5(c) and 5(d) we show results related to pair B, for the conditions C2.0 and C2.1, respectively. The behavior is similar to the one obtained for the C1 condition, since the metronome seems to help the musicians, who however need more effort in keeping their tempo closer to the target one of 104 BPM.

3.3 Experimental study 2

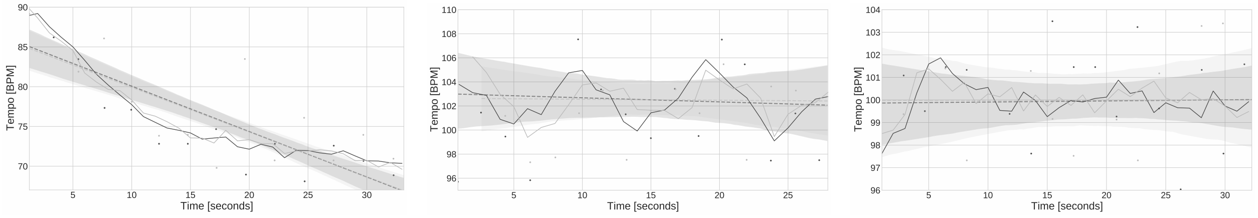
Table 4. Participants to experiment 2

	Couple E	**Couple F**	**Couple G**
Musician 1	Bass Guitar	Bass Guitar	Drum Pad
Musician 2	Drum Clap	Hand Clapping	Hand Clapping

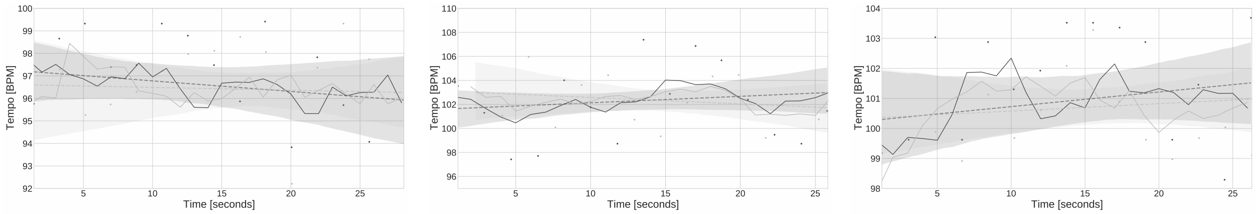
Table 5. The conditions of the three repetitions in experiment 2

Condition label	**M0**	**M1**	**M2**
Adaptive metronome	No	CBT	UMVC
One-way latency [ms]	68		

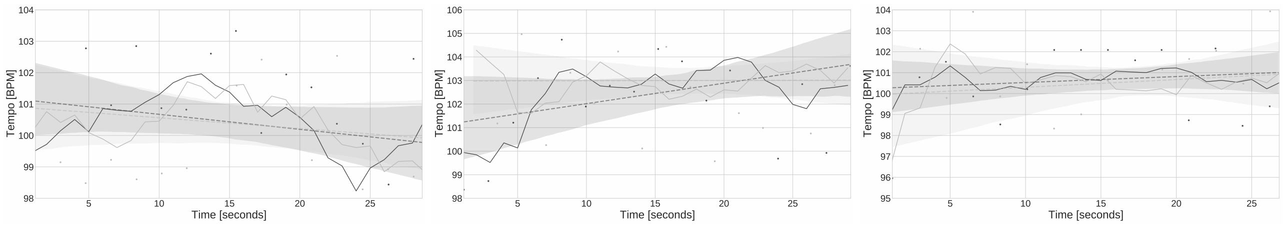
The second experimental study was conducted in a private building in Como, Italy. Due to the COVID-19 pandemic, a more extensive experimentation was not possible. However we believe that these results, although limited in



(a) Couple E, condition M0 (b) Couple E, condition M1 (c) Couple E, condition M2



(d) Couple F, condition M0 (e) Couple F, condition M1 (f) Couple F, condition M2



(g) Couple G, condition M0 (h) Couple G, condition M1 (i) Couple G, condition M2

Fig. 6. Experiment 2. Tempo trend analysis for couples E,F and G, considering all the proposed conditions. Black and gray lines refer to musician 2 and 1, ● is the BPM trend $\overline{\delta(n)}$ ----- is the BPM tempo slope κ, —— is the BPM smoothed trend. The bands around the BPM tempo slope represent the corresponding confidence intervals.

number and scope, can provide a first insight in the applicability of this technique to more realistic scenarios.

The setup of the experiment is shown in Fig. 4b, the musicians were located in two acoustically isolated rooms and connected by means of a Yamaha TF5 mixing console. During audio transmission, a digital delay was introduced in order to simulate a fixed network latency. Each musician worn headphones for listening to the performance of the partner and to the adaptive metronome. The metronomes ran in room 1, since the inputs for the beat trackers were provided by the mixing console, and their signals were added to the outputs of the headphones. The processing latency introduced by the mixing console and the connections was negligible. A Digital Audio Workstation, which consisted of an Logic Pro X running on an Apple MacBook and the Yamaha TF5 as audio card, recorded the performance for the post-experiment analysis.

Differently from experiment 1, the musicians did not follow a specific score, as they had to simply keep the tempo by playing their instrument every quarter note beat, similarly to what done in [5]. The target tempo was defined as $\tau_t = 100$ BPM and the artificially introduced one-way latency corresponds to $d = 68$ms. We define the conditions according to Table 5 by devising one condition for both the CBT andd UMVC conditions and for the case where no metronome was applied.

The CBT method was initialized with $\beta = 8$ and $\varepsilon = 0.1$, therefore the maximum allowed difference from the target tempo was of 10 BPM and the beat tracking was performed every 8 beats. The parameters for the UMVC technique were set to $\beta = 8$, $\varepsilon_{min} = 0.01$ and $\varepsilon_{max} = 0.2$. This means that the virtual conductor was active in presence of a difference from the target in the range from 1 to 20 BPM, avoiding abrupt changes during the performance. The adaptive weight computed by the virtual conductor could vary from $\lambda_{min} = 0.2$ s to $\lambda_{max} = 1$ s.

The three conditions were proposed to musicians in random order. Each repetition had an approximate duration of 30 s, and the musicians had to play at least $48 - 50$ quarter-note beats.

The pairs are shown in Table 4. Pairs composed by all the possible combinations of three amateur musicians have been used.

In Table 6 we report absolute value asymmetry results averaged over all the couples for each condition. The results suggest that the condition MO is the one that corresponds to the lowest asymmetry, i.e. the one where no metronome was applied. Asymmetry values related to M1 and M2 are higher, but still lower than the corresponding values found for C2.0 an C2.1 in experiment 1, where musicians were facing a latency of a similar value.

In Fig. 6 we report tempo trend results for all the couples and conditions involved in experiment 2. As it is possible

Table 6. Asymmetry results - experiment 2

Condition label	M0	M1	M2
$\lvert\alpha^{(12)}\rvert$[ms]	8	33	14

to see, while condition M0 is associated to low asymmetry values, it is associated to a substantial decrease in the tempo during the performance. This is extremely noticeable in Fig 6(a), but also visible in Fig 6(d) and Fig 6(g). This behavior could be explained by the hypothesis that when no metronome is active each musician relies more on the audio feedback of the other performer, trying to reciprocally match the tempo. While this makes it possible to attain a lower asymmetry, it causes the musicians to slow down each other.

The CBT technique instead seems to avoid excessive slow-downs in the musicians performance, differently from M0. Indeed, the tempo seems to be more stable around the target one of $t_t = 100$ BPM. Couple E, Fig 6(b) shows a decrease of less than 1 BPM during the whole performance, while couple F, Fig 6(e) exhibits an acceleration towards the end of the performance, but amounting again to less than 1 BPM. In the case of couple G, Fig 6(h), the couple seems to be constantly accelerating, however they keep performing at pace not too far from the target tempo τ_t.

In the case of the UMVC technique, the musicians seem to be able to maintain the tempo for the whole duration of the performance, better than in the case of M0 and M1. The change in tempo for the duration of the performance never exceeds 1 BPM, which can be considered as a negligible change in tempo. Also, in all cases, the couples seem to be able to stick around the target tempo $\tau_t = 100$ BPM. Couple E, Fig 6(c) seems to be able to maintain the tempo almost steady around 100 BPM for the whole duration of the session, couple F, Fig 6(f), slightly accelerates, however by less than 1 BPM, finally the behavior of couple G, Fig 6(i), is similar to the one of couple F.

4 CONCLUSION

In this manuscript we presented three techniques that apply adaptive metronomes in the context of Networked Music Performance, in order to diminish the impact of latency on the interaction between musicians. Specifically, we presented three techniques that increase in complexity w.r.t how the information due to the tracking of the musicians is combined. The SBT techniques presents the case where a single beat tracker is used in order to implement a guided master-slave latency compensation technique. The CBT technique uses two adaptive metronomes in order to track simultaneously the two musicians, while the UMVC technique performs an additional step by merging this information in order to act similarly to a virtual conductor. Experimental results, although limited in number and scope, demonstrate the potential of such techniques and motivate us to a more extended experimental campaign.

5 ACKNOWLEDGMENT

This study was partially conducted as part of 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. We also would like to thank Stefano Delle Monache, Enrico Pietrocola and Filippo Berbenni for the help in setting up the experiment performed in the conservatory.

6 REFERENCES

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

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

[3] 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 *Proceedings of the XXII Colloquium on Musical Informatics (CIM), Udine*, pp. 20–23 (2018).

[4] M. Gurevish, C. Chafe, G. Leslie, S. Tyan, “Simulation of Networked Ensemble Performance with Varying Time Delays: Characterization of Ensemble Accuracy.” presented at the *ICMC* (2004).

[5] C. Chafe, M. Gurevich, “Network time delay and ensemble accuracy: Effects of latency, asymmetry,” presented at the *Audio Engineering Society Convention 117* (2004).

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

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

[8] S. Farner, A. Solvang, A. Sæbo, 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).

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

[10] 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 Audio Engineering Society*, vol. 63, no. 11, pp. 864–875 (2015).

[11] J.-P. Cáceres, 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).

[12] C. Drioli, C. Allocchio, N. Buso, “Networked performances and natural interaction via LOLA: Low latency high quality A/V streaming system,” presented at the *International Conference on Information Technologies for Performing Arts, Media Access, and Entertainment*, pp. 240–250 (2013).

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

[14] L. Turchet, C. Fischione, G. Essl, D. Keller, M. Barthet, “Internet of musical things: Vision and challenges,” *IEEE Access*, vol. 6, pp. 61994–62017 (2018).

[15] L. Turchet, “Smart Musical Instruments: vision, design principles, and future directions,” *IEEE Access*, vol. 7, pp. 8944–8963 (2018).

[16] L. Turchet, M. Barthet, “Co-design of Musical Haptic Wearables for electronic music performer’s communication,” *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 2, pp. 183–193 (2018).

[17] A. Tanaka, A. Parkinson, “Haptic wave: A cross-modal interface for visually impaired audio producers,” presented at the *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2150–2161 (2016).

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

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

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

[21] r. hupke, l. beyer, m. nophut, s. preihs, j. peissig, “effect of a global metronome on ensemble accuracy in networked music performance,” *journal of the audio engineering society* (2019 october).

[22] R. Hupke, J. Peissig, A. Genovese, S. Sridhar, A. Roginska, “Impact of Source Panning on a Global Metronome in Rhythmic Networked Music Performance,” presented at the *2020 27th Conference of Open Innovations Association (FRUCT)*, pp. 73–83 (2020), [Online]. Available: 10.23919/FRUCT49677.2020.9211004.

[23] R. Battello, L. Comanducci, F. Antonacci, A. Sarti, S. Delle Monache, G. Cospito, E. Pietrocola, F. Berbenni, “An Adaptive Metronome Technique for Mitigating the Impact of Latency in Networked Music Performances,” presented at the *2020 27th Conference of Open Innovations Association (FRUCT)*, pp. 10–17 (2020).

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

[25] P. Bos, D. Reidsma, Z. Ruttkay, A. Nijholt, “Interacting with a virtual conductor,” presented at the *International Conference on Entertainment Computing*, pp. 25–30 (2006).

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

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

[28] L. Deng, H. Zheng, X.-Y. Liu, X. Feng, Z. D. Chen, “Network Latency Estimation With Leverage Sampling for Personal Devices: An Adaptive Tensor Completion Approach,” *IEEE/ACM Transactions on Networking*, vol. 28, no. 6, pp. 2797–2808 (2020).

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

[30] S. Delle Monache, L. Comanducci, M. Buccoli, M. Zanoni, A. Sarti, E. Pietrocola, F. Berbenni, G. Cospito, “A presence-and performance-driven framework to investigate interactive networked music learning scenarios,” *Wireless Communications and Mobile Computing*, vol. 2019 (2019).

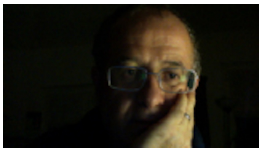
[31] B. Bartók, *Mikrokosmos, 153 Progressive Piano Pieces, New Definitive Edition* (Boosey & Hawkes Music Publishers Limited, London, UK) (1987).

THE AUTHORS

Riccardo Battello

Luca Comanducci

Fabio Antonacci



Giovanni Cospito

Augusto Sarti

Riccardo Battello received the M.S. Degree in Computer Science and Engineering from the Politecnico di Milano, Milan, Italy in 2020. He conducted his work on the application of adaptive metromes to Networked Music Performance problems at the Image and Sound Processing Lab of the Elettronica, Informazione e Bioingegneria (DEIB) at the Politecnico di Milano, Milan, Italy.

●

Luca Comanducci was born in Genova, Italy on May 07, 1991. He received the B.S. degree in Music Information Science from the University of Milan, Milan, Italy, in 2014 and the M.S. degree in Computer Science and Engineering from Politecnico di Milano, Milan, Italy, in 2018, where he is currently working toward the Ph.D. degree in information engineering within the Dipartimento di Elettronica, Informazione and Bioingegneria. His main research interests concern the application of Deep Learning techniques to space-time audio signal processing.

●

Fabio Antonacci was born in Bari, Italy, on July 26, 1979. He received the Laurea degree in 2004 in telecommunication engineering and the Ph.D. degree in information engineering in 2008, both from the Politecnico di Milano, Milan, Italy. He is currently an Assistant Professor at the Politecnico di Milano. His research focuses on space-time processing of audio signals, for both speaker and microphone arrays (source localization, acoustic scene analysis, rendering of spatial sound) and on modeling of acoustic propagation. He is a member of the IEEE Audio and Acoustic Signal Processing Technical Committee and of the EURASIP SAT on Audio, Speech and Music Signal Processing.

●

Giovanni Cospito is a professor of Electroacoustic Music Composition, and Coordinator of the Department of Music and New Technologies and the School of Electronic Music at the Music Conservatory of Milano. His research interests span education and computing technologies for music production, which he promotes through cooperations with the Politecnico of Milano, the Swiss SSPM, the UNIMI University of Milano and the University of Trento. He also collaborated with several music research centers, including the CSC (Padua, Italy), GMVL (Lyon, France), AGON (Milano, Italy), EMS (Stockholm, Sweden), and IRCAM (Paris, France), among the others. He is currently member of the CEDIM (Center for Electroacoustics and Digital Interactions) of the San Fedele Cultural Foundation-Milan, and Coordinator of Erasmus+ research project INTERMUSIC (INTERactive environment for MUSIC learning and practising, 2017 - 2020).

●

Augusto Sarti received his Ph.D. in Information Engineering from the University of Padova, Italy, in 1993 (joint program with University of California, Berkeley). After a few months as Post-Doc researcher at UC Berkeley, he joined the Faculty of the Politecnico di Milano (PoliMI), Italy, where he is currently a Full Professor. In 2013, he joined the University of California, Davis, as a Full Professor (adjunct). He is the co-founder of the Image and Sound Processing Group at PoliMI, with which he promoted/coordinated and contributed to numerous European projects. He is also the founder and coordinator of PoliMI's M.Sci. program in Music and Acoustic Engineering. He coauthored over 300 scientific publications on international journals and congresses and numerous patents in the multimedia signal processing area. He has been an Active Member of the IEEE Technical Committee on Audio and Acoustics Signal Processing and of the Editorial Board of the IEEE.