Audio Engineering Society

Convention e-Brief 49

Presented at the 147th Convention
2019 October 16 – 19, New York

A Latency Measurement Method for Networked Music Performances

Robert Hupke[1], Sripathi Sridhar[2], Andrea Genovese[2], Marcel Nophut[1], Stephan Preihs[1], Tom Beyer[2], Agnieszka Roginska[2], and Jürgen Peissig[1]

[1]*Leibniz University Hannover, Germany*
[2]*New York University, NY, USA*

Correspondence should be addressed to Robert Hupke (`hupke@ikt.uni-hannover.de`)

ABSTRACT

The New York University and the Leibniz University Hannover are working on future immersive Networked Music Performances. One of the biggest challenges of audio data transmission over IP-based networks is latency, which can affect the interplay of the participants. In this contribution, two metronomes, utilizing the Global Positioning System to generate a globally synchronized click signal, were used as a tool to determine delay times in the data transmission between both universities with high precision. The aim of this first study is to validate the proposed method by obtaining insights into transmission latency as well as latency fluctuations and asymmetries. This work also serves as baseline for future studies and helps to establish an effective connection between the two institutions.

1 Introduction

New York University (NYU) and Leibniz University Hannover (LUH) are both investigating novel interactive concepts for the immersive audiovisual linking of distributed live events at different geographic locations. A joint collaborative project was thus started between the Live Interactive PMSE Services (LIPS) [1] and Holodeck [2] project to test and establish NMP connections.

As Networked Music Performances (NMP) are based on audio data transmission over IP-based networks, they exhibit a certain latency, latency fluctuation, limited bandwidth, packet loss and asymmetry within the data transmission. Several publications [1, 2] have shown that one of the biggest challenges of playing music together over Wide Area Network (WAN) connections is the network latency, which has a high impact on the interplay of the participants. According to [3], the overall delay of a NMP perceived by the participating musicians can be divided into different stages: the delay introduced by the propagation of sound in physical space, AD/DA conversion, buffering and packaging on the sender/receiver side; the delay in data processing of the intermediate network nodes between the source and destination as well as the propagation delay over the physical transmission medium; and playout buffering which may be required to compensate the effects of jitter to achieve a sufficiently low packet loss rate.

While the round-trip delay of a NMP is easy to determine, it is more difficult to estimate the contributions of individual stages to the overall latency. In

[1]http://www.lips-project.de
[2]http://www.nyu-x.org/holodeck.html

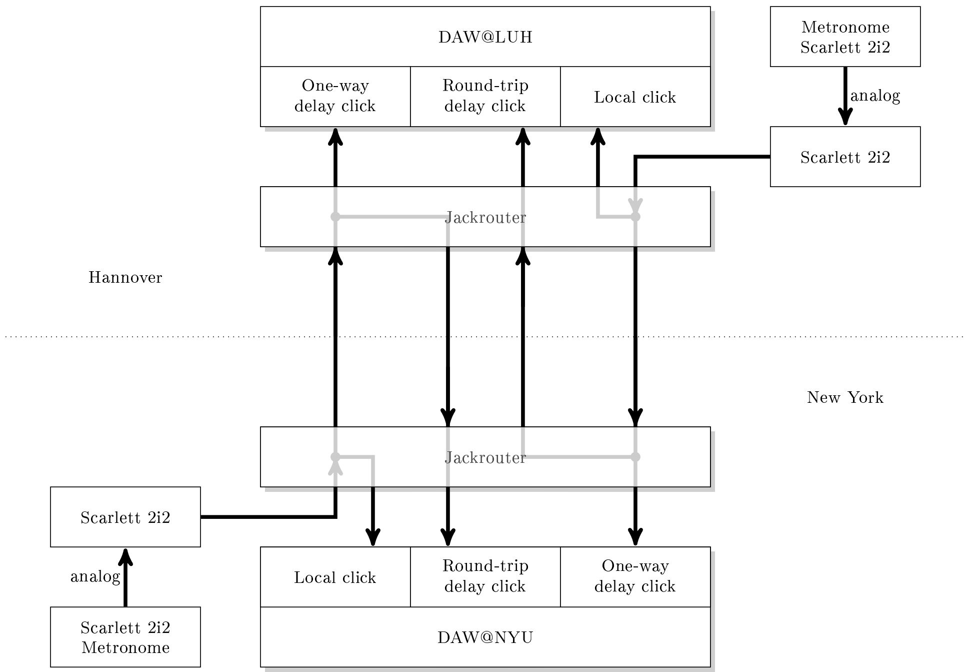


Fig. 1: Latency measurement setup for measurements between NYU and LUH.

this study, first measurements were performed in order to obtain information about latency as well as its fluctuations and asymmetries between the two nodes. For these tests, two global metronome units [4] implemented with Raspberry Pi devices were used. The global metronome uses the Global Positioning System (GPS) to accurately synchronize the system time of several devices independent of their geographic locations [5]. This globally-shared synchronized time can be used to trigger the generation of a highly synchronized audio click impulse signal in every connected device [3]. The click can in turn be used as a global conductor and assist with rhythmic synchronization on both sides of a distributed performance.

Although the metronome was mainly developed to enable psychoacoustic investigations in NMPs [6] it can also be used as a tool to determine delay times in the signal processing chain. Here we propose the use of this system to build a latency measurement method between geographically remote nodes of an NMP. By comparing the arrival time of a "remote" click signal to the "local" one, it is possible for each node to determine the latency of the incoming transmission since the clicks are generated simultaneously. Accordingly, a round-trip delay can be measured by a loopback of the click at the remote end which provide insights useful for some particular connections [7]. Additionally, jittery network conditions may create signal delay fluctuations, possibly asymmetrically, which is worth investigating. Such measurements are useful to NMP performers as an instrument of assessing the quality and performance implications of a connection, while also testing its network stability over time, transmission settings, and variations to the signal processing pipeline. Particular performance strategies [8] may thus be systematically considered by performers according to the measurement metrics found at the current conditions.

2 Measurements

As shown in Figure 1, the metronome click signal was fed into the system as an analog input on both sides of the network connection. Two audio interfaces per node were used, one as the audio card to translate the metronome numerical output into an analog signal, and another to convert and parse that signal into a computer (both ends used two Fireface Scarlett 2i2). Each node, connected to a metronome device, used a Digital Audio Workstation (DAW) to record the locally produced

[3]Please refer to [4] for the implementation details of the metronome

click, which serves as the time reference. The received one-way (remote) click and the round-trip click looped back from the receiver node to the sender node were also recorded in the DAW. *JackTrip* [9] was used to establish a connection between the two university nodes. *JackTrip* is a software tool developed at Stanford University that allows peer-to-peer bidirectional streaming of uncompressed audio data through IP-based networks. At each peer, the connection to local audio hardware is configurable via *JackPilot*, where buffer sizes, number of channels, and sample rate are set. The signal routing is handled by *JackRouter*.

It should be noted that occasional issues were found during the metronome setup at the New York node. The received GPS signal was found to be imprecise, possibly due to excessive signal reflections from nearby buildings to the antenna, making it difficult to synchronize the metronome with the GPS signal. This was solved by moving the system to a different location. However, it highlights possible system instability in urban environments. For the type of analysis that we wished to perform, it was necessary to centralize the control of the metronomes' starting time through Secure Shell (SSH). This step facilitated the appropriate latency measure of the received signals as the clicks could be more easily matched to their respective reference click. The measurement protocol used is detailed below:

1. Ensure Raspberry Pi is receiving the GPS and PPS signals. Connect to a network to enable remote configuration via SSH.
2. Change *JackPilot* settings to the desired interface device, buffer size, sample rate and number of channels.
3. Establish a *JackTrip* connection between the two locations.
4. Define the routing scheme to create one-way and round-trip connections through *JackRouter* and pass the signals to the DAW.
5. Begin recording the tracks in the DAW, and start both metronomes at the same time via SSH for a later synchronization of the recordings.

During the tests, the I/O buffer size was varied between 1024, 512, 256, 128, 64, 32 samples. Tests at different sample rates (48k and 96k) and increasing number of virtual channels (4 to 16) were also conducted.

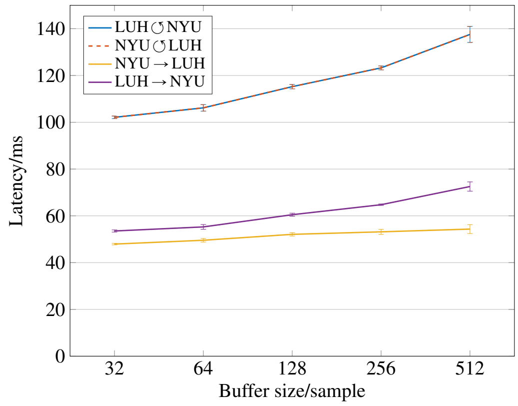
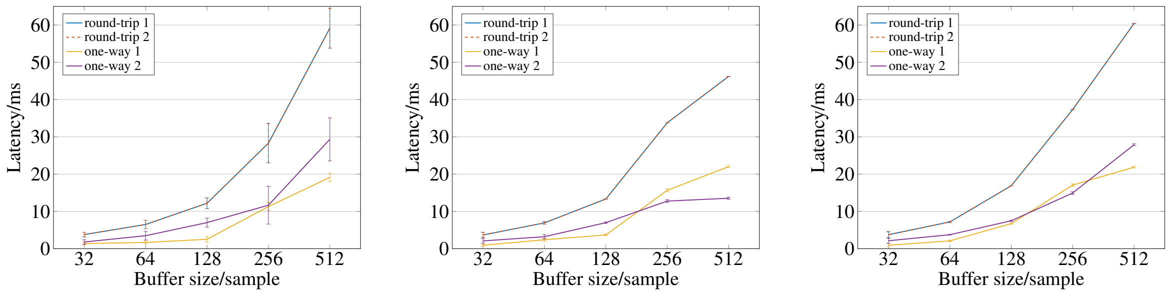


Fig. 2: Measured round-trip latencies (LUH ↻ NYU, NYU ↻ LUH) and one-way latencies (NYU → LUH, LUH → NYU) for different buffer sizes using Jacktrip, the global metronome and two Focusrite Scarlett 2i2.

In addition to the latency measurements across the Atlantic Ocean, the same measurements were conducted locally at LUH, by having the whole measurement chain in the same room. The setup and procedure was left unchanged, the only difference being that the two peer machines were connected on a local area network with only one network switch between them. Thus, the network latency and jitter was reduced to a minimum and the contributions of other components in the measurement chain could be investigated in more detail. Furthermore, these insights reveal the separate contributions of the equipment and the network to the overall latency. Two different interfaces, the Focusrite Scarlett 2i2 (identical to the remote measurement) and the RME MADIface XT, were compared in order to assess the impact of different hardware on the setup. Since this local setup hosted all the equipment in the same room, the RME MADIface interfaces allowed the option to be synchronized by the same word clock signal (WCK). To inspect the effect of synchronized clocks, data was collected with and without this option enabled.

3 Results

All measurements were evaluated by comparing the locally produced click signal to the received one-way



(a) Focusrite Scarlett 2i2. **(b)** RME MADIface XT /wo WCK. **(c)** RME MADIface XT /w WCK.

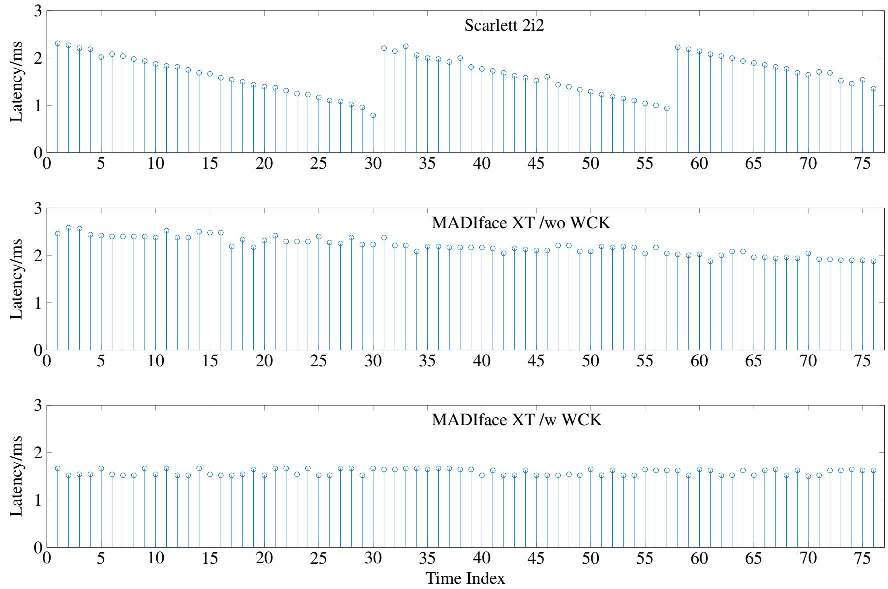
Fig. 3: Local network latency measurements for different buffer sizes and three different cases of interface setup.

click signal and the round-trip click signal. Therefore, the latency was calculated for each beat of the corresponding click signal separately. The results were averaged over three measurement sessions taken on the same day to consider temporal changes in the network latency. This section only shows the results obtained from changing the buffer size setting, and omits those obtained for different sample rates and number of channels as those elements did not show any significant impact.

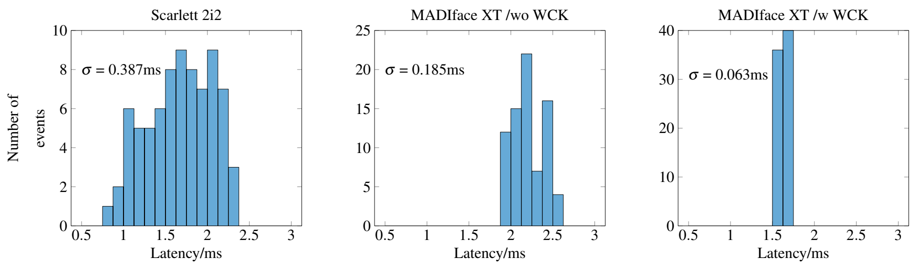
Figure 2 shows the results of the remote measurement between LUH and NYU. The mean of round-trip latencies (red and blue dashed line) and the corresponding one-way latency (violet and yellow line) at each node, as well as the standard deviations are shown. As expected, the latency increases with larger buffer sizes as this means more samples are to be collected for transmission. Moreover, the variance of the measurements also increases with higher buffer sizes. A closer look at the results shows that the variance at each different buffer size is mainly due to latency jumps in multiples of the temporal buffer size. Consistent results were obtained for the round-trip latencies at both ends, while the one-way latencies reveal an asymmetry for the two measurement paths where one direction of transmission seems slower than the other one.

For the local measurements conducted at LUH (same setup but with all end-nodes in the same room), the same interface as the remote measurement case (Focusrite Scarlett 2i2, non-synchronized) was measured for comparison in low network latency conditions. Compared to the remote case, the latency variances for the local measurement showed a higher range, proportionally increasing in accordance to the buffer size (Figure 3a). To examine the hardware effect more closely, the same measurement was conducted with a professional-grade interface: the RME MADIface XT (Figure 3b). It can be seen that the variance is observed to be much smaller and has almost no influence on the performed delay measurement. Possible reasons for the discrepancy between the interfaces may be linked to device-related factors such as the performance of audio drivers or current CPU-load or even JackTrip-specific behavior. Finally, we also inspected the case where the internal sampling clock for the audio interfaces was synchronized via word clock signals (WCK). Even if it is not possible to use this configuration during remote measurements, the results in Figure 3c imply that the use of WCK would only bring very small improvement in latency measurement.

In order to assess the effects of the WCK on clock-drift and stability, a continuous one-way latency measurement over a time period of 50 s at a buffer size of 64 samples is shown in Figure 4 for all three examined setups. Figure 4a depicts the offset of two corresponding clicks of one take for a local one-way latency measurement. It can be seen that the non-synchronized Focusrite Scarlett 2i2 show a significant clock drift with a standard deviation of $\sigma = 0.387$ ms. This leads to a circular signal alignment error, causing the periodical sawtooth-patterned change in the latencies observed. Whereas, the RME MADIface shows a much slower rate of drift ($\sigma = 0.185$ ms). As expected, the latency variance pattern caused by the clock drift is stabilized when using the WCK ($\sigma = 0.063$ ms). These results indicate that the clock drift of the audio interfaces has only a small influence on the accuracy of the measurement method. It must be pointed out that in Figure 4



(a) One-way latency over a time period of 50 s.



(b) Histogram of one-way latencies with σ, the standard deviation of each.

Fig. 4: One-way latencies in local measurement scenario over a time period of 50 s at a buffer size of 64 samples for different interface setups.

only the time period of 50 s of a single measurement was considered which results in low variances. In Figure 3 the mean and variance was averaged over several independent measurements, resulting in a variance in the range of several milliseconds. This is due to the fact that some takes had deviations in the mean values for the same buffer size. Nevertheless, these inaccuracies can be reduced by the choice of the audio interface. Our results show that the signal latency depends on the performance of the chosen interface equipment and the stability of the audio signal transmission path.

4 Discussion and Future Work

The GPS-based metronome is primarily intended to serve as a rhythmic synchronization tool, or performance conductor. Nevertheless, the presented method is shown to be suitable for the measurement of one-way and round-trip delay times in NMPs. This is proved by the fact that the local measurements were able to capture the influence of using different interfaces, and show that latency variance is more linked to device-related factors (audio driver, CPU load or Jacktrip behaviour) than the metronome unit.

It can be summarized that the choice of the audio interface has quite an influence on the measurements. The choice of the Focusrite Scarlett 2i2 was not optimal, but for the first investigation, due to the large local separation between LUH and NYU and the importance of conducting tests with identical interfaces at the two ends, it was not possible to realize it in another way. Professional-grade interfaces are more likely to produce less variability in the results. However, the use of WCK synchronization only brings negligible improvements and is not a key factor to the quality of the measurement.

Performers using distributed networks may find this method useful to learn more about their connection and coarsely dissect some of the delay stages in their pipeline, informing the choice of musical strategies or appropriate equipment. If asymmetric conditions are found in the one-way delays, further inspection of the transmission components may reveal whether the asymmetry is grounded in the network conditions or the processing components.

Even if the use of the WCK has not resulted in a significant improvement of our latency measurements, the time-synchronous sampling of audio devices at separate locations could provide great advantages for future professional NMPs. New platforms and possibilities could be created, especially with regard to the ongoing technological convergence of production and distribution networks and the availability of new mobile radio and network technologies.

5 Acknowledgements

The project LIPS-CISS is co-funded by the German Federal Ministry for Economic Affairs and Energy under the reference number 01MD18010G.

References

[1] Chafe, C., Gurevich, M., Leslie, G., and Tyan, S., "Effect of Time Delay on Ensemble Accuracy," *Proceedings of the International Symposium on Musical Acoustics*, 39, 2004.

[2] Farner, S., Solvang, A., Asbjørn, S., and Svensson, U. P., "Ensemble Hand-clapping experiments under the influence of delay and various acoustic environments," *AES: Journal of the Audio Engineering Society*, 57(12), 2009.

[3] Rottondi, C., Chafe, C., Allocchio, C., and Sarti, A., "An Overview on Networked Music Performance Technologies," *IEEE Access*, 4, 2016.

[4] Hupke, R., Beyer, L., Nophut, M., Preihs, S., and Peissig, J., "A Rhythmic Synchronization Service for Music Performances over Distributed Networks," in *Fortschritte der Akustik : DAGA 2019, 45. Jahrestagung für Akustik, Rostock*, 2019.

[5] Oda, R. and Fiebrink, R., "The Global Metronome: Absolute Tempo Sync For Networked Musical Performance," *Proceedings of the International Conference on New Interfaces for Musical Expression*, 16, 2016.

[6] Hupke, R., Beyer, L., Nophut, M., Preihs, S., and Peissig, J., "Effect of a Global Metronome on Ensemble Accuracy in Networked Music Performance," in *Audio Engineering Society Convention 147*, 2019.

[7] Boren, B. and Genovese, A., "Acoustics of virtually coupled performance spaces," Georgia Institute of Technology, 2018.

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

[9] Cáceres, J.-P. and Chafe, C., "JackTrip: Under the Hood of an Engine for Network Audio," in *Proceedings of International Computer Music Conference*, p. 509–512, International Computer Music Association, International Computer Music Association, San Francisco, California, 2009.