A Rhythmic Synchronization Service for Music Performances over Distributed Networks

Robert Hupke[1], Lukas Beyer[1], Marcel Nophut[1], Stephan Preihs[1], Jürgen Peissig[1]
[1] *Institut für Kommunikationstechnik, Leibniz Universität Hannover, Deutschland*
Email: hupke@ikt.uni-hannover.de

Introduction

Within the project "Live Interactive PMSE Services" (LIPS)[1] the Institute of Communication Technology (IKT) is investigating novel interactive concepts for the immersive audiovisual linking of live events at different geographic locations. The project has the aim to develop solutions to improve access and participation of rural areas in the creative, cultural and political operation of metropolitan areas. Therefore, a low-latency connection of the utilized audio systems (e.g. microphones, in-ear monitors, sound reinforcement systems) and video systems (e.g. displays, projectors, cameras, spotlights, effects) is required for live interaction. The underlying scenario of the project is depicted in Figure 1. In order to assist the interaction of musicians at different geographic locations, suitable technological services for musical synchronization are developed and analyzed at the IKT.

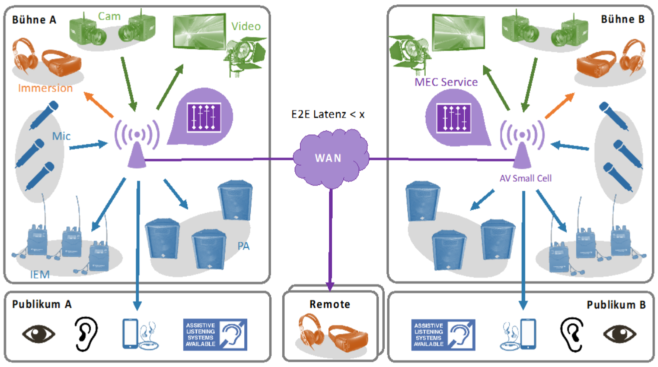


Figure 1: Basic technical concept of the project "Live Interactive PMSE Services" (LIPS).

A so-called Networked Music Performance (NMP) is based on audio data transmission over IP-based networks and always exhibits a certain latency, latency fluctuation, a limited bandwidth and packet losses which can have an impact on the synchronicity of the participants [1]. One of the biggest challenges when playing music together over Wide Area Network (WAN) connections is latency. Different approaches to evaluate musical synchronization over WAN were evaluated in the last decade. Experiments focusing on rhythmic patterns performed by handclapping over various fixed network delays (0 ms to 77 ms) show that delays longer than 11.5 ms cause a deceleration of the tempo, while for shorter delays, musicians tend to speed up their tempo [2, 3, 4]. By the use of a distributed metronome and the possibility of an adjustable user delay (Delayed Auditory Feedback), other investigations show that musicians were able to synchronize in the range of several seconds, depending on the structural duration of the evaluated musical piece [5]. By creating a metronome pulse as a feedback signal which is adjusted dynamically to the network delay, a music performance at an average uni-directional latency of 110 ms has been achieved for NMPs [6].

Our synchronization service is based on the approach to use the Global Positioning System (GPS) to generate a globally shared synchronized time signal, which can be used as a global conductor within a NMP [7, 8]. Although our global conductor was mainly developed to enable future psychoacoustic investigations in NMPs, it can be also a useful tool for determining delay times in the signal processing chain between different geographic locations. The goal of our future psychoacoustic experiments will be to investigate whether the so called global metronome shifts the limits of the defined playable delay times, as defined in [3], to higher ones while stabilizing the rhythmical interaction. The focus of this publication is the base technology of these investigations and concentrates on the technical evaluation in terms of time synchronization of the implemented metronomes. Furthermore, a feature to change the tempo of several metronomes simultaneously was rudimentary implemented and evaluated. Finally, an outlook on the procedure for future psychoacoustic investigations within the project LIPS is given.

Approach

The implemented metronome ticks with the tempo of $\mathrm{f_{bpm}}$ beats per minute, which is actually a frequency referred to a time interval of one minute. This can be represented as a rotating phasor which rotates $2\,\pi$ radians once every bar. This results in $\mathrm{n_{beat}}$ metronome click for every bar distributed on the unit circle as shown in Figure 2. The phase angle $\phi = 0$ marks the beginning of each new bar. Under the assumption that the metronome has been ticking since the beginning of the Unix time $t_{0,Unix}$ with an constant tempo $\mathrm{f_{bpm}}$, the time of the *k-th* metronome tick t_k can be calculated $\forall k \in \mathbb{N}$ as

$$t_k = t_{0,Unix} + (k-1)\,\frac{60}{f_{bpm}}. \qquad (1)$$

In order to calculate the first tick of the *k-th* bar, which is the segment of time corresponding to the specific num-

---

[1] http://www.lips-project.de

ber of n_{beat} clicks, Eq. (1) has to be multiplied by the factor n_{beat}. So the beginning of every *k-th* bar can be calculated $\forall k \in \mathbb{N}$ as

$$t_{k,bar} = t_{0,Unix} + (k-1)\,\frac{60}{f_{bpm}} \cdot n_{beat}. \qquad (2)$$

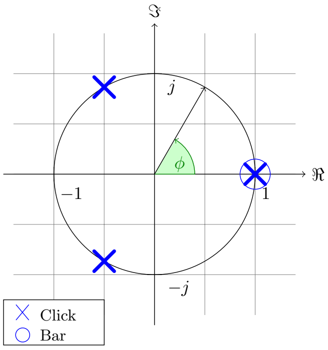


Figure 2: A phasor which rotates by 2π once every bar represents the metronome click. The clicks are distributed on the unit circle. Example for $\mathrm{n_{beat}} = 3$: clicks are located at $\phi = 0°, 120°, 240°$.

Implementation

A Raspberry Pi (RPi) was used to implement the metronome as described in [7]. It is inexpensive and provides many possibilities to connect additional hardware with its freely usable GPIO pins. The basic concept of the global metronome is to use the GPS signal to synchronize the system time of the devices accurately. This allows the previously described method of calculating the next click on every device. The standard RPi is not able to evaluate GPS information. Therefore a GPS expansion board and an active GPS antenna were used. Since the digital-to-analog audio converter of the Raspberry is unreliable and several tested methods for generating an analog audio signal did not produce a reliable, jitter-free audio output, an additional audio interface (Focusrite Scarlett 2i2) was connected via USB for the audio generation of the click. Figure 3 shows the measurement setup for evaluating the temporal synchronicity of two devices.

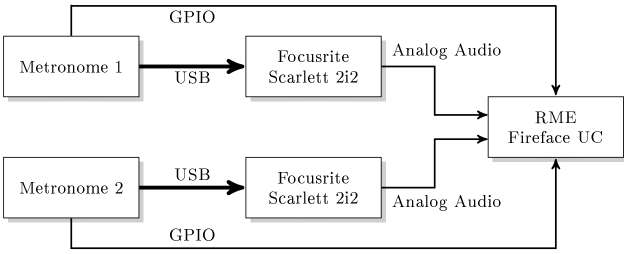


Figure 3: Setup of the temporal synchronization measurement for two global metronomes. GPIO-pins are directly connected to a RME Fireface. For the generation of the analog audio click signal two additional audio interfaces (Focusrite Scarlett 2i2) were connected via USB.

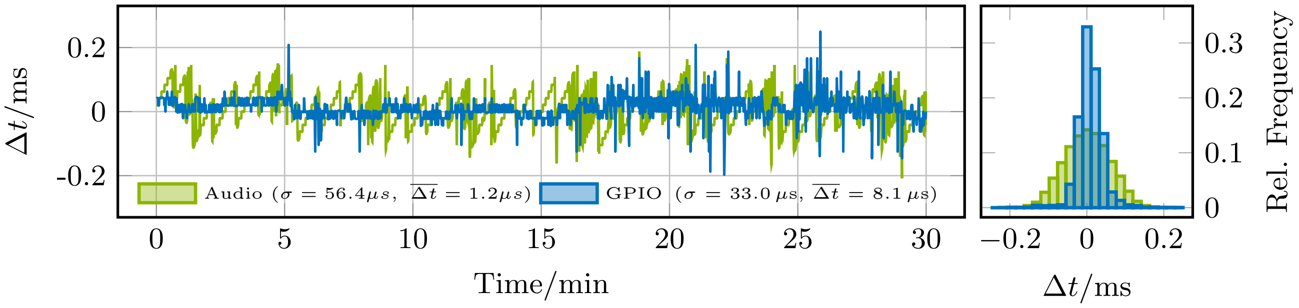
Evaluation

Three different configurations of the implemented metronome were investigated. Therefore, similar configurations were chosen as described in [7] in order to ensure comparability. Although time synchronization via GPS is theoretically in the range of nanoseconds, the adjustment of the system time, the calculation of the click signal, as well as the generation of the audio click signal results in a deviation of the temporal synchronicity. With the goal to estimate the time accuracy for each configuration of the metronome, a beat with a frequency of $\mathrm{f_{beat}} = 120\,\mathrm{bpm}$ was generated on both devices and recorded for 30 min with a sampling frequency of $\mathrm{fs} = 96\,\mathrm{kHz}$. The time differences of the corresponding clicks for both metronome were calculated by dividing the measured signals into parts of length $\mathrm{l} = 48\,000$ samples, which corresponds to the time range between two successive clicks. Afterwards, the time difference was computed for each click using a cross-correlation of both signals. The generated audio output as well as the output of one GPIO-pin were measured. The output of the GPIO-pin was measured to evaluate the smallest possible deviation between two synchronized devices. Thus, the GPIO-pin was switched from low to high for 10 ms and back to low for each click. In comparison to the audio output signal this allows to estimate the time which is needed for audio processing. Under the assumption that the GPS-receivers have a location-independent synchronization accuracy, both devices were located at the same place for all measurements. The three different evaluated settings were:

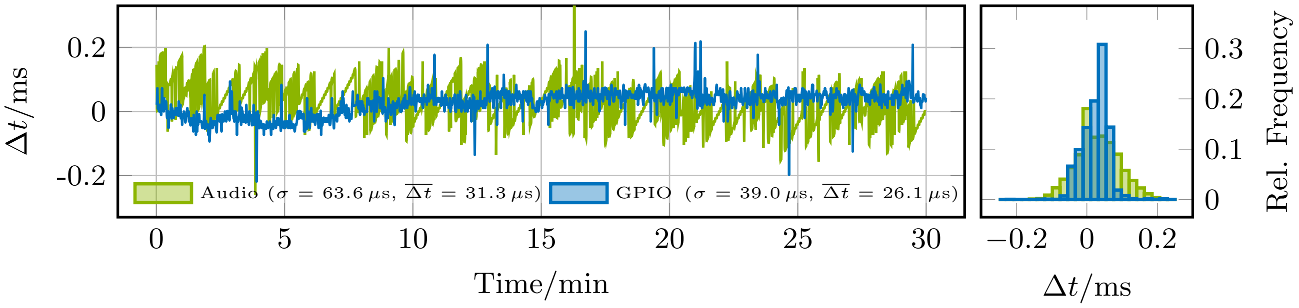
1. Both devices synchronized via GPS,
2. Both devices synchronized via a NTP time server,
3. Both devices not synchronized with each other generating the metronome signal by their internal clock.

Results

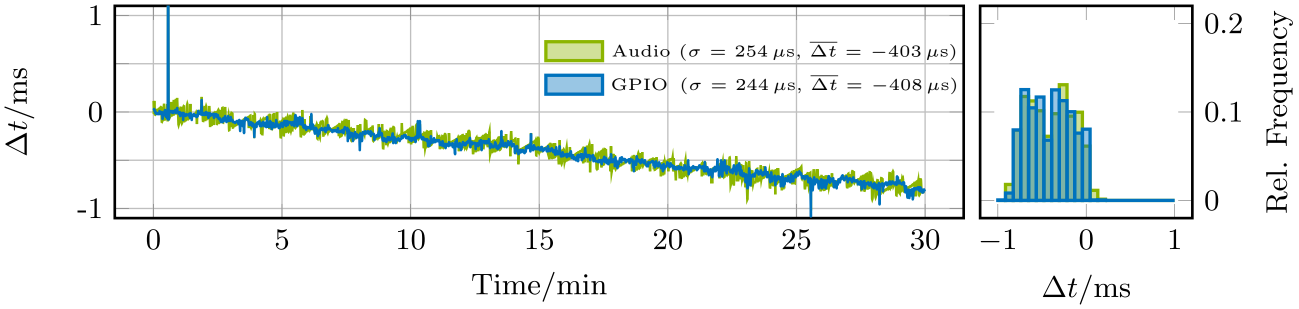
Figure 4 shows the results of the time difference measurement for all listed configurations. The measurements of the two GPS-synchronized devices are depicted in Figure 4a. In direct comparison, the audio click has a slightly higher deviation in form of a recurring saw-tooth with an standard deviation of $\sigma = 56.4\,\mu\mathrm{s}$ which is the result of the audio click generation since the GPIO output shows no saw-tooth. The GPIO measurement via GPS synchronization achieves the most accurate standard deviation of $\sigma = 33\,\mu\mathrm{s}$. Due to the fact that both devices had permanently good coverage of at least three satellites during the measurement, no spikes occur in both GPS measurements. Thus, the signal generation of the audio click increases the standard deviation by $23\,\mu\mathrm{s}$. In comparison, the results of the NTP synchronization setting in Figure 4b show higher time deviations in form of occurring spikes in a range less than 0.5 ms. The same recurring saw-tooth occurs for the audio click generation. Nevertheless, the standard deviation for the NTP setting is nearly as low as for the GPS setting with a standard



(a) Synchronization via GPS.



(b) Synchronization via NTP.



(c) No synchronization.

Figure 4: Measurement results for investigating the temporal synchronicity of two global metronomes which were synchronized via GPS (a), NTP (b) or not at all (c). Measurements were taken over 30 min with a generated tempo of $f_{\mathrm{beat}} = 120\,\mathrm{Hz}$. The generated audio click signal and a square wave signal generated at the GPIO-pin were measured. The left figures show the course of the time difference Δt between the corresponding metronome clicks generated by both devices. The right figures show the corresponding relative frequency distribution. Note the different axis scaling in graphic (c).

deviation of $\sigma = 63.6\,\mu\mathrm{s}$ for the audio click generation and $\sigma = 39\,\mu\mathrm{s}$ for the GPIO measurement. Furthermore, the results of the NTP-measurement show an average offset of $\overline{\Delta t} = 31.3\,\mu\mathrm{s}$ for the audio click and $\overline{\Delta t} = 26.1\,\mu\mathrm{s}$ for the GPIO-pin. The offset of the NTP measurement changes over time and is therefore not the same for both GPIO and audio measurements. Since both measurements were not conducted simultaneously, this illustrates that the synchronization accuracy of NTP-setting highly depends on network route asymmetry, network latency and the accuracy of the time server. For our measurement we used a nearby time server of the Leibniz Universität Hannover, which enables almost ideal network conditions. Nevertheless, high accuracy can be stated for both GPS and NTP measurements. The “No-Sync” measurement in Figure 4c is intended to illustrate that the click generation by the internal clock of the devices can not be used for the time synchronization since both measurements diverge with time. The “No-Sync” setting in Figure 4c shows an offset of 403 μs which will increase over time.

Tempo Change

A feature to change the tempo on several metronomes was rudimentary implemented. As recommended in [7] a tempo change has to be divided into equal steps of tempo change. The step size has to be small enough for a perceptual smooth tempo change. This results in an recommendation of 1 bpm for each tempo step, as it is below the 2 % just noticeable difference for tempos higher than 50 bpm. In our implementation the step size as well as the new target tempo can be manually modified. Figure 5 depicts the tempo of two devices during a tempo change. After 8 s the tempo of both metronome clicks were simultaneously set from 120 bpm to 70 bpm. The tempo

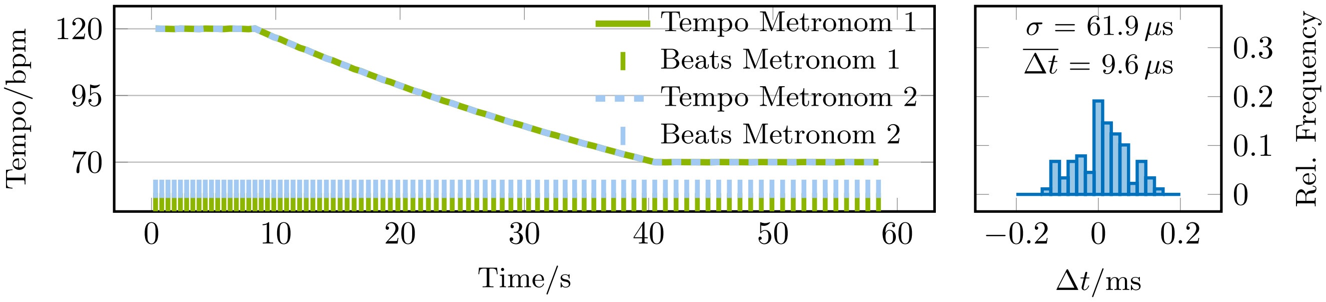


Figure 5: Tempo change of two metronomes. After 8 s the tempo of the metronome clicks were set from 120 bpm to 70 bpm. The tempo was reduced by 1 bpm with each metronome click until after 40 s the target tempo was reached. The histogram shows the time difference between the corresponding clicks produced by both metronomes during the tempo change.

is reduced by 1 bpm with each metronome click until the target tempo is reached after 40 s. The histogram shows the time difference between the corresponding clicks produced by both metronomes during the tempo change. The standard deviation of $\sigma = 61.9\,\mu$s during the tempo change is slightly higher than in our measured GPS synchronization setting. Nevertheless, it is small enough for a temporal synchronization and shows that if both devices receive the tempo change command at the same time, a common tempo change is possible.

Conclusion and Future Work

Our measurements show that by using the GPS signal as a global time reference for our acoustical synchronization service, extremely accurate time synchronization and click generation is possible with our implemented metronomes. Since our experiment was conducted in the same geographical environment, we assume that the GPS receivers that are far apart are expected to have the same synchronization accuracy as those that are side by side [7]. Nevertheless, this has to be investigated in further experiments. Even if the audio click generation adds a time deviation of 23.4 μs to the system, the standard deviation of the generated audio click signal is $\sigma = 56.4\,\mu$s for two GPS-synchronized metronomes. This is sufficiently small for the acoustical synchronization of musicians. However, the synchronization via Precision Time Protocol (PTP) has to be implemented and evaluated in future research. Furthermore, a feature to change the tempo of the metronome was presented and it was shown that it works simultaneously on two devices. This feature is still in need of further development. A direct communication between the two devices is absolutely necessary and an adaptive tempo adjustment is also conceivable. In future work, we want to investigate whether the global metronome can increase the previously defined delay times [3] for NMP. For this, comprehensive subjective experiments with 15 volunteers has been already conducted. Individual pairs of subjects had to play a predefined rhythm at geographical separated places from each other. The audio transmission between the subjects was artificially delayed by different delay times. The experiment was performed with the help of the global metronome and without any assistance for the subjects. First results show that a tempo stabilization by the use of the metronome is quite possible. However, it also seems as if the metronome has a counter-productive effect on the musical interplay above a certain transmission delay. This has to be investigated in further research.

Acknowledgments

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

References

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

[2] Chris Chafe, Michael Gurevich, Grace Leslie, and Sean Tyan, “Effect of time delay on ensemble accuracy,” *Proceedings of the International Symposium on Musical Acoustics*, vol. 39, 2004.

[3] Chris Chafe, Juan-Pablo Cáceres, and Michael Gurevich, “Effect of temporal separation on synchronization in rhythmic performance,” *Perception*, vol. 39, no. 7, pp. 982–992, jan 2010.

[4] Chris Chafe, “Living with net lag,” in *Audio Engineering Society Conference: 43rd International Conference: Audio for Wirelessly Networked Personal Devices*, Sep 2011.

[5] Nicolas Bouillot, “njam user experiments: Enabling remote musical interaction from milliseconds to seconds,” in *NIME*. 2007, pp. 142–147, nime.org.

[6] *To the Edge with China: Explorations in Network Performance*, Porto, Portugal, 2008.

[7] Reid Oda and Rebecca Fiebrink, “The Global Metronome: Absolute Tempo Sync For Networked Musical Performance,” *Proceedings of the International Conference on New Interfaces for Musical Expression*, vol. 16, pp. 26–31, 2016.

[8] Reid Kei Oda, “Tools and techniques for rhythmic synchronization in networked musical performance,” 2017.