Using Metrically-entrained Tapping to Align Mobile phone sensor measurements from In-person and Livestream Concert Attendees

Finn Upham (a)

(a) RITMO, University of Oslo, Norway, finn.upham@imv.uio.no

Music is often made and enjoyed in large groups, but simultaneously capturing measurements from dozens or hundreds of people is technically difficult. When measurements are not constrained to wired or continuous connected wireless systems, we can record much bigger groups, potentially taking advantage of the wearable sensors in our phones, watches, and more dedicated devices. However, aligning measurements captured by independent devices is not always possible, particularly to a precision relevant for music research. Phone clocks differ and update sporadically, wearable device clocks drift, and for online broadcast performances, exposure times can vary by tens of seconds across the remote audience. Many measurement devices that are not open to digital synchronisation triggers still include accelerometers; with a suitable protocol, participant movement can be used to imbed synchronisation cues in accelerometry measurements for alignment regardless of clock times. In this paper, we present a tapping synchronisation protocol that has been used to align measurements from phones worn by audience members and a variety sensors worn by a symphony orchestra. Alignment with the embedded cues demonstrate the necessity of such a protocol, correcting offsets of more than 700 ms for devices supposedly initialised with the same computer clock, and over 10 s for online audience participants. Audience tapping performance improved cell phone measurement alignment to a median of 100 ms offset, and professional musicians tappings improved alignment precision to around 40 ms. While the temporal precision achieved with entrained tapping is not quite good enough for some types of analyses, this improvement over uncorrected measurements opens a new range of group coordination measurement and analysis options.

Keywords: Concert research, Audience research, Technology, Motion

1. Introduction

Large group musical performances to large audiences are extreme examples of hundreds and thousands of humans feeling and acting in tight temporal coordination. However, simultaneous measurement of experiences of groups this size is made difficult by a wide range of technical challenges. Wearable wireless sensors are common place and yet these devices rarely allow for continuous simultaneous signal capture. Instead, these devices rely on their own clocks to track sample times, and these clocks can differ in their alignment to external clocks (like satellite time) and tendency to drift. In order to make use of such devices' measurements during music performances, some kind of synchronisation trigger is to accurately alignment measurements with the music performed by compensating for clock differences.

Besides the drift between devices in a performance space, we are also seeing many more music performances broadcast online, allowing audiences of thousands to share a performance in near synchrony without traveling to the same place. Measurement through mobile apps offer an exciting opportunity to capture these geographically distributed experiences, if again there is a way to align their exposures to the shared broadcast signal. When privacy issues and transmission delays making existing synchronisation strategies untenable, entrained participant actions may be enough to embed a usable synchronisation trigger in a shareable signal.

This paper describes a tapping synchronisation protocol and its impact on inter-device temporal alignment in two concert experiments. At each performance experiments, a sequence of beeps was played to participants who were instructed to tap along on the sensor devices they wore. The timing of these taps in individual device time was used to assess the quality of initial alignment and compensate to bring the sensor measurements to the highest alignment possible, under the circumstances.

2. Methods

Two experiments used a similar entrained tapping protocol to embed a detectable synchronisation cue in the accelerometer measurements of independent sensor systems worn by participants.

2.1.Participants

In the first experiment (Copenhagen), the participants were audience members attending a string quartet concert. Voluntary participants in the concert hall (84) and watching remotely (24) work sensors on the chests to track body sway, most using their own mobile phones in a special holder. Most were experience classical music concert goers with an average age of 52.

In the second experiment (Stavanger), members of a professional symphony orchestra (55) wore sensor vests during their dress rehearsal and five performances of a children's concert. A subset of string players (13) also had small accelerometers on their bowing arm.

2.2. Sensors

In the Copenhagen concert, most accelerometer measurements were collected via the MusicLab app (Høffding, 2021). With participants permission, this app sent internal phone sensor measurements, including accelerometry, in one minute increments to dedicated servers at the University of Oslo. Timestamps on these devices are set according to the mobile phones clock, each updated according to

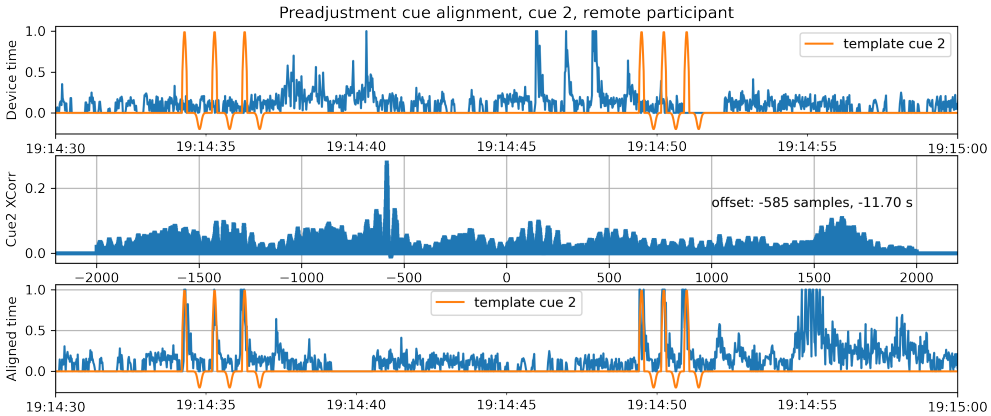

Figure 1. Initial and corrected alignment for remote participants motion recordings with second synchronization cue. Top: Remote participant motion signal in device time and Cue Template in concert time. Middle: Cross-correlation between signal excerpt and cue template. Bottom: Remote participant motion signal aligned to concert time with Cue Template.

individual device settings. Most devices recorded this signal at around 60 Hz, with the occasional gap between increments.

In the Stavanger experiments, participating orchestra members wore Equivital Monitor vests on stage. These record a suite of physiological signals along accelerometry for body sway at 256 Hz. The sensors on bowing arms were Axivity AX3 sensors set to log motion at 200 Hz. Both of these devices are supposed to align their internal clocks to that of the computer initializing a recording session.

2.3. Procedure

The procedure to embed the tapping signals and assess its location in accelerometer data was much the same for the two experiments.

2.3.1. Synch cue

Entrainment to an isochronous beat is very quick at normal tempi. Two or three beats are all that is needed for a human to begin adapting their movement to the predictable sound. Tapping accuracy to an isochronous beat also tends to improve after the first tap. The entrained tapping task was designed to be as short as possible while allowing for some of these advantages to improve the alignment between heard cue and participant action.

For each synchronisation cue, participants were played two sets of six isochronous beeps, first slow then faster. They were instructed to listen to the first three and tap along to the last three for each tempo. Thus the synch cue embeded in their device's accelerometer measures should be three taps, pause, then three slightly faster taps. The Stavanger orchestra heard midi woodblock hits at 72 and 80 BPM. Copenhagen participants heard 440 Hz sine tone beep with a manually shaped amplitude envelope of 5 ms rise, 10 ms sustain, and 55 ms decay in imitation of a percussive sound profile, first at 60 Hz then at 80 hz.

The synchronisation cue was played to the Copenhagen audience participants twice: before each half of the concert. Despite the novelty of the task and sensor performance issues, most participants' recordings contained both cues.

The Stavanger orchestra heard and performed the cue during their dress rehearsal and before each of 5 concerts set over 3 days. On the first two concert days, this short concert program was performed twice within the same recording interval. Additionally, Clapping music by Steve Reich was on the program, with everyone on stage clapping together for the first cycle of the piece. These clapping sequences offered an alternative synchronisation reference to check on alignment quality. Most recordings on double concert days include four potential cues of alignment: First concert synchronisation cue, first concert's onset of Clapping music about 10 minutes later, Second concert's synchronisation cue an hour later, and then the start second clapping music performance.

2.3.2. Cue detection

For each recording, excerpts of 3D acceleration values were taken around the times synchronisation cues were expected to appear. These resampled (50 Hz Copenhagen, 100 Hz Stavanger), reduced to the magnitude of jerk and cross-correlated with a template cue built to match the timing of taps according to the audio cue. Figure 1 shows the signal and template cue on a measurement from a remote participant of the Copenhagen concert and the offset identified by cross-correlation.

Variation is tapping performance sometimes complicated automatic assessment of offsets. Detected cue times in each recording were manually reviewed and corrected where necessary to a second or third highest peak in the cross correlation.

Synch cue times were recorded in original device times and then used to generate new timestamps in "concert time" using constant or linear alignment

Table 1. Inter-device clock variance at alternative cues per measurement condition before and after clock corrections: Constant (offset at first cue) and Linear (scaling for drift).

Participants	Sensor	Inital alignment				Corrected alignment			
		Assessment Cues	StD	Median |offset|	Max |offset|	Correction	STD	Median |offset|	Max |offset|
Audience, In hall	Mobile Phones	65	0.58 s	0.20 s	2.34 s	Constant	0.2 s	0.1 s	0.9 s
Audience, Online	Mobile Phones	14	12.8 s	12.42 s	36.12 s	Constant	1.6 s	0.71 s	4.3 s
Musicians, strings	AX3 on wrist	51	0.06 s	0.044 s	0.143 s	Constant	0.05 s	0.04 s	0.1 s
						Linear	0.040 s	0.03 s	0.09 s
Musicians, orchestra	Equivital vest	102	0.84 s	0.42 s	5.1 s	Constant	0.15 s	0.59 s	0.84 s
						Linear	0.044 s	0.032s	0.13 s

corrections to bing the synchronisation cues into focus across devices.

2.3.3. Alignment assessment

The characteristics of the clock variance across devices changes by sensors and conditions. This distribution is described with three statistics: Standard deviation of all cues times in device time relative to their respective average (aggregating across cues and recordings per condition), Median of the absolute offset on individual device time from each cues' average device time, and Maximum offset (early and late) from cue averages.

Variance in across devices at synch cues are assessed all measured synch cues. However, assessment of clock correct can only be measured in recordings with multiple cues. For the Copenhagen measurements, the second synch cues are used for clock correction assessment. (Though otherwise these measurements are readjusted to align at these times too). For the Stavanger measurements, scaling was performed with the outer most cues on days with two concerts, thus alignment quality was assessed the first Clapping music onset and the second Synch cue on double concert days.

3. Results

These tapping synchronisation cues first serve to mark the variation between device clock times in these different conditions. In the total of 145 synch cues in the 80 measurements in the concert hall at Copenhagen, the standard deviation across cue times of 2.44 s. The total 47 cues captured in remote audiences (33 recordings) were delayed by as much as 53.8 s, with variance metrics all over 10 seconds. Figure 2 shows the spread of mobile phone accelerometer measurements at the first synch cue of in initial device time and after correction, both in hall and remote participants.

For the devices that were supposed to already be synchronised in each recordings, clock differences were still substantial. In 89 tapping cues measured over 51 recordings, the standard deviation of cue times in AX3 accelerometers was 0.167 s, with at most 1.02 s offset from the average. And the Equivital vests varied by 0.77 s (standard deviation) across 349 tapping cues captured in 210 recordings. These results underline the importance of clock corrections in studies using measurements across independent of clocks.

The effect of correcting clocks with these entrained cues is reported in Table 1. Constant corrections in the mobile phone measurements produced a substantial reduction in variance at the assessment cues, both for in hall and remote audience participants. The large max offset after correction seems to be from non-linear clock shift, likely phones that had received an external clock correction during the concert. The substantial variance measured across remote participants after constant correction at the first synch cue is consistent with non-linear shifts in concert livestream delays on top of the occasional phone update.

The corrections on AX3 clocks during the Stavanger experiment are much more modest. These long-running dataloggers showed variable drift, with the time between device initalisation and experiment impacting the amount of disagreement between devices. The impact of linear correction was greater on later concert days.

The Equivital monitor vests worn by the orchestra had the most measurable clock drift, making linear corrections per recording essential for alignment to the recorded music and between devices.

4. Discussion

The entrained synchronisation cues embedded in the accelerometer measurements allows for significant improvement of alignment over device clock times, with the scale of improvement dependent on the devices and conditions. While these measurements are not practical for assessment of leader-follower roles and some other high precision comparisons of intra-performer timing, there are many signals that can still be looked at more closely with the incremental improvement of participant-participant and participant-music alignment quality. Figure 3 shows a 5 second snapshot of concurrent respiration measurements at the start of an orchestral work. As respiration cycles on periods of 2 to 6

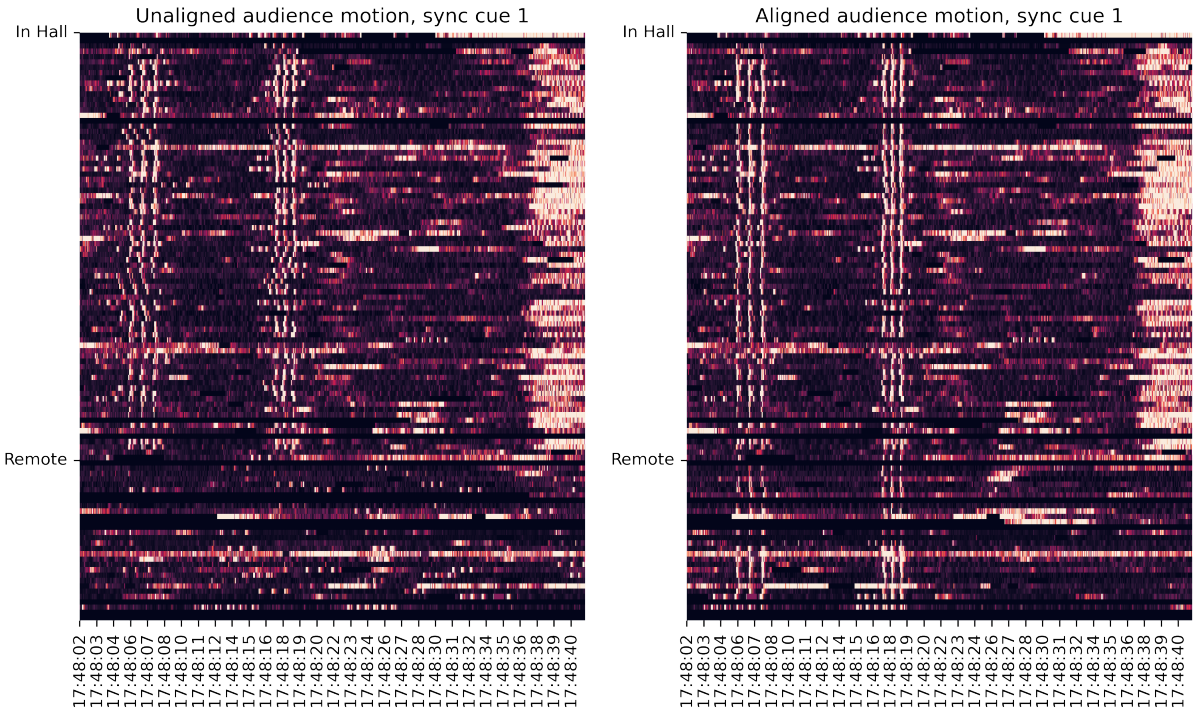

Figure 2. Normalized motion feature according to original device timestamps around the first synchronization cue (40 s), stacking each individual recording from in Hall and Remote participants with interpretable cues. Right: The same normalized motion in concert time, after realignment with the tapping cue.

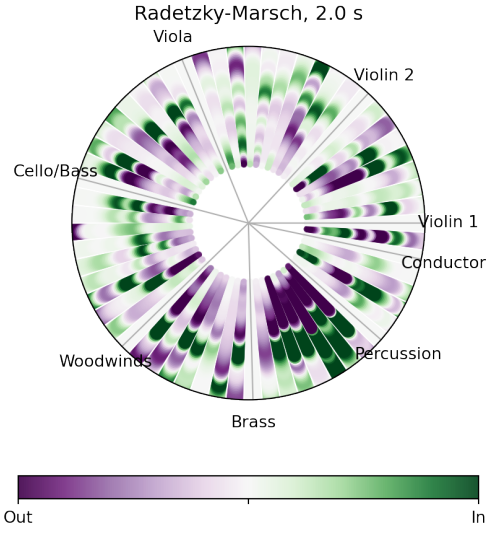

Figure 3. Equivital chest stretch measurements capturing performer respiration at the beginning of Strauss's Radetzky March (-3 s to 2s) after alignment using tapping protocal.

seconds (or longer during some activities) alignment precision of around 40 ms is enough to bring concurrent breaths into focus that would be more difficult to catch in spreads around 700 ms.

There are, however, a few important issues with this entrainment protocol.

4.1.Precision limites

There are two precision constraints on this synchronization strategy: measurement sample rate, and tapping quality. At a sample rates of 60 or 50 Hz on mobile phones, the precision of measurement timestamps is only to within 40 ms. The reliability of humans tapping to an isochronous beep sequence of 60 to 80 BMPs is not far from that range, with variation in inter-tap-intervals expected around 4% for non-musicians (Repp, 2005) on longer tapping sequences than used in this task. The precision of taps measured from the orchestra are more consistent, but not without variation. However, tapping reliability is complicated by the additional factor of negative mean asynchrony, anticipatory offsets of 50 ms common in non-musician who are tapping these rates (Repp & Su, 2013). The blur on alignment across the professional musicians may have a similar complication through more on the order of 10-20 ms, at the edge of measurable with alignments calculated at 100 Hz. And yet, even in experienced musicians, the first tap is often late relative to the remaining cues. Assessment over a longer interval reduces the impacts of these local

deviations, however there is no way to entirely remove the variability of human rhythm production.

4.2. Device and stream reliability

While collecting multiple synchronisation cues in accelerometer measurements make it possible to assess the clock corrections and use higher order adjustments, there are some forms of noise that can't be corrected for so easily. It is very lucky the apparent drift rate on mobile phones is usually closer to 60 ms an hour, as unpredictable clock updates make it necessary to treat each synchrony cue as a chance for constant updates. The additional variability in the remote audience measurements suggests after corrections suggest that a more synchronisation cue might be needed for reliable concert-long alignment.

4.3. Improving the task

The choice to use two tapping sequences at different tempi came from observing variation in tapping behaviours during a pilot experiment. Often participants start tapping after two beeps, or keep taping after six. Presenting sequences at different tempi allows for the time between sequences to be used for more definitive detection of when the audio cue was presented. Thus manual oversight or more elaborate automated alignment can compensate for many of the variations of the task performed in concert experiment conditions.

The shape of the tapping cue could be improved, both to yield more reliable taps from audience participants with a range of musical experience and to aid automated cue detection. Extending the entrainment and tapping sequences from three to four beeps/taps may be helpful for mechanical and cultural reasons. Many cultures have a strong bias towards 4/4 over 3/4 interpretations of isochronous beeps, and this could reduce instances of anticipatory and extraneous taps. Also the two tempi should be selected to reduce risk of cycle locking. By chance, spacing between the two sequences in the first cue of the Copenhagen resulted in avoidable ambiguities. The faster and more similar BPM used at the Stavanger concerts also resulted in some confusion, including some participants tapping at the same rate both times, though this behaviour subsided after the first performances.

Having an audience of humans act as a synchronization trigger across concurrently recording devices is not fool proof, even a set of professional musicians tap with variability, but the results of these experiments demonstrate that a substantial improvement of signal alignment is possible when leveraging metrical entrainment and participants' good will. The precision achieved with this technique is already supporting analysis of inter-participant coordination in these and concurrently sampled signals.

References

Anglada-Tort, M., Harrison, P., & Jacoby, N. (2022.) REPP: A robust cross-platform solution for online sensorimotor synchronization experiments. *Behavior research methods,* 1–15.

Høffding, S., Rebecca, J. F.; Bishop, L.; Bravo P. L., Burnim K., Cancino-Chacón, C., Clim, A., Good, M. Hansen, N. C., Karlsen, E. S., Laeng B., Lartillot, O., Lippert, E., Martin, R., Nielsen, N., Nørgaard, A., Omprakash, R., Rosas, F., Sjölin, F., Swarbrick, D., Sørby, S., Sørensen, R. T., Upham, F., Vrasdonk, A.; Vuoskoski, J., Yi, W., Øland, F., & Jensenius A. F. 2021. *MusicLab Copenhagen Dataset.* Open Science Framework. https://osf.io/v9wa4/

Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychonomic bulletin & review,* 12 (6), 969–992.

Repp, B. H. and Su, Y. H. (2013) Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonomic bulletin & review* 20 (3), 403–452.