REAL-TIME MUSIC TRACKING USING MULTIPLE PERFORMANCES AS A REFERENCE

Andreas Arzt, Gerhard Widmer
Department of Computational Perception, Johannes Kepler University, Linz, Austria
Austrian Research Institute for Artificial Intelligence (OFAI), Vienna, Austria
andreas.arzt@jku.at

ABSTRACT

In general, algorithms for real-time music tracking directly use a symbolic representation of the score, or a synthesised version thereof, as a reference for the on-line alignment process. In this paper we present an alternative approach. First, different performances of the piece in question are collected and aligned (off-line) to the symbolic score. Then, multiple instances of the on-line tracking algorithm (each using a different performance as a reference) are used to follow the live performance, and their output is combined to come up with the current position in the score. As the evaluation shows, this strategy improves both the robustness and the precision, especially on pieces that are generally hard to track (e.g. pieces with extreme, abrupt tempo changes, or orchestral pieces with a high degree of polyphony). Finally, we describe a real-world application, where this music tracking algorithm was used to follow a world-famous orchestra in a concert hall in order to show synchronised visual content (the sheet music, explanatory text and videos) to members of the audience.

1. INTRODUCTION

Real-time music tracking (or, score following) algorithms, which listen to a musical performance through a microphone and at any time report the current position in the musical score, originated in the 1980s (see [8, 24]) and still attract a lot of research [4, 6, 11, 15, 17, 21, 23]. In recent years this technology has already found use in real-world applications. Examples include Antescofo [1], which is actively used by professional musicians to synchronise a performance (mostly solo instruments or small ensembles) with computer realised elements, and Tonara [2], a music tracking application focusing on the amateur pianist and running on the iPad.

A common approach in music tracking, and also for the related task of off-line audio to score alignment (see e.g. [9, 19, 20]), is to start from a symbolic score representation (e.g in the form of MIDI or MusicXML). Often, this score representation is converted into a sound file using a software synthesizer. The result is a 'machine-like', low-quality rendition of the piece, in which we know the time of every event (e.g. note onsets). Then, a tracking algorithm is used to solve the problem of aligning the incoming live performance to this audio version of the score – thus, the problem of real-time music tracking can be treated as an on-line audio to audio alignment task.

In this paper we follow a similar approach, but instead of using the symbolic score directly, we propose to first automatically align a recording of another performance of the same piece to the score. Then, we use this automatically annotated 'score performance' as the new score representation for the on-line tracking process (for the related task of off-line performance to performance alignment see e.g. [18]). Our motivation for this is twofold. First of all, we expect the quality of the features to be higher than if they were computed from a synthesised version of the score. Also, in a performance a lot of intricacies are encoded that are missing in the symbolic score, including (local) tempo and loudness changes. In this way we implicitly also take care of special events like trills, which normally are insufficiently represented in a symbolic score representation.

As will be seen in this paper, this approach proves to be promising, but the results also depend heavily on which performance was chosen as a reference. To improve the robustness we further propose a multi-agent approach (inspired by [25], where a related strategy was applied to off-line audio alignment), which does not depend on a single performance as a reference, but takes multiple 'score performances' and aligns the live performance to all these references simultaneously. The output of all agents is combined to come up with the current position in the score. As will be shown in the evaluation, this extension stabilises our approach and increases the alignment accuracy.

The paper is structured as follows. First, in Section 2 we give an overview on the data we use to evaluate our music tracker. For comparison, we then give results of the original tracking algorithm that our approach is based on in Section 3. In Section 4 we present a tracking strategy based on off-line aligned performances, which shows promising but unstable results. Then, in Section 5 we propose a multi-agent strategy, which stabilises the tracking process and

[1] repmus.ircam.fr/antescofo
[2] tonara.com

ID	Composer	Piece Name	# Perf.	Groundtruth
CE	Chopin	Etude Op. 10 No. 3 (excerpt)	22	Match
CB	Chopin	Ballade Op. 38 No. 1 (excerpt)	22	Match
MS	Mozart	1^{st} Mov. of Sonatas KV279, KV280, KV281, KV282, KV283, KV284, KV330, KV331, KV332, KV333, KV457, KV475, KV533	1	Match
RP	Rachmaninoff	Prelude Op. 23 No. 5	3	Manual
B3	Beethoven	Symphony No. 3	1	Manual
M4	Mahler	Symphony No. 4	1	Manual

Table 1. The evaluation data set.

Error	CE	CB	MZ	RP	B3	M4
≤ 0.05	0.33	0.33	0.55	0.45	0.42	0.23
≤ 0.25	0.96	0.92	0.97	0.90	0.84	0.71
≤ 0.50	0.99	0.96	0.98	0.96	0.91	0.83
≤ 0.75	1	0.98	0.99	0.98	0.94	0.87
≤ 1.00	1	0.98	0.99	0.98	0.95	0.91

Table 2. Results for the *original on-line tracking algorithm*. The results are shown as proportion of correctly aligned pairs of time points (note times or downbeat times, respectively), for different error tolerances (in seconds). For instance, the first number in the first row means that for the Chopin Etude the alignment was performed for 33% of the notes with an error smaller than or equal to 0.05 seconds.

Error	CE	CB	MZ	RP	B3	M4
≤ 0.05	0.92	0.87	0.93	0.75	0.54	0.38
≤ 0.25	0.99	0.97	0.99	0.97	0.93	0.86
≤ 0.50	1	0.97	1	0.99	0.96	0.94
≤ 0.75	1	0.98	1	0.99	0.97	0.97
≤ 1.00	1	0.98	1	1	0.98	0.98

Table 3. Results for the *off-line alignments*. The results are shown as proportion of correctly aligned pairs of time points (note times or downbeat times, respectively), for different error tolerances (in seconds). For instance, the first number in the first row means that for the Chopin Etude the alignment was performed for 92% of the notes with an error smaller than or equal to 0.05 seconds.

improves the results for all test pieces. Next, we compare the results of the previous chapters to each other (Section 6). Finally, we describe a real-life application of our algorithm at a world-famous concert hall, where it was used to track Richard Strauss' *Alpensinfonie* (see Section 7).

2. DATA DESCRIPTION

To evaluate a real-time music tracking algorithm, a collection of annotated performances is needed. Table 1 gives an overview on the data that will be used throughout the paper. It is important to note that the dataset includes two orchestral pieces (symphonies by Beethoven and Mahler), which in our experience are difficult challenges for music tracking algorithms, due to their high polyphony and complexity. The table also indicates how the ground truth was compiled. For the Chopin Ballade and Etude, and for the Mozart piano sonatas we have access to accurate data about every note onset ('matchfiles') that was played, as these were recorded on a computer-monitored grand piano (see [12] and [26] for more information about this data). For the Prelude by Rachmaninoff as well as for the Symphonies by Beethoven and Mahler we have to rely on manually annotated performances (at the note level for the prelude and at the downbeat level for the two symphonies).

Furthermore, we collected a number of additional performances of the pieces in our dataset. For these we do not have any annotations, and their sole purpose is to be processed fully automatically. These will act as replacements for the symbolic scores. We collected 7 additional performances for each piece in the dataset. We made an exception for the excerpts of the Ballade and the Etude by Chopin, as we already have 22 performances of those. We thus reused these performances accordingly, randomly selected 7 additional performances for each performance in the evaluation set, and treated them in the same way as the other additional data (i.e. we did not use any part of the ground truth, everything was computed automatically when they were used as a 'score performance'). We also took care not to use additional performances of the same performer(s) that occur in our evaluation set.

3. STANDARD MUSIC TRACKING BASED ON A SYMBOLIC SCORE REPRESENTATION

Our approach to music tracking is based on the standard dynamic time warping (DTW) algorithm. In [10] extensions to DTW were proposed that made it applicable for on-line music tracking: 1) the path is computed in an incremental way, and 2) the complexity is reduced to being linear in the length of the input sequences. Later on, this algorithm was extended with a 'backward-forward' strategy, which reconsiders past decisions, increasing the robustness [4], and a simple tempo model (see [3]), which greatly increases the ability of the algorithm to cope with tempo differences.

To make music tracking possible, some internal repre-

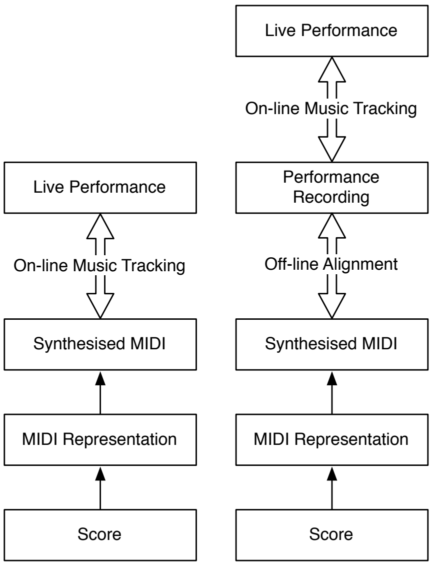

Figure 1. Standard music tracking (left) vs. music tracking via an off-line aligned reference performance (right).

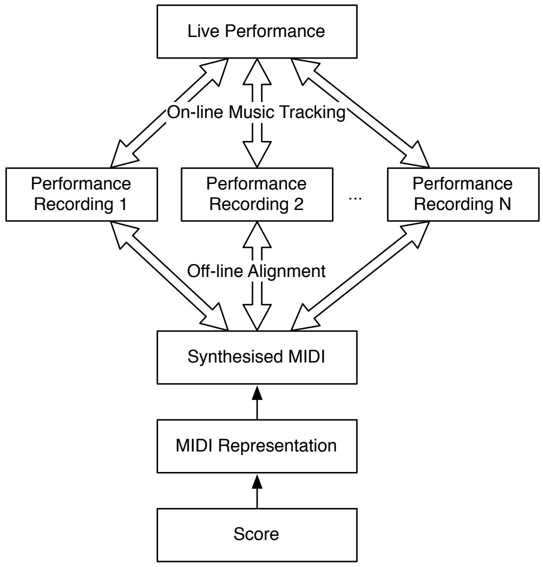

Figure 2. Multi-agent tracking based on off-line aligned performances as a reference.

sentation of the musical score is needed. In this case we start with a MIDI version of the score, which is converted into an audio file using a software synthesizer. Thus we actually treat this task as an audio-to-audio alignment problem, with additional knowledge about the score audio file (i.e. the exact timing of each note). See Figure 1 (left) for a sketch of this setup. In our approach we use the features (a mix of chroma features and 'semi-tone onset' features) and the distance computation method presented in [5].

For comparison, we re-evaluated this algorithm on our data. Each performance from our evaluation set was aligned to the symbolic score representation. The results are given in Table 2. The goal of this paper is to improve on these results, both regarding tracking precision and, especially, robustness (i.e. reduce the amount of big mistakes made by the music tracker). As can be seen, the algorithm works particularly well on the piano pieces, but shows problems with the two symphonies. A reason for this is that it is relatively easy to synthesise piano pieces from MIDI in acceptable quality, but it is much harder to do this automatically for orchestral pieces.

4. MUSIC TRACKING VIA A SINGLE PERFORMANCE AS A REFERENCE

As we are effectively treating the task of music tracking as an on-line audio-to-audio alignment task, we can actually use any annotated audio recording of a performance as a score representation. Using a real performance as a 'score' has some advantages.

First of all, an audio file synthesised from a deadpan MIDI file may sound bad compared to a real performance, thus also the features are of relatively low quality (i.e. they differ sometimes quite heavily from the features computed from the live performance we want to track). Despite obvious differences between performances, their respective features tend to be more similar to each other. This is especially true for orchestral pieces, which often include instruments that are hard to synthesise in high quality (or at least this would demand for expensive sound fonts and a lot of effort by a trained audio engineer).

Secondly, a performance implicitly encodes a lot of information that is missing in the symbolic score. This includes detailed information about tempo, loudness and articulation. Again we want to stress that of course performances differ from each other quite heavily, but compared to the differences between a performance and an audio synthesised from the MIDI, these differences are small.

There is also one big disadvantage: the symbolic information linking time points in the audio to beat times in the score, which we get for free when we use a MIDI file as the basis for the score audio, is missing. Thus, this information needs to be generated. There are two possible ways to do that: (1) by manual annotation, which can be very laborious, or (2) by automatic off-line alignment of the performance to the score – which is the option we decided on, as we are interested in an automatic method to improve tracking results (see Section 4.1 below).

Figure 1 shows a sketch of the intended setup. On the left, 'normal' music tracking is shown, where the live performance is aligned to the symbolic score (via a synthesised audio). On the right, another performance is first aligned to the symbolic score. This performance is then used as the new reference in the on-line alignment process.

4.1 Offline Alignment

To use a performance as a 'score' we have to generate the necessary symbolic information, linking time points in the audio to beat times in the score. As we are interested in an automatic way to improve the tracking results, we decided to use off-line audio alignment to align the 'score perfor-

Error	CE	CB	MZ	RP	B3	M4
≤ 0.05	0.39	0.35	0.52	0.25	0.35	0.27
≤ 0.25	0.98	0.96	0.97	0.87	0.85	0.80
≤ 0.50	0.99	0.97	0.99	0.97	0.93	0.92
≤ 0.75	1	0.98	0.99	0.99	0.95	0.95
≤ 1.00	1	0.98	1	1	0.97	0.96

Table 4. Results for *on-line music tracking* based on a *single off-line aligned performance as a reference*. The results are shown as proportion of correctly aligned pairs of time points (note times or downbeat times, respectively), for different error tolerances (in seconds). For instance, the first number in the first row means that for the Chopin Etude the alignment was performed for 39% of the notes with an error smaller than or equal to 0.05 seconds.

Error	CE	CB	MZ	RP	B3	M4
≤ 0.05	0.39	0.35	0.58	0.19	0.44	0.32
≤ 0.25	0.99	0.98	0.99	0.92	0.90	0.84
≤ 0.50	1	0.98	1	1	0.95	0.94
≤ 0.75	1	0.98	1	1	0.96	0.96
≤ 1.00	1	0.99	1	1	0.97	0.97

Table 5. Results for the *multi-agent tracking* approach based on a *set of off-line aligned performances as a reference*. The results are shown as proportion of correctly aligned pairs of time points (note times or downbeat times, respectively), for different error tolerances (in seconds). For instance, the first number in the first row means that for the Chopin Etude the alignment was performed for 39% of the notes with an error smaller than or equal to 0.05 seconds.

mance' to the symbolic score, which gives us the needed mapping as a result. As off-line audio alignment is far more accurate than on-line tracking, our intuition was that the increase in feature quality outweighs the introduced error by the off-line alignment process.

The off-line alignment is computed with the music tracking algorithm from Section 3 above, with the only difference being that in the end we compute the backward path, as it is done in the standard DTW algorithm. As this path is based on more information (i.e. it is computed in a non-causal way), the results are generally much more accurate than in the on-line case. Of course any off-line audio score alignment algorithm could be used for this task (see e.g. [16, 19, 20].

Just to get a rough idea of how much error will be introduced by the off-line alignment, we ran an experiment on our test data and aligned it to the symbolic scores (later on, off-line alignments of the additional data will be used, but we expect a similar behaviour). Unsurprisingly, the results show that there is a gap between the results of the off-line approach (see Table 3) and the on-line music tracking approach (see Table 2). As we will use the off-line algorithm during data preparation, we strongly expect that the higher quality of the features and the additional information encoded in the performances will outweigh the error that is introduced during this step.

Thus, we aligned all the additional performances from Section 2 to the respective symbolic scores, resulting in performances with linked symbolic information. In the following sections, we will use these performances as new references ('score performances') for the music tracking algorithm.

4.2 Tracking based on an aligned Performance

Given the automatically computed 'score performances', we can now use them in the tracking process as shown in Figure 1. In this experiment, each performance from the evaluation set is aligned to the score via each respective 'score performance', resulting in 7 on-line alignments for each performance.

The results are given in Table 4 and should be compared to the numbers in Table 2. As can be seen, the general trend is an improvement in robustness, especially for the complex orchestral pieces (e.g. the percentage of aligned downbeats with an error smaller than 250 ms increased from 71% to 80% for the Mahler Symphony).

Unfortunately, the results also proved to be unstable. Some performances are more similar (or at least easier to align) to each other, which also results in good tracking results – but the use of some of the 'score performances' led to results that were worse than our basic approach. A closer look at the positions where tracking errors occurred showed that some of them happened at the same points in time over all alignments of the piece – basically showing that some parts are harder to track than others. But there were also many alignment errors that occurred only for one or two of the 'score performances', but not for the others. This led us to the idea to combine individual on-line alignments in such away, that it would smooth out these errors.

5. MUSIC TRACKING VIA A SET OF PERFORMANCES AS REFERENCE

The analysis of the results from Section 4 above showed that a combination of a number of on-line alignments might further improve the tracking results. Here, we propose a simple multi-agent strategy (see Figure 2 for an illustration). During a live concert n trackers run in parallel and each tracker tries to align the incoming live performance to its score representation, each producing its own, independent hypothesis of the current position in the score. Finally, the hypotheses are combined to form one collective hypothesis of the music tracking system.

Many different ways of combining the hypotheses would be possible, e.g. based on voting or on the current alignment error of the individual trackers. Here, we decided on a very simple method: taking the median of the positions that are returned by the individual trackers. The reasoning behind this is that trackers tend to make mistakes in both directions – i.e. 'running ahead' (reporting events to early), and 'lagging behind' (reporting events with some delay) –

with about the same frequency. Thus, trackers that stay safely in the middle of the pack tend to give a robust estimate of the position in the score.

Furthermore, using the median also means that as long as $\frac{n}{2} + 1$ trackers stay close to the actual position, the system would still come up with a reasonable position estimate – while this is not directly reflected in the evaluation results, this extra robustness is convenient when the tracking algorithm is used in real-world applications. Further strategies to increase the robustness are possible, like the automatic replacement of trackers that got lost, but were not used in our experiments.

For the evaluation we set $n = 7$, as this was a good trade-off between robustness and computation time (7 on-line alignments can still be easily computed in real-time on a conventional consumer laptop). The results, given in Table 5, show that our approach is working well. Errors of more than 1 second are rare, and the multi-agent approach even improved the alignment precision for all pieces (with the exception of the Prelude by Rachmaninoff).

Piece	Offline	Standard	Via 1	Via 7
CE	99.06%	95.62%	97.92%	**98.78%**
CB	97.13%	92.10%	96.00%	**97.93%**
MZ	99.35%	96.88%	97.46%	**99.04%**
RP	96.62%	90.14%	87.47%	**92.47%**
B3	92.88%	83.67%	85.04%	**89.55%**
M4	86.74%	71.15%	80.06%	**83.66%**

Table 6. Comparison of the results (error tolerance 250 ms). The results are shown as percentage of matching pairs of time points (note times or downbeat times, respectively). For instance, the first number in the first row means that for the Chopin Etude the off-line alignment was performed for 99.06% of the notes with an error smaller than or equal to 0.25 seconds. The results of the *offline* alignment algorithm are only shown for comparison. *Standard* refers to the basic on-line music tracker (see Section 3), *Via 1* to the tracker using a single 'score performance' as a reference, *Via 7* to the multi-agent approach based on 7 trackers.

6. DISCUSSION

The main goal of our approach was to increase the robustness of the algorithm, i.e. to decrease the frequency of 'large' errors and to make sure that the tracker does not get lost, even when following difficult orchestral pieces. For convenience, we give a summary of the results (see Table 6) based on a common measure in the evaluation of music tracking algorithms: the percentage of notes that were aligned with an error less than or equal to 250 ms (see [7]). As can be seen, the multi-agent approach based on automatically aligned reference performances improves the results heavily – in fact for CB the results of the on-line alignment even surpassed the off-line alignment. For the results on the Chopin data (CE and CB) one has to take into account that we used 22 performances which were recorded by different performers, but still on the same piano and with the same recording setup, which will have a positive influence on the alignment results. Still, as the remaining results show, even when completely unrelated performances of the same piece were used as references, the alignment results improved drastically.

Especially for the orchestral pieces (B3 and M4), we can see that our intuition proved to be correct: the error introduced by the off-line alignment had a lot less impact than the better quality of the features and the additional tempo and loudness information provided by the performances. In addition, the multi-agent approach proved to be very effective regarding the increase in robustness. It smooths out some of the bigger errors that occur when using just a single performance as a score reference.

7. REAL-LIFE SCENARIO: MUSIC TRACKING IN THE CONCERTGEBOUW AMSTERDAM

The multi-national European research project PHENICX [3] provided us with the unique opportunity (and challenge) to demonstrate our score following technology in the context of a big, real-life symphonic concert (for a full description of this experiment see [2], a similar study was presented in [22]). The general goal of the project is to develop technologies that enrich the experience of classical music concerts. In the experiment to be described, this was done by using the live performance tracker to control, in real time and via WiFi, the transmission and display of additional visual and textual information, synchronised to the live performance on stage. The user interface and the visualisations were provided by our project partner Videodock [4]. Some impressions can be seen in Figure 3.

The event took place on February 7th, 2015, in the Concertgebouw in Amsterdam. The Royal Concertgebouw Orchestra, conducted by Semyon Bychkov, performed the *Alpensinfonie* (Alpine Symphony) by Richard Strauss. This concert was part of a series called 'Essentials', during which technology developed within the project can be tested in a real-life concert environment. All the tests during this concert series have to be as non-invasive as possible. For the demonstration during the concert in question, a test audience of about 30 people was provided with tablet computers and placed in the rear part of the concert hall.

In contrast to the experiments presented in this paper so far, we did not even have access to a symbolic score. Instead, we annotated a single performance manually (on the level of downbeats) and used it as a score representation. Then, to add extra robustness, we aligned 6 more performances to this reference, resulting in 7 instances that can be used for the tracking process.

The event in the Concertgebouw was a big success. The tracking went smoothly and there were no glitches, only some minor inaccuracies, and the accuracy was more than sufficient to trigger the visualisation in time.

After the event we annotated an audio recording of the concert to be able to perform quantitative experiments (see

[3] http://phenicx.upf.edu

[4] http://videodock.com

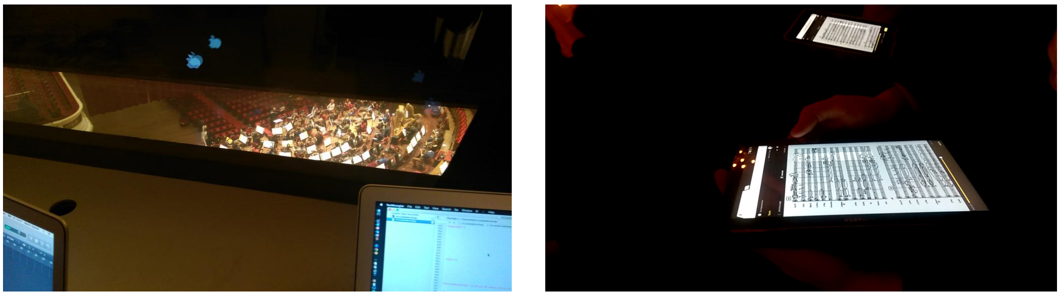

Figure 3. Left: View from the control room onto the stage (during orchestra rehearsal); right: synchronised score display in the audience during the concert.

Err. (sec)	Single	Multi-agent
≤ 0.25	78.25%	81.80%
≤ 0.50	92.20%	93.24%
≤ 0.75	95.57%	96.44%
≤ 1.00	97.49%	98.01%

Table 7. Real-time alignment results for the single tracker (using only on manually annotated performance), and the multi-agent tracker, shown as percentages of correctly aligned pairs of downbeats. For instance, the first number in the first row means that the single tracker aligned 78.25% of the downbeats with an error smaller than or equal to 0.25 seconds.

Table 7). The first column shows the results of the tracking using only the manually annotated performance as a reference. The second column shows the results of the multi-agent approach. Also in this case using multiple performances as a reference improved the tracking results: extra robustness and a slight increase in accuracy were achieved without any extra manual efforts as the additional data was prepared by automatic methods.

8. CONCLUSION

In this paper we presented an alternative approach to real-time music tracking. Instead of tracking directly on a symbolic score representation, we first use off-line alignment to match other performances of the piece in question to the symbolic score. We then use these performances as our new score representation, which results in high quality features, and implicitly also adds extra information about how this piece generally is performed. Together with a multi-agent tracking strategy, which smooths out most of the major errors, we achieve increased robustness and also increase the accuracy of the live tracking, especially for complex orchestral music. We also reported on a successful real-world test of our algorithm in a world-famous concert hall.

In the future, we will also look at other options to combine tracking results of the individual trackers. While taking the median seems like a natural choice, more sophisticated strategies also based on alignment costs might be promising. A further problem which deserves a closer look is the automatic selection strategy of the 'score performances'. For this paper we simply decided on 7 additional performances of the pieces based on availability. With a bigger database, automatic selection of the 'best score performances' for an on-going live performance becomes an interesting question, and a good selection strategy might further improve the tracking results.

A common problem of real-time music tracking and audio to score alignment are structural differences between the score and the performance. For example, if a piece has some repeated sections, the performers might decide to play the repetition or to leave it out. For the experiments in this data we chose the additional 'score performances' manually, such that they have the same structure as the piece we want to track, but in the future we will try to cope with this automatically – in the preparation phase via the technique used in [13] or [14] (maybe in combination with the method described in [25], to bring the benefit of using multiple performances also to the preprocessing stage), and in the live tracking phase with the approach presented in [1], extended to orchestral music.

9. ACKNOWLEDGEMENTS

This research is supported by the Austrian Science Fund (FWF) under project number Z159 and the EU FP7 Project PHENICX (grant no. 601166).

10. REFERENCES

[1] Andreas Arzt, Sebastian Böck, Sebastian Flossmann, Harald Frostel, Martin Gasser, and Gerhard Widmer. The complete classical music companion v0.9. In *Proc. of the AES Conference on Semantic Audio*, London, England, 2014.

[2] Andreas Arzt, Harald Frostel, Thassilo Gadermaier, Martin Gasser, Maarten Grachten, and Gerhard Widmer. Artificial intelligence in the concertgebouw. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, Buenos Aires, Argentina, 2015.

[3] Andreas Arzt and Gerhard Widmer. Simple tempo models for real-time music tracking. In *Proc. of*

the Sound and Music Computing Conference (SMC), Barcelona, Spain, 2010.

[4] Andreas Arzt, Gerhard Widmer, and Simon Dixon. Automatic page turning for musicians via real-time machine listening. In *Proc. of the European Conference on Artificial Intelligence (ECAI)*, Patras, Greece, 2008.

[5] Andreas Arzt, Gerhard Widmer, and Simon Dixon. Adaptive distance normalization for real-time music tracking. In *Proc. of the European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, 2012.

[6] Arshia Cont. A coupled duration-focused architecture for realtime music to score alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):837–846, 2009.

[7] Arshia Cont, Diemo Schwarz, Norbert Schnell, and Christopher Raphael. Evaluation of real-time audio-to-score alignment. In *Procedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, 2007.

[8] Roger Dannenberg. An on-line algorithm for real-time accompaniment. In *Proc. of the International Computer Music Conference (ICMC)*, Paris, France, 1984.

[9] Roger Dannenberg and Ning Hu. Polyphonic audio matching for score following and intelligent audio editors. In *Proc. of the International Compter Music Conference (ICMC)*, Singapore, 2003.

[10] Simon Dixon. An on-line time warping algorithm for tracking musical performances. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, Edinburgh, Scotland, 2005.

[11] Zhiyao Duan and Bryan Pardo. A state space model for on-line polyphonic audio-score alignment. In *Proc. of the IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011.

[12] Sebastian Flossmann, Werner Goebl, Maarten Grachten, Bernhard Niedermayer, and Gerhard Widmer. The magaloff project: An interim report. *Journal of New Music Research*, 39(4):363–377, 2010.

[13] Christian Fremery, Meinard Müller, and Michael Clausen. Handling repeats and jumps in score-performance synchronization. In *Proc. of the International Conference on Music Information Retrieval (ISMIR)*, Utrecht, The Netherlands, 2010.

[14] Maarten Grachten, Martin Gasser, Andreas Arzt, and Gerhard Widmer. Automatic alignment of music performances with structural differences. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Curitiba, Brazil, 2013.

[15] Filip Korzeniowski, Florian Krebs, Andreas Arzt, and Gerhard Widmer. Tracking rests and tempo changes: Improved score following with particle filters. In *Proc. of the International Computer Music Conference (ICMC)*, Perth, Australia, 2013.

[16] Marius Miron, Julio José Carabias-Orti, and Jordi Janer. Audio-to-score alignment at note level for orchestral recordings. In *Proc. of the International Conference on Music Information Retrieval (ISMIR)*, Taipei, Taiwan, 2014.

[17] Nicola Montecchio and Arshia Cont. A unified approach to real time audio-to-score and audio-to-audio alignment using sequential montecarlo inference techniques. In *Proc. of the IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011.

[18] Meinard Müller, Frank Kurth, and Michael Clausen. Audio matching via chroma-based statistical features. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, London, Great Britain, 2005.

[19] Meinard Müller, Frank Kurth, and Tido Röder. Towards an efficient algorithm for automatic score-to-audio synchronization. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Barcelona, Spain, 2004.

[20] Bernhard Niedermayer and Gerhard Widmer. A multi-pass algorithm for accurate audio-to-score alignment. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, The Netherlands, 2010.

[21] Takuma Otsuka, Kazuhiro Nakadai, Toru Takahashi, Tetsuya Ogata, and Hiroshi G. Okuno. Real-time audio-to-score alignment using particle filter for co-player music robots. *EURASIP Journal on Advances in Signal Processing*, 2011(2011:384651), 2011.

[22] Matthew Prockup, David Grunberg, Alex Hrybyk, and Youngmoo E. Kim. Orchestral performance companion: Using real-time audio to score alignment. *IEEE Multimedia*, 20(2):52–60, 2013.

[23] Christopher Raphael. Music Plus One and machine learning. In *Proc. of the International Conference on Machine Learning (ICML)*, Haifa, Israel, 2010.

[24] Barry Vercoe. The synthetic performer in the context of live performance. In *Proc. of the International Computer Music Conference (ICMC)*, Paris, France, 1984.

[25] Siying Wang, Sebastian Ewert, and Simon Dixon. Robust joint alignment of multiple versions of a piece of music. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Taipei, Taiwan, 2014.

[26] Gerhard Widmer. Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence*, 146(2):129–148, 2003.