MIDI-to-Score Alignment Tool

Miao Zhang
Center for Computer Research in Music and Acoustics
Stanford University
miaoz18@stanford.edu

Abstract

MIDI-to-score alignment is a technique to automatically match a note in a MIDI file of music performance to the corresponding note in the score. It is an important pre-processing step for analysis of performance since the timings of notes can be related to the meter and melodies. This work shows an algorithm to link notes in MIDI file and Humdrum file using the context of each note as reference. The algorithm doesn't include complicated machine learning structure but showed very similar alignment results with the state-of-art MIDI-to-Score alignment tool from work (Nakamura et al. 2017) on four alignment tasks.

1 Introduction and Motivations

The process of music alignment is to match notes in a music performance(aligned signal) to those in a reference musical score or another performance signal (reference signal). And the automatically music alignment has become a popular topic for research and a fundamental technique for music information processing. Audio-to-Score alignment and MIDI-to-Score alignment are two main topics of this technique. This project focuses on MIDI-to-score alignment, where the digital score file is in Humdrum format (Sapp. 2005).

Turning music performances to data files that can be computationally analyzed is valuable, and that can be used in performance analysis. For example, the analysis on hand breaking in arpeggiation. The timing of arpeggiated chords involving both hands sequentially can be measured in performances. (Repp. 1997) measured ten pianist students' performance and found that Arpeggio durations depended more on the preceding note value and position in the phrase than on the number of notes. The author showed that there were consistent individual differences in the music students' timing patterns. The alignment tool can be helpful to the further analysis on their difference of hand breaking in arpeggiation and the investigation of the origins of those differences.

Additinally, music alignment can be used to analyze phrasing and pedaling patterns in music. Phrase is an important music element and it has been researched in many relevant fields. Nan, Y., Knösche, T. R., Friederici, A. D. (2006) used Electroencephalography (EEG) in a cross-cultural music study to investigate phrase boundary perception, where Chinese and German musicians performed a cultural categorization task under Chinese and Western music listening conditions. Phrasing and pedaling analysis can also be performed with alignment tool which has access to the timing information of the notes in the performance.

Furthermore, music alignment is useful in ornaments recognition and. For example, in a score there will be one note representing a trill, but many notes in the MIDI file for the realization of the trill, as shown in Figure 1. Thus, a task to do after alignment is finished is to identify trilled notes, and then we can describe how the trill was performed (how many notes, when did the trill stop, etc).

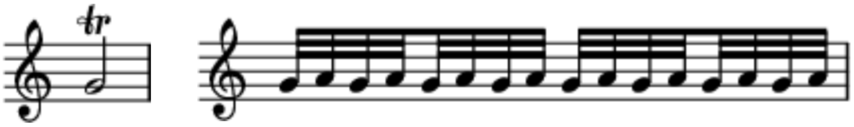


Figure 1: Trill in score and in performance

This study deals with offline symbolic MIDI-to-Score alignment, with particular focus on piano performances.I will mainly consider Western classical music(Beethoven, Mozart, Chopin) or similar music styles where musical scores exist behind the performances.

2 Related work

Eita Nakamura' work Symbolic Music Alignment (Nakamura et al. 2017) (or score-to-performance matching) is a technique to automatically match a note in a music performance to the corresponding note in the score. He focused on reducing alignment errors by detecting performance errors in the preliminary alignment result, then the algorithm automatically finds regions in the performance that can potentially contain alignment errors and corrects them by the hidden Markov model approaches.

HMM is a method to find the optimal alignment when handling deviations in music performances. For example, an HMM is constructed for each reference signal, in which note insertions and deletions, repeats, and skips are described by transition probabilities, and pitch errors are described by output probabilities. The aligned signal is considered as an output sequence from the HMM and the most probable sequence of latent states is estimated with the Viterbi algorithm for alignment. The main alignment algorithm consists of three parts: preliminary alignment by hidden Markov models (HMMs); performance error detection; post-processing realignment by merged-output HMMs.

The author believes that alignment errors are often connected with performance errors. So by detecting performance errors in a given result of automatic alignment, it's possible to select limited regions in the aligned signal that may contain alignment errors. He first develop a performance error detection algorithm that recognizes pitch errors, extra notes, and missing notes in a given alignment result. Error regions are then defined as segments of aligned and reference signals around performance errors and he investigate how much alignment errors are contained in these regions with various sizes of the regions. Next he develop a post-processing realignment method to correct limited segments of preliminary results, which combines the method using merged-output HMMs with a voice(hand) separation method. The preliminary alignment method is the one based on temporal HMMs. With the realignment method combined with an HMM-based method, the author achieved the highest accuracies, with short computation time in all tested data and for both score-to-MIDI and MIDI-to-MIDI alignment cases. Figure 2 shows the outcome of this method.

Arzt, A., Lattner, S. (2018) performed the audio-to-score alignment using novel transposition-invariant audio features. The transposition-invariant audio features are low-dimensional features representing local pitch intervals. And they were learned in an unsupervised fashion by a gated autoencoder, the model raised by authors. The results showed that the proposed features were indeed fully transposition-invariant and enable accurate alignments between transposed scores and performances, also, they can outperform widely used features for audio-to-score alignment on 'untransposed data'. Kwon, T., Jeong, D., Nam, J. (2017) offered another way to extract features for audio-to-score alignment. They used two recurrent neural networks that work as the AMT(automatic music transcription) based feature extractors to the alignment algorithm. The note prediction output can be regarded as a learned feature representation that is directly comparable to MIDI note. Dynamic time warping was applied as the alignment algorithm without any additional post-processing. And the authors obtained the significantly improved accuracy than previous work with the proposed alignment framework and learned features on the MAPS data set.

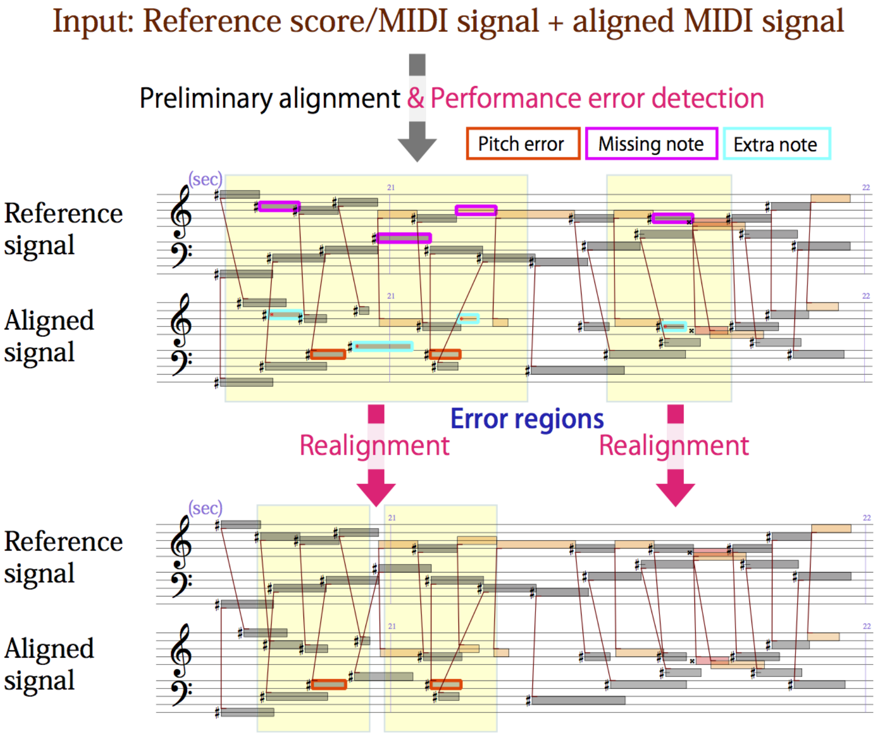


Figure 2: An outcome of the proposed method. Errors in preliminary alignment caused by reordered note pairs in the aligned signal are corrected by the realignment method (Nakamura et al. 2017).

3 Method

3.1 Algorithm Design

The basic idea of the algorithm is to process each MIDI note in time order by linking it to a best score note (humdrum format). In this work, I only consider pitch to decide whether the two notes can be linked. I use four assistant parameters to help choose the "best" humdrum note in the procedure:

(1) Window length: All surrounding notes which happened not earlier than and not later than the current note for the time length. In MIDI note list, the time length is described in SEC (the start time of the note in seconds), and it is described in QTIME (quarter-note start time of the note relative to the start of the music) in humdrum note list. The parameter is used to decide candidate notes to be linked.

(2) Y1 value. (3)Y2 value. (4) Y3 value: The parameters used to describe the position of the note in the list. As shown in Figure 3, if we have one MIDI note and one humdrum note to be linked, the ratio of the time difference between current note and the previous note at the same pitch to the time difference between current note and the next note at the same pitch can be calculated in both MIDI list and humdrum list. Then, the ratio of the MIDI note ratio to the humdrum note ratio is the so called Y1 value for this alignment pair. Similarly, the Y2 value and Y3 value can be calculated, and the only difference is that instead of searching for the previous same-pitch note and the next same-pitch note, we search for the previous two same-pitch notes and next two same-pitch notes to calculated the ratio. The parameters are used to pick the best note to link among the candidate notes.

I use the time window with the length initialized as 1 for both MIDI notes list and humdrum notes list to search for a suitable link. Take the MIDI to score alignment for example, at the very beginning, the first note in the MIDI list is linked to the first note having the same pitch in humdrum list. Then, for each MIDI note after the first one, check all other successfully matched MIDI notes around it (within the window length) and their linked humdrum notes, and the target humdrum note should be around (within the window length). After

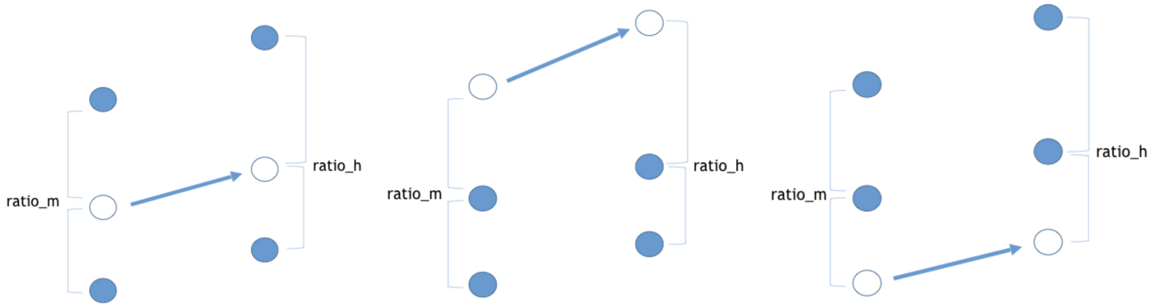


Figure 3: Computation of Y1 value, Y2 value and Y3 value.

having a range in humdrum list including candidate humdrum notes, to choose the best one to link among them, compute Y1 value, Y2 value and Y3 value to evaluate how good the link is. Ideally, the three Y values should be exactly 1 representing that the MIDI note and the humdrum note are exactly the same one, but a deviation within a specific range is tolerable. Here we pick the humdrum note having the Y values most close to 1 to link to the MIDI note. And there should be a threshold for the Y values to help decide whether a MIDI note-humdrum note link is good enough to be valid.

In this way, all notes in the MIDI list searches for and stores a possible link to the humdrum list. If a MIDI note fails to find a humdrum note to link, the stored value is negative 1. The possible failed situations are: First, A note in the MIDI list was not found in the humdrum list; Second, a note in the humdrum list was not found in the MIDI list; Third, a combination of the first and the second cases where a note is at one pitch in the MIDI data by a different one in the humdrum data. This would be two independent cases of the first case and the third case as well as dependent cases which represent a "wrong" note played by the pianist.

The github link of the project can be found and downloaded here: `https://github.com/zm17943/MIDI-Score-Alighment-Tool`.

3.2 Problems and Solutions

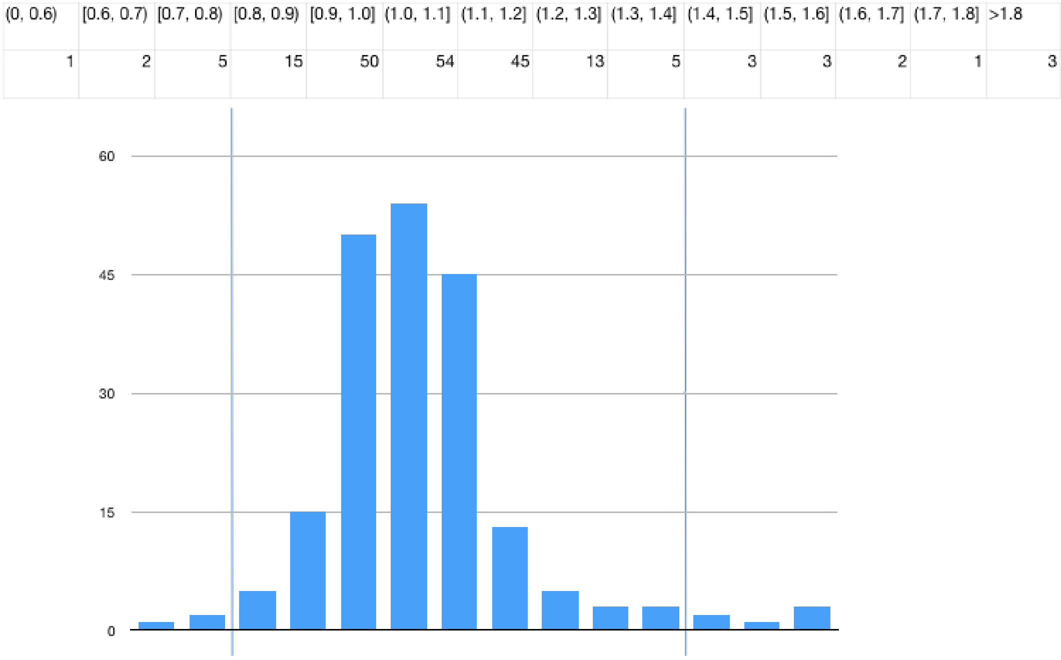
Deviations in music performances will bring challenge to music alignment. Possible deviations include tempo changes, performance errors (e.g. pitch errors, note insertions and deletions), ornamentation, and global structural differences (repeats and skips). Furthermore, it has been found that in the case of polyphonic piano performances, deviations in performances due to asynchronies between hands/voices result in reordering of notes with different score times, which is the main cause of alignment errors for the preliminary time-window system that is not specially designed to handle them.

Here are several potential problems of the proposed algorithm: First, the initialized window length and Y value range might not work for the current task and need to be adjusted. To solve this, the algorithm do the realignments after the first alignment. For example, if the result shows too many failed linked notes, the window length will be lengthened and the tolerance of the Y values will be increased; Second, if there is one extra MIDI notes, for example, an ornamentation, which is very close to and have the same pitch with the original note, Y1 value will be negatively influenced, since the small distance between them push the value far away from the ideal 1. Thus, the algorithm check the Y2 value and Y3 value if Y1 value is bad. If one of the Y values is good, the link is regarded as successful; Third, if the MIDI note to be linked is the first note in a measure, it is very likely that the time difference between it and the previous note is longer than the window length. The note has no reference linked note to help decide the candidates humdrum note, so the alignment tends to fail. The preliminary solution is to lengthen the window length temporarily for that note by adding the time difference between two measures.

However, the solutions showed above are not perfect since it can't handle all situations. For example, all three Y values of a note are affected by extra or wrong notes around leading to the failure of the linking; There is a wrong note where the note is at one pitch in the MIDI data by a different one in the humdrum data, then the algorithm is difficult to decide and may recognize the wrong note as an extra MIDI note and an extra score note. There are many more details to consider in the alignment and it's necessary to be robust to all possible situations.

4 Result and conclusion

In this stage, the algorithm is tested on four MIDI-to-Score alignment tasks which are Prelude op. 28, Nr. 15, Prelude op. 28, no. 17, Prelude op. 28, no. 18 and Prelude op. 28, no. 20 from Chopin. Figure 4 shows the statistical distribution of the average of Y1 value, Y2 value, and Y3 value for all MIDI notes in Prelude op. 28, no. 15 after adjustment. The tolerance range of Y values representing a good link were set to bigger than 0.7 and smaller than 1.6, which is reasonable for this task. I compared the alignment results of the proposed tool to the alignment tool from Nakamura et al., (2017) which has been proved to have the accuracy higher than 0.98. I found that the detection for not linked abnormally notes (including extra note, missing note and wrong note) of these two systems have the similarity higher than 0.9 which was counted by hand, showing that the algorithm is effective and accurate on these four examples, but the alignment tool need to be tested on much more examples in the future.



(0, 0.6)	[0.6, 0.7)	[0.7, 0.8)	[0.8, 0.9)	[0.9, 1.0]	(1.0, 1.1]	(1.1, 1.2]	(1.2, 1.3]	(1.3, 1.4]	(1.4, 1.5]	(1.5, 1.6]	(1.6, 1.7]	(1.7, 1.8]	>1.8
1	2	5	15	50	54	45	13	5	3	3	2	1	3

Figure 4: Statistical distribution of Y values.

5 Future work

As shown in "Problems and Solutions" session, there are still many unsolved problems left for this algorithm. What I can try in the future is a more complicated but solid way to find the optimal alignment which is the sequence matching method dynamic time warping (DTW). What is showed in the figure below is the constrained DTW path, where x-axis is the index of the current MIDI data point to be searched, and y-axis is the index of score note. The Euclidean distance is used to compute the difference between them. D is the accumulated distance when the MIDI is at the t note and the reference signal is at the j note, and α is the weighting of the diagonal steps. This constrained path inhibits successive occurrences of vertical or horizontal steps, and smooth path is estimated. To configure the real-time version of DTW for the on-line search without backtracking. We can simply select the score note which has the smallest accumulated distance with the current MIDI note.

After linking the MIDI and score, I can use the alignment results to do the performance analysis. Music is a performing art, and the differentiation between the score and its performance is hard. Music Performance Analysis (MPA) aims at studying the performance of a musical score rather than the musical score itself. It deals with the observation, extraction, description, interpretation, and modeling of music performance parameters as well as the analysis of attributes and characteristics of the generation and perception of music performance (Rao Hari, 1989). The alignment data show the inconsistency between the performance and the score and it is possible to find rules for specific performers and songs, which is useful in MPA tasks, for example, the analysis topics shown in "Introduction and Motivation" session.

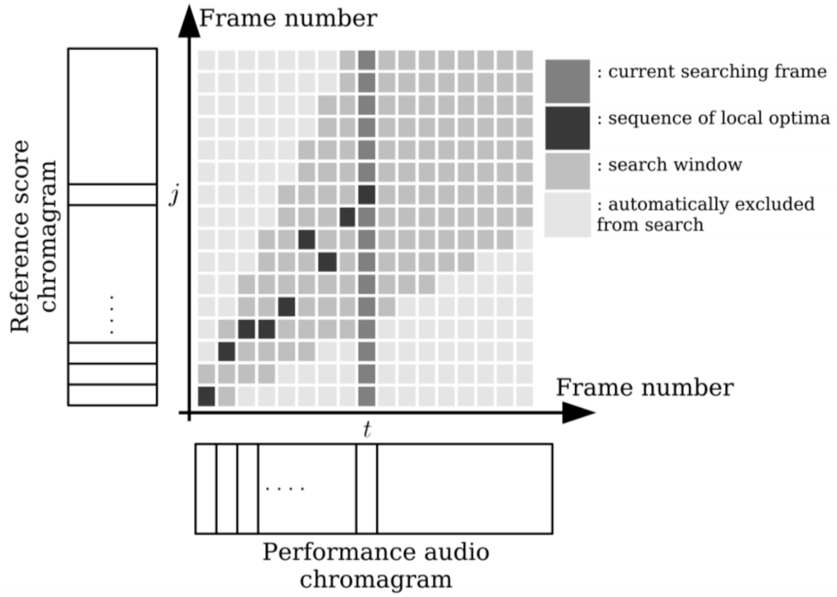

Frame number
: current searching frame
: sequence of local optima
: search window
: automatically excluded from search
Reference score chromagram
j
Frame number
t
....
Performance audio chromagram

Figure 5: The principle of Dynamic Time Warping.

6 Acknowledge

Thanks to Craig Sapp for his guidance on the whole project, offering technical assistance in implementing the loading of MIDI files and Humdrum files, suggesting the idea for the algorithm, and preparing the materials to test the alignment. Thanks to Eleanor Selfridge-Field for giving me advice and related works for reference.

References

[1] Nakamura, E., Yoshii, K., Katayose, H. (2017). Performance Error Detection and Post-Processing for Fast and Accurate Symbolic Music Alignment. In ISMIR (pp. 347-353).

[2] Sapp, C. S. (2005, September). Online Database of Scores in the Humdrum File Format. In ISMIR (pp. 664-665).

[3] Repp, B. H. (1997). Some observations on pianists' timing of arpeggiated chords. Psychology of Music, 25(2), 133-148.

[4] Nan, Y., Knösche, T. R., Friederici, A. D. (2006). The perception of musical phrase structure: a cross-cultural ERP study. Brain research, 1094(1), 179-191.

[5] Hug, F., Dorel, S. (2009). Electromyographic analysis of pedaling: a review. Journal of electromyography and Kinesiology, 19(2), 182-198.

[6] Arzt, A., Lattner, S. (2018). Audio-to-score alignment using transposition-invariant features. arXiv preprint arXiv:1807.07278.

[7] Kwon, T., Jeong, D., Nam, J. (2017). Audio-to-score alignment of piano music using RNN-based automatic music transcription. arXiv preprint arXiv:1711.04480.

[8] Rao, B. D., Hari, K. S. (1989). Performance analysis of root-MUSIC. IEEE Transactions on Acoustics, Speech, and Signal Processing, 37(12), 1939-1949.