

THE BATIK-PLAYS-MOZART CORPUS: LINKING PERFORMANCE TO SCORE TO MUSICOLOGICAL ANNOTATIONS

Patricia Hu¹

Gerhard Widmer^{1,2}

¹ Institute of Computational Perception, Johannes Kepler University Linz, Austria

² LIT AI Lab, Linz Institute of Technology, Austria

firstname.lastname@jku.at

ABSTRACT

We present the *Batik-plays-Mozart* Corpus, a piano performance dataset combining professional Mozart piano sonata performances with expert-labelled scores at a note-precise level. The performances originate from a recording by Viennese pianist Roland Batik on a computer-monitored Bösendorfer grand piano, and are available both as MIDI files and audio recordings. They have been precisely aligned, note by note, with a current standard edition of the corresponding scores (the New Mozart Edition) in such a way that they can further be connected to the musicological annotations (harmony, cadences, phrases) on these scores that were recently published by [1].

The result is a high-quality, high-precision corpus mapping scores and musical structure annotations to precise note-level professional performance information. As the first of its kind, it can serve as a valuable resource for studying various facets of expressive performance and their relationship with structural aspects.

In the paper, we outline the curation process of the alignment and conduct two exploratory experiments to demonstrate its usefulness in analyzing expressive performance.

1. INTRODUCTION

Music performance is a complex and nuanced activity that involves the interplay of various expressive features such as timing, dynamics, and articulation. Expressive performance research in music information retrieval (MIR) focuses on modeling expressive aspects of music performance by analyzing how performers use nuances in timing, dynamics, articulation, and other expressive features to convey their musical intentions, with the aim of developing computational models that can analyze, recognize, or synthesize expressive performances [2].

Recent research in this field for Western classical piano has focused on data-driven approaches both for performance generation [3,4] and data creation in the form of

large-scale MIDI performance data transcribed from audio recordings [5,6]. While such data corpora can be useful for comparative performance analyses and related tasks (e.g., performer identification, performance style transfer), they lack the necessary precision and alignment information (with the underlying musical score) required to precisely map expressive intentions and parameters to underlying score features.

Compared to these large-scale transcribed MIDI datasets, precise MIDI data (as recorded on computer controlled grand pianos such as the Yamaha Disklavier or Boesendorfer SE/Ceus series) along with their corresponding score alignment is somewhat limited in quantity and size [7–9]. The performances in such datasets are typically sourced from advanced piano students or piano competitions, whereas the digital scores are often obtained from open-source, user-curated online libraries such as MuseScore¹.

Regarding the performance-to-score alignment, one would ideally want to have note-by-note correspondence information; unfortunately, in the case of the largest of these datasets [7], score-performance alignments are only given at a rather coarse level of beats. Score annotations conveying structural information such as underlying harmony or phrases are even more scarce.

To address these limitations, we introduce the *Batik-plays-Mozart* dataset², in which we provide a set of expert performances of 12 complete Mozart piano sonatas (36 distinct movements) in MIDI format by concert pianist Roland Batik, precisely aligned, at a note-by-note-level, to a standard edition (the New Mozart Edition) of the score, thereby linking the performance information to a previously published dataset [1] of expert annotations of the scores in terms of harmony, cadence, and phrase structure. To the best of our knowledge, this is the first corpus of its kind, combining high quality digital score and structural annotations with expert performances in recorded MIDI format. We report two preliminary experiments to demonstrate the benefits of having precise performance–score–structure annotation alignments.

The remainder of this paper is organised as follows: Section 2 presents a list of comparable expressive performance datasets currently publicly available. Section 3 de-



© P. Hu and G. Widmer. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** P. Hu and G. Widmer, “The Batik-plays-Mozart Corpus: Linking Performance to Score to Musicological Annotations”, in *Proc. of the 24th Int. Society for Music Information Retrieval Conf.*, Milan, Italy, 2023.

¹ <https://musescore.com/sheetmusic>

² https://github.com/huispaty/batik_plays_mozart

Dataset	Pieces	Size Performances	Modality		Alignment	Annotations
			MIDI	Score		Other
ASAP [7]	222	1,068	recorded	MusicXML	beat	time and key signature
Vienna4x22 [8]	4	88	recorded	MusicXML	note	-
CrestMuse PEDB [9]	35	411	recorded	MusicXML	note	phrase
MazurkaBL [10]	44	2,000	-	MusicXML	beat	dynamics, tempo markings
<i>Batik-plays-Mozart</i>	36	36	recorded	MusicXML	note	phrase, harmony, cadence

Table 1. An overview of publicly available comparable piano performance datasets for which precise recorded MIDI data, score-performance alignments and/or musicological annotations are available.

scribes the data origins, the used data formats, and the curation process. Section 4 gives an overview of the dataset, and Section 5 describes two preliminary experiments to demonstrate the benefits of performance–score–structure annotation alignments. Finally, Section 6 concludes the paper with some remarks for future work.

2. RELATED WORK

Several piano performance datasets have been published in the context of expressive performance analysis and performance rendering. While recently published datasets are considerably larger than *Batik-plays-Mozart*, they provide performance recordings solely in the form of MIDI transcribed from audio recordings [5, 6] or do not include a high-quality digital score ground truth [11]. Despite the encouraging results demonstrated by recent transcription models, they often introduce inaccuracies, such as incorrect note fragmentation, missed note onsets, and falsely identified notes [12]. Similarly, certain expressive performance aspects such as (micro-)timing and tempo can only be measured given either a temporal or note-wise score-performance mapping [2]. Nevertheless, these datasets remain useful for various related tasks such as symbolic music generation, music transcription and tagging, or high-level comparative performance analysis.

Table 1 presents an overview of comparable piano performance datasets currently publicly available, for which precise (recorded) MIDI data, score-performance alignments and/or musicological annotations are available. Among these datasets, ASAP [7] stands out as the most extensive one, both in terms of musical pieces and performer range, with 1,068 performances beat-aligned to 222 scores, each annotated with key and time signature. In comparison to ASAP, all other publicly accessible datasets are significantly smaller: The Vienna 4x22 corpus [8] contains 22 different performances for excerpts of four different pieces, each aligned on a note level and provided in MusicXML³, MIDI and audio format. The CrestMuse PEDB v2.0 [9] provides 35 pieces note-aligned to 411 performances, with scores provided in MusicXML and MIDI and performances in MIDI and WAV. The dataset also contains phrase structure annotations, however, merely in the format of PDF and plain text files, somewhat limiting their (re)usability.

The MazurkaBL dataset [10] consists of a corpus of 44

Chopin Mazurkas with MusicXML scores that have been beat-aligned to 2000 performances. The performances themselves are not provided (neither as MIDI nor as audio); only beat positions and corresponding loudness values are given, along with the positions of tempo/dynamics markings in the score.

3. CURATION PROTOCOL AND FILE FORMATS

3.1 File origins

The MIDI performance files originate from a performance of twelve Mozart piano sonatas by Viennese concert pianist Roland Batik on a computer-controlled Bösendorfer SE290 grand piano, the predecessor of the CEUS model. The Bösendorfer SE series measures each individual keystroke and pedal movement precisely, with onset and offset times being captured at a time resolution of 1.25ms. Hammer velocity values are captured in a proprietary file format, and converted and mapped to the 128 dynamics MIDI values (see [13] for conversion details). The audio recordings corresponding to those MIDI files can be purchased commercially⁴.

These MIDI performance data were originally aligned manually, on a note-to-note level, to a symbolic encoding of the score produced by our team [14, 15]. In order to make it possible to link the performance data in an unequivocal way to the musicological score annotations provided in the *Annotated Mozart Sonatas* dataset by Hentschel et al. [1], we decided to replace our score encoding in the alignments entirely by the score notes as given in the their dataset, which link to their annotations directly via absolute temporal score position. The scores in the *Annotated Mozart Sonatas* dataset conform to the New Mozart Edition⁵ and are given in MuseScore format, with the harmony, phrase and cadence label annotations provided in tabular format, as tab-separated values (TSV) files.

3.2 The match alignment format

We provide the alignment between the above-mentioned score and performance files in the match file format [16], a file format for symbolic music alignment in a human-understandable textual form. It is structured sequentially, and the alignment information is given at the level of individual notes.

⁴ <https://www.gramola.at/products/9003643987012>

⁵ <https://dme.mozarteum.at/DME/nma/start.php?l=2>

³ <https://www.musicxml.com/>

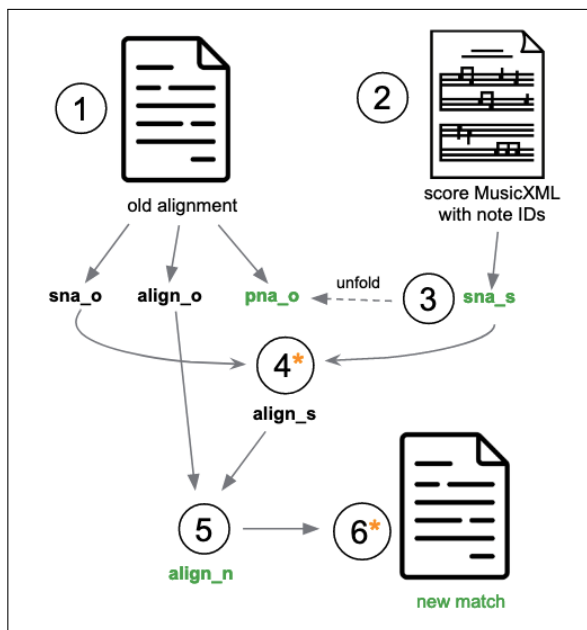


Figure 1. Visual illustration of the alignment process. Each step in the alignment process is numbered according to the textual description in Section 3.3. Steps marked * indicate manual correction / post-processing. Elements highlighted in green are combined in the new alignment match files.

The encoded alignment is complete in the sense that all performance and all score notes are captured. Each performance and each score note is represented with their respective note ID, and their respective alignment can be recorded with one out of three potential tuples: 1. A *match* between score note and performance note, i.e., $(score_id, performance_id)$, 2. a *deleted* score note $(score_id,)$ which represents a score note omitted in the performance, or 3. an *inserted* performance note $(, performance_id)$, which marks a performed note for which there is no corresponding score note.

Following this alignment encoding, each line in a match file corresponds to either a *match*, a *deletion* or an *insertion*. Additional lines express (sustain or soft) pedal information, or encode meta information about the musical piece and performer. While the performance part in match corresponds to a lossless encoding of a corresponding performance in MIDI format, the score part captures essential information including onset, offset and duration in beats, and pitch, pitch spelling, and octave information for each score note.

3.3 Curation protocol

To create note-level score-to-performance alignments, encoded in the match file format, between the performance MIDI data by pianist Roland Batik and scores and musicological annotations by Hentschel et al. [1], we follow the workflow as outlined below (see Fig. 1):

1. **Retrieve information from old alignment.** Given an old alignment file, we use `partitura` [17] to re-

trieve a score and performance representation which we parse into score and performance note arrays, `sna_o` and `pna_o`, to sequentially capture each (notated and performed) note with a unique note ID. In addition we retrieve a score-to-performance alignment, `align_o`, in the encoding format explained above (i.e., a list of note ID tuples expressing either a match, deletion or insertion).

2. **Retrieve score note array from MusicXML.** In the next step, we convert the annotated MuseScore format scores provided by Hentschel et al. [1] to MusicXML, assign unique note IDs to each note, and convert this score representation into a second score note array (`sna_s`).
3. **Unfold score note array.** We update the score note array obtained from MusicXML, `sna_s`, by unfolding it in accordance to the repetition structure found in the performance note array, `pna_o`.⁶
4. **Create score-score alignment.** In this step, we create a score-to-score alignment (`align_s`) by matching each note in the two score note arrays `sna_o` and `sna_s` using its pitch, onset and duration information in beats. Any notes in `sna_o` and `sna_s` not matched automatically need to be aligned manually. Missed alignments at this stage can occur due to:

- **Score mistakes.** These reflect mistakes in the score (e.g., a missing note, incorrect pitch, octave, missing modifier, missing repetition or ending markings) and require a manual correction of the score file.
- **Differing score versions.** For certain sonata movements, the notated score provides an alternative score version reflecting the first edition (“Erstdruck”) for certain segments of a piece, expressing the composer’s impromptu ornamentation.⁷ For the current dataset, such ornamented versions exist in K.284iii, K.332ii, K.457iii.
- **Double-voiced score notes.** These occur frequently in notated music, and describe a score note that is notated doubly in two different voices but corresponds to one performed note.
- **Grace notes.** Grace notes in notated music can occur in multiple forms to reflect different types of ornaments such as trills, acciaccature, mordents, turns etc. Depending on the ornament type and the underlying score encoding format, this may result in several notes occurring at the same (notated) onset (and hence

⁶ To reflect the same note occurring in a repeated segment, a suffix is added to the ID to reflect the number of occurrence, i.e. for a note with ID `n14`, the repeat structure unfolding is expressed as `n14-1` for the first, and `n14-2` for the second occurrence, respectively.

⁷ <https://www.henle.de/en/music-column/mozart-piano-sonatas/>



Figure 2. An example of a cadenza within a piano sonata starting in measure 198 in KV333, 3rd movement.

with zero duration) to ensure a regular measure according to the time signature of that piece. Without onset and duration information, these notes must then be manually aligned to their corresponding performed notes.

- **Cadenza and *ad libitum* measures.** Both cadenza measures and those marked *ad libitum* correspond to irregular measures, that is, measures that contain more beats than indicated in the time signature (see Fig. 2). Digitally encoded, the notes in such measures are commonly notated without duration to allow for error-free parsing, and thus share the same beat onset and need to be aligned manually.

5. Update score-performance alignment. Here we update the score note IDs in the old alignment (`align_o`) according to the score-score alignment (`align_s`) to create new score-performance alignments, `align_n`. For each alignment in `align_o`, we then need to ensure the validity of the original alignment type (match, insertion or deletion). In particular, for notes in the original score note array (`sna_o`) that could not be aligned to notes in the MusicXML-based score note array (`sna_s`), we consider two cases:

- If the note in `sna_o` corresponds to type ‘match’ in `align_o`, the alignment type for the formerly matched performance note is changed accordingly into an insertion.
- If the note in `sna_o` corresponds to type ‘deletion’ in `align_o` (i.e., a score note that was not performed), it is discarded in `align_n`.

Notes in `sna_s` that could not be aligned with notes in `sna_o`, on the other hand, are recorded as type ‘deletion’ in `align_n`.

6. Create match files. Using the updated performance-to-score alignment `align_n`, we

create new match files, and manually add attributional information (e.g., ‘diff_score_version’, ‘voice_overlap’) to score notes to reflect edge cases described in step 4.

4. DATASET OVERVIEW

The *Batik-plays-Mozart* dataset contains performances by pianist Roland Batik of twelve Mozart sonatas (see Table 2 for the list of sonatas), corresponding to approx. 102,400 played notes and 223 minutes of music, for which the performances are provided in MIDI, musical scores in MusicXML, and the alignment in match file format. Approximately 98,300 (95.36%) of all performed notes are aligned with a corresponding score note, the remaining 4,100 (4.44%) represent insertions (reflecting mostly ornaments). Roughly 200 score notes have been omitted in the performances.

For each performance, we also provide the performance note arrays, which capture each played note with its note ID along with onset and duration information in seconds and MIDI ticks, as well as velocity and pitch information. Likewise, the dataset includes the score note array (unfolded according to the repeats as played by the pianist and reflected in the alignment), which captures each score note with its (MusicXML) note ID (including repeat suffixes, where applicable), onset and duration information in terms of beats (reflecting the time signature), and quarter notes (reflecting a “normalized” score time unit), and pitch and voice information.

We link our aligned score note arrays to the musicological annotations in [1] via their temporal position in the following way: In the second version⁸ of the dataset, each annotation label for harmonies, cadences, and phrases is unequivocally referenced to a temporal score position represented in terms of quarterbeats and measure number, where the first expresses the distance of the label from the beginning of the piece in quarter note units. We leverage these two temporal parameters to link each note-aligned score note array by first reducing it to its shortest form (without any unfolded repeats), aligning it temporally with the musicological annotations, and eventually unfolding it according to the performed repetition structure.

5. DATASET DEMONSTRATIONS

This section presents two simple examples of the kinds of studies that are made possible by our dataset. The first is motivated by a directly related study in the Annotated Mozart Sonata corpus paper [1]; the second shows how precise performance alignments permit more detailed investigations relating to cadences and their performance.

5.1 Global tempo and harmonic density

In a first study, we replicate the second experiment in Hentschel et al. [1], aimed at investigating the relationship

⁸ https://github.com/DCMLab/mozart_piano_sonatas

Sonata	Performed Notes	Duration (min)	Match Notes	%	Insertion Notes	%	Deletion Notes	%
KV279	7,789	16.21	7,385	94.087	404	5.780	11	0.130
KV280	6,277	14.69	6,070	95.793	207	3.983	13	0.223
KV281	7,030	14.43	6,396	90.450	634	9.393	11	0.160
KV282	5,761	14.77	5,552	96.197	209	3.467	20	0.337
KV283	8,231	17.39	7,915	95.657	316	4.233	9	0.107
KV284	13,386	25.92	12,691	93.763	695	6.033	27	0.203
KV330	7,869	18.47	7,589	96.857	280	3.047	7	0.100
KV331	11,760	22.64	11,595	98.283	165	1.370	45	0.347
KV332	9,013	17.84	8,660	93.417	353	6.210	24	0.373
KV333	9,137	20.40	8,827	96.723	310	3.120	16	0.157
KV457	7,290	18.24	7,022	96.043	268	3.843	9	0.110
KV533	8,878	22.12	8,616	97.027	262	2.837	15	0.137
Total	102,421	223.12	98,318	95.358	4,103	4.443	207	0.199

Table 2. List of sonatas in the *Batik-plays-Mozart* dataset. The bottom row represents the sum in all columns except for those expressing percentages, for which the mean is shown.

between tempo and harmonic change rate. The basic question asked in [1] was whether the rate at which the harmony changes in a piece is correlated with the piece’s typical performance tempo. Their study involved determining the average (median) performance duration of each sonata movement from 6 complete commercial sonata recordings, and correlating harmonic label density (rate of harmonic labels in their annotations, per performance time unit) with average overall performance tempo (number of quarter notes per performance time unit). We repeat the same experiment with our pianist’s performances and our alignment files instead of 6 pianists’ audio recordings.

We apply the same procedure as in [1], unfolding the score according to the repeat structure of the piece in order to calculate the actual piece length (in terms of quarter notes). The only difference is that we do this according to the repeats actually performed by the pianist (which are expressed in our match files, thus omitting the need for a dedicated “unfolding” step), whereas [1] seem to have assumed that all repeats were played by all pianists.

Comparing our results (Fig. 3) to Fig. 10 in [1], we see a similar general trend, in the form of a roughly linear increase in harmonic label density with performed tempo (slope = .43, $r = .75$, compared to .48 and .80, respectively, in [1]).⁹ However, we also immediately see a marked difference in the performance tempo distribution: in [1], Fig. 10, there is a relatively large cloud of points (sonata movements) with conspicuously high tempos of 180–200 (quarters per minute), which does not appear in our plot, and which we believe may point to a systematic problem in their way of estimating playing tempo: assuming that all notated repeats are played out by the performers leads them to overestimate the tempo in all cases where some or a majority skipped some repeats.¹⁰

⁹ Note that we have a somewhat smaller set of points, because we only have 12 of the 18 sonatas in our dataset.

¹⁰ Of course, the authors explicitly acknowledge the problem: “Also, some of the initial assumptions might have to be revisited. For example, the extreme outlier suggesting a tempo of 239 quarter notes per minute is due to the fact that for this particular piece – the first movement of K. 533/494 – there seems to be a convention among pianists to repeat the first part of the piece, but not the second (as the score would suggest), which of course reduces the performance duration.” [1] (p.76), but a comparison

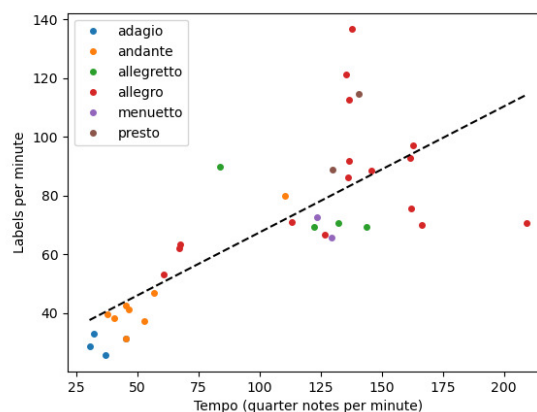


Figure 3. Correlation between global tempo (as measured in quarter notes per minute) and harmony label density

We thus see an immediate advantage of our more precise performance-aligned corpus: the match files naturally give correct tempo and score duration information, being based as they are on score-performance alignments that reflect the actual repeat structure played by our performer. Still, we can say that our results support and confirm the overall hypotheses proposed there, showing a more or less linear relationship between harmonic label density and global performance tempo.

5.2 Performance of different cadence types

Our data permits much more detailed investigations into relationships between structural aspects of a piece, and how these are translated into performance decisions by a pianist. As a simple example, we investigate variations in local tempo before various types of cadences. Specifically, we compare the local tempo prior to a cadence annotation across different tempo classes for authentic (perfect and imperfect, i.e., PAC and IAC) and half cadences (HC), and differentiate between the cases when a cadence falls on either a downbeat or a weak beat. The hypothesis to be tested

with our distribution implies it might be more severe than expected.

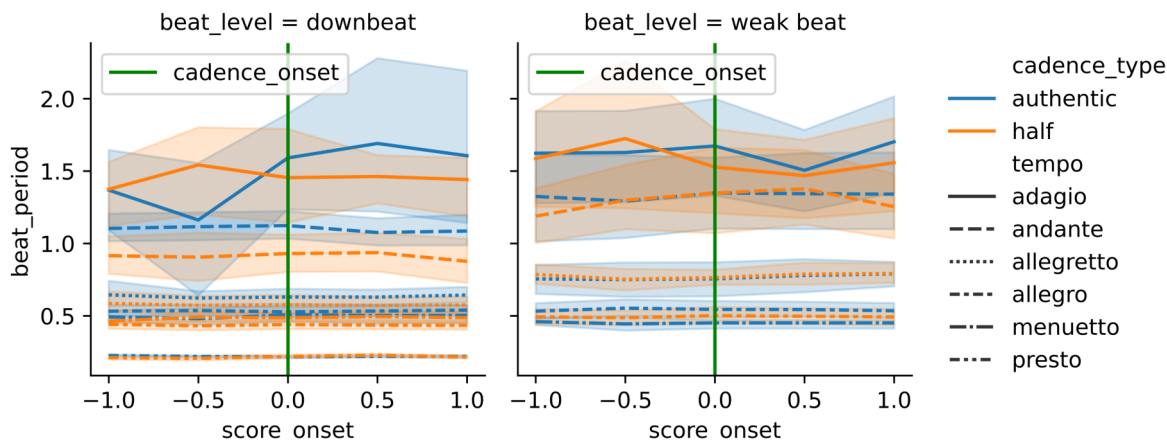


Figure 4. Comparison of local timing strategies one quarter note before and after authentic and half cadence labels, over different tempo classes (in increasing tempo from top to bottom), for cadences falling on a downbeat (left) or weak beat (right). Colour identifies cadence type, line style notated tempo class.

here is that a performer will tend to shape cadences differently, in terms of tempo, depending on their type and degree of ‘finality’.

To compute the local tempo curves, we consider a uniform window spanning one quarter note each preceding and following a cadence label.¹¹ For each score-note-aligned performed note in that window, we define the local tempo via the *beat period* (*BP*), which we calculate as the ratio of the inter-onset-interval (IOI) between the current *performed* onset and the subsequent one, and the IOI between the current *notated* onset and subsequent one. We exclude grace notes and their corresponding performed notes from this calculation in order to remove outliers.

Next, we perform time-wise interpolation on these tempo curves to obtain beat period values at eighth note intervals within the window. Given that we are most interested in the local timing strategy immediately before a label (that is, an eighth note before the label position), we discard those curves where that particular time point is interpolated. Following this procedure, we obtain a total of 3,540 local tempo values (corresponding to 708 curves), of which 251 (7.09%) values are interpolated.

Figure 4 shows the mean of local tempo curves across different tempo classes, for cadence labels annotated on a downbeat (left) and on a weak beat (right), respectively. For both authentic and half cadence types, the differences in local tempo diminish with increasing global tempo for both downbeat and weak beat cadences. Likewise, the tempo profiles tend to flatten out with increasing global tempo, suggesting that the pianist takes more liberty, in terms of expressive timing, in slow pieces. For this reason, we focus our analysis on the *adagio* tempo class, the slowest tempo (the solid line plots in Fig. 4).

The influence of the beat level on the local tempo for half cadences seems to be negligible, with the local beat period decreasing slightly prior to the cadence (causing an

increase in local tempo, i.e. an *accelerando*), regardless of whether it falls on a downbeat or weak beat. For authentic cadences, we can see a substantial difference in expressive tempo depending on whether or not the label falls on a downbeat: for authentic cadences falling on a downbeat, the mean tempo curve for the *adagio* tempo class corresponds mostly to what one would expect (i.e., a very clear *ritardando* in preparation of the cadence) based on the underlying harmonies and their notion of tension and release. Interestingly, this *ritard* seems to continue somewhat after the resolution into the tonic, suggesting a lengthening of the tonic arrival. For weak-beat authentic cadences, a similar significant preparation or anticipation is largely missing.

6. CONCLUSION AND FUTURE WORK

We have presented *Batik-plays-Mozart*, a piano performance dataset linking professional Mozart piano sonata performances to expert-labelled musical scores, at the level of notes. The resulting dataset is the first of its kind to combine professional performances in precise, recorded MIDI with curated musical scores and expert musicological and structural annotations [1] at this level of detail.

We presented two preliminary experiments, intended to demonstrate the benefits of having such precise, note-aligned performance–score–structure annotation data for studying expressive features and their relation to the underlying musical structure.

Our plan for future work includes the transcription of the remaining six sonatas of the Mozart piano sonatas corpus from audio recordings by the same pianist, and their subsequent alignment to the musical scores using state-of-the-art transcription and alignment models. By doing so, we hope to advance our understanding of the differences between transcribed and recorded MIDI, and to evaluate the potential benefits of incorporating an alignment step to improve the quality of transcription.

¹¹ In the *Annotated Mozart Sonatas Corpus* [1], cadence labels are placed at the onset of the final target harmony (e.g., *I/i* for authentic cadences).

7. ACKNOWLEDGMENTS

We wish to express our gratitude to pianist Roland Batik for his gracious permission to publish the detailed measurements of his performances. We also want to thank the authors of the *Annotated Mozart Sonatas Corpus* for their tremendous efforts, and for permitting us to link our data to theirs. This work receives funding from the European Research Council (ERC), under the European Union’s Horizon 2020 research and innovation programme, grant agreement No. 101019375 (*Whither Music?*). The LIT AI Lab is supported by the Federal State of Upper Austria.

8. REFERENCES

- [1] J. Hentschel, M. Neuwirth, and M. Rohrmeier, “The Annotated Mozart Sonatas: Score, Harmony, and Cadence,” *Transactions of the International Society for Music Information Retrieval (TISMIR)*, vol. 4, no. 1, pp. 67–80, 2021. [Online]. Available: <https://doi.org/10.5334/tismir.63>
- [2] C. E. Cancino-Chacón, M. Grachten, W. Goebel, and G. Widmer, “Computational Models of Expressive Music Performance: A Comprehensive and Critical Review,” *Frontiers in Digital Humanities*, p. 25, 2018.
- [3] D. Jeong, T. Kwon, Y. Kim, K. Lee, and J. Nam, “VirtuosoNet: A Hierarchical RNN-based System for Modeling Expressive Piano Performance.” in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2019, pp. 908–915.
- [4] A. Maezawa, K. Yamamoto, and T. Fujishima, “Rendering Music Performance with Interpretation Variations using Conditional Variational RNN,” in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2018.
- [5] H. Zhang, J. Tang, S. R. M. Rafee, and S. D. G. Fazekas, “ATEPP: A Dataset of Automatically Transcribed Expressive Piano Performance,” in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2022.
- [6] Q. Kong, B. Li, J. Chen, and Y. Wang, “GiantMIDI-Piano: A Large- Scale MIDI Dataset for Classical Piano Music,” *Transactions of the International Society for Music Information Retrieval (TISMIR)*, May 2022.
- [7] F. Foscarin, A. Mcleod, P. Rigaux, F. Jacquemard, and M. Sakai, “ASAP: A Dataset of Aligned Scores and Performances for Piano Transcription,” in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2020, pp. 534–541.
- [8] W. Goebel. (1999) The Vienna 4x22 Piano Corpus. [Online]. Available: <http://dx.doi.org/10.21939/4X22>
- [9] M. Hashida, E. Nakamura, and H. Katayose, “Crest-MusePEDB 2nd Edition: Music Performance Database with Phrase Information,” in *Proceedings of the 15th Sound and Music Computing (SMC) Conference*, 2018.
- [10] K. Kosta, O. F. Bandtlow, and E. Chew, “MazurkaBL: Score-Aligned Loudness, Beat, Expressive Markings Data for 2000 Chopin Mazurka Recordings,” in *Proceedings of the Fourth International Conference on Technologies for Music Notation and Representation (TENOR) (Montreal, QC)*, 2018, pp. 85–94.
- [11] C. Hawthorne, A. Stasyuk, A. Roberts, I. Simon, C.-Z. A. Huang, S. Dieleman, E. Elsen, J. Engel, and D. Eck, “Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset,” in *International Conference on Learning Representations*, 2019. [Online]. Available: <https://openreview.net/forum?id=r11YRjC9F7>
- [12] A. Ycart, L. Liu, E. Benetos, and M. Pearce, “Investigating the Perceptual Validity of Evaluation Metrics for Automatic Piano Music Transcription,” *Transactions of the International Society for Music Information Retrieval (TISMIR)*, 2020.
- [13] W. Goebel and R. Bresin, “Measurement and Reproduction Accuracy of computer-controlled Grand Pianos,” *The Journal of the Acoustical Society of America*, vol. 114, no. 4, pp. 2273–2283, 2003.
- [14] G. Widmer, “Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries,” *Artificial Intelligence*, vol. 146, no. 2, pp. 129–148, 2003.
- [15] E. Cambouropoulos, “From MIDI to traditional musical notation,” in *Proceedings of the AAAI Workshop on Artificial Intelligence and Music: Towards Formal Models for Composition, Performance and Analysis*, vol. 30, 2000.
- [16] F. Foscarin, E. Karystinaios, S. D. Peter, C. Cancino-Chacón, M. Grachten, and G. Widmer, “The match file format: Encoding Alignments between Scores and Performances,” in *Proceedings of the Music Encoding Conference (MEC)*, 2022.
- [17] C. Cancino-Chacón, S. D. Peter, E. Karystinaios, F. Foscarin, M. Grachten, and G. Widmer, “Partitura: A Python Package for Symbolic Music Processing,” in *Proceedings of the Music Encoding Conference (MEC)*, 2022.