ENABLING EMPIRICAL ANALYSIS OF PIANO PERFORMANCE REHEARSAL WITH THE RACH3 MIDI DATASET

Alia Morsi[1*] **Suhit Chiruthapudi**[2*] **Silvan Peter**[2] **Ivan Pilkov**[2]
Laura Bishop[3] **Akira Maezawa**[4] **Xavier Serra**[1] **Carlos Cancino-Chacón**[2]

[1] Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[2] Institute of Computational Perception, Johannes Kepler University Linz, Austria
[3] RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Norway
[4] Yamaha Corporation, Hamamatsu, Japan

alia.morsi@upf.edu, carlos.cancino_chacon@jku.at

ABSTRACT

The study of piano rehearsals can offer interesting insights into the strategies adopted by a pianist in order to learn, interpret and eventually perform musical pieces. The analysis of rehearsal processes requires computational methods that differ from those used for piano performance, due to challenges like mistakes, repetitions of musical segments, or forward and backward skips to sections in the piece. The scarcity of publicly available rehearsal data limits the empirical understanding of these challenges. We release the Rach3 MIDI Dataset, an openly available collection of MIDI files containing more than 750 hours of recordings of piano rehearsals and corresponding MusicXML scores by four pianists (3 advanced, 1 beginner), collected over a period of more than 4 years. This dataset records the progression of pianists learning new repertoire, as well as practicing familiar pieces, all in the Western Classical tradition. We describe the rehearsal piece identification process used for automatically labeling a portion of the data in this release. Furthermore, we use the Rach3 data to highlight several challenges and future research directions pertaining to the computational analysis of piano rehearsals, specifically symbolic rehearsal-to-score alignment, rehearsal structure analysis, and automatic mistake identification.

1. INTRODUCTION

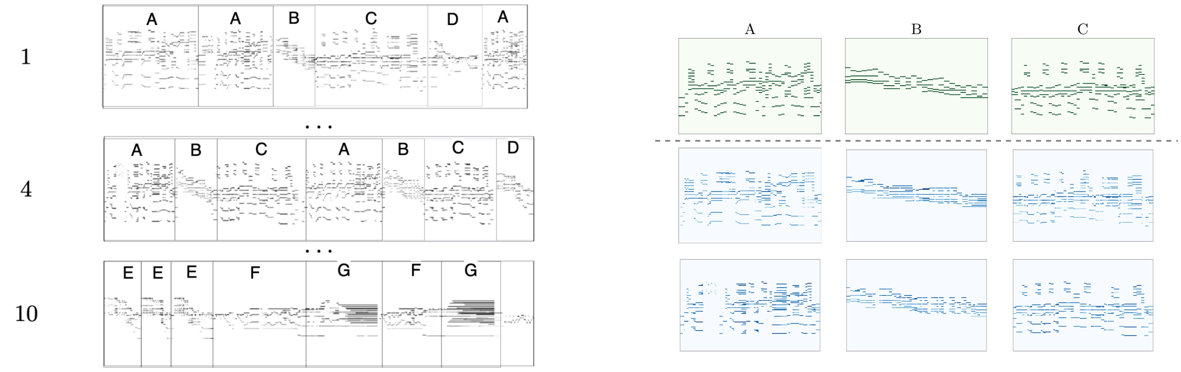
Computational analysis of music performance has traditionally focused on the end product, that is, the outcome of a rehearsal process, rather than rehearsal itself. Yet musicians spend substantial time on rehearsal. Analysis of rehearsal has the potential to improve understanding of music learning and expertise development and support the development of pedagogic tools. We define *rehearsal* as goal-oriented, systematic practice with the aim of learning and becoming proficient in playing specific repertoire. While the same families of music analysis approaches are applicable for data from either performances or rehearsals, rehearsal data poses specific challenges that are not present in polished performance. The most notable examples are the presence of mistakes, jumps between different parts of a piece, and non-compositional repetitions (i.e., playing the same passage repeatedly).

To date, research on music rehearsal has been hindered by a lack of data and appropriate computational tools. For advanced musicians, rehearsal is a process that can span months or years, and understanding that process requires a longitudinal perspective, with data collected at different stages. This paper introduces the Rach3 MIDI dataset, which contains more than 750 hours of recordings of piano rehearsals by four pianists, mostly involving music from the Western classical tradition. The dataset allows for a comprehensive and ecologically valid computational, data-driven analysis of piano rehearsal over an extended period, which has been limited in previous research due to technical constraints and data availability (cf., the scale and scope of the studies by Chaffin and colleagues [1, 2]). The dataset will be made publicly available, and, to the best of our knowledge, comprises the largest collection of piano rehearsal data. Existing symbolic datasets for analysis of piano performance (e.g., (n)ASAP [3], Vienna 4x22 [4] and Batik [5]), focus on polished performances.

The Rach3 MIDI dataset will contribute to rehearsal research by enabling systematic study of rehearsal decisions. As musicians develop expertise on their instrument, they also develop more effective rehearsal strategies. Beginners are more likely to repeat individual notes, whereas more experienced musicians tend to repeat musically coherent sections or measures [6]. Among musicians of the same level, some organize their rehearsal sessions according to learning goals, for example, focusing separately on technical challenges and musical understanding, while others work in a more undifferentiated way [7]. Figure 1 shows

* Equal contribution.

 Attribution: A. Morsi, S. Chiruthapudi, S. Peter, I. Pilkov, L. Bishop, A. Maezawa, X. Serra and C. Cancino-Chacón, “Enabling Empirical Analysis of Piano Performance Rehearsal with the Rach3 MIDI Dataset”, in *Proc. of the 26th Int. Society for Music Information Retrieval Conf.,* Daejeon, South Korea, 2025.



(a) Changing rehearsal structure across time for Pianist 1 rehearsing Rapsodia Mexicana No. 2 by Manuel Ponce. Rehearsal number indicated on the left.

(b) Variations of music segments A, B, and C within and across rehearsals. Green: MIDI score reference. Blue: Performance instances across rehearsals.

Figure 1: Evolution of practice structure across the different phases of learning a piece.

a pianist's rehearsal structured into segments that reflect changes in focus over time.

This paper describes data collection (Section 3) and automatic labeling of rehearsal files using fingerprinting methods (Section 4). We highlight limitations of state-of-the-art symbolic performance-to-score alignment for rehearsal data (Section 5) and propose an alternative approach for rehearsal structure analysis inspired by pattern discovery and music structure segmentation (Section 6), including preliminary attempts at automatic score-independent piano mistake identification (Section 7). We conclude with future research directions in the computational analysis of piano rehearsals (Section 8). [1]

2. RELATED WORK

Research on music rehearsal has been sparse to date, though a few studies have examined rehearsal behaviors like decision-making, goal-setting, and practice strategies [8–10]. Ericsson et al. highlighted the role of deliberate practice in achieving expertise [11]. Studies by Hallam [12,13], Sokolovskis [14] and Chaffin [1,2,15–17] investigated rehearsal through observations and retrospective accounts, tracking how practice strategies evolve over time. The broader literature on musical learning has examined how different practice schedules affect memory for pitch and timing [18, 19]. Some research has also investigated how visual attention (eye gaze) is split between the score and the hands during learning of piano pieces, and how this is affected by the music structure [20].

Despite this, rehearsal remains understudied in a data-driven way, with much of the literature based on case studies. As noted in Miksza's review [9], no studies have involved more than 40 hours of rehearsal recordings (see Table 1 in [9]). This is partly due to technical and logistical limitations in capturing long-term rehearsal data and the lack of efficient algorithms to extract relevant information and patterns from such a large source of data. Winters et al. [21] introduced an audio-based method for automatic practice logging, to keep track of which pieces were performed during a rehearsal session. Tools have also been developed for the automatic quantitative assessment of performance quality [22, 23].

3. RACH3 MIDI DATASET

The Rach3 MIDI dataset contains over 3,000 MIDI files from piano rehearsals performed mostly on acoustic pianos equipped with systems to enable MIDI capturing. The dataset aims to be *representative* of typical rehearsal practices, ensure *ecological validity* by reflecting natural rehearsal conditions, and remain *comprehensive in scope* through diverse (i.e., multimodal) data sources for quantitative and qualitative analysis [24]. The full Rach3 dataset is a multimodal dataset that includes synchronized audio (captured with microphones), MIDI, video from a camera positioned over the keyboard, and written logs about practice strategies and focus. This paper focuses solely on the MIDI data and other modalities will be addressed and released in future publications.

Data collection began in Fall 2020 and now includes over 750 hours of recorded rehearsal sessions from four pianists (three advanced, one beginner; three of the pianists are co-authors on this paper), making this the largest synchronized piano MIDI dataset to date, 3.9 times larger than the MAESTRO dataset (see Table 1). Figure 2 shows a cumulative distribution of the performed notes and duration over time. The advanced pianists average 12.7 ± 11.2 years of formal training at the conservatory level, with Pianists 1 and 2 holding undergraduate or conservatory degrees in piano performance. Pianist 3, a beginner, started lessons as part of the project in Summer 2024. Pianist 4 has undergraduate-level training in piano performance.

Rehearsals are conducted on acoustic pianos equipped with Silent systems, allowing for MIDI capture while preserving the natural acoustic sound via condenser microphones. Pianist 1 uses a Yamaha GB1K Silent, Pianist 2 an Essex EUP-116E, and Pianist 4 a Yamaha Disklavier

[1] The dataset can be downloaded from the companion website `https://r3midi.rach3project.com/` where further examples and visualizations are available.

Pianist	Total hours	Total notes (millions)	Avg. hours per session
All	769.6	20.9	0.94
P1	487.1	14.8	1.01
P2	142.4	3.6	0.85
P3	38.9	1.0	0.69
P4	101.2	1.5	0.93
MAESTRO v3	198.7	7.0	–
Batik	3.0	0.1	–

Table 1: Size comparison of piano-centric datasets with synchronized MIDI and audio.

C1X. Pianist 3 records on a Yamaha Clavinova digital piano, with the volume slider kept fixed on their teacher's recommendation.

Pianists organize their rehearsal sessions freely; typical rehearsals include technical warmups (e.g., Hanon exercises, scales) and repertoire practice. The repertoire selection focuses on two areas: learning new pieces from scratch and maintaining previously learned works. This allows for analysis of different rehearsal strategies: initial learning, ongoing maintenance, and relearning. Each advanced pianist focuses primarily (though not exclusively) on specific repertoire: Pianist 1 on Rachmaninoff's *Piano Concerto No. 3, Op. 30*, Pianist 2 on Grieg's *Piano Concerto*, and Pianist 4 on Beethoven's Piano Sonatas. For practical reasons, contributing pianists concentrate on music from the Common Practice Period. [2] Over 100 pieces have been played (counting individual movements separately). The dataset includes rehearsal of some four-hands piano duets. For these, each part (primo and secondo) is counted as a separate piece.

In addition to MIDI, the dataset includes MusicXML scores for the performed works. Most scores were sourced from MuseScore; [3] where unavailable, we created them manually using MuseScore based on printed editions or IMSLP [4] scans (preliminary tests with OMR were unsuccessful for the complex piano works included in the dataset). This manual score entry is ongoing, with over half of the dataset currently covered. A full repertoire list is provided in the Appendix. [5]

The dataset also includes live performances from the *Dress Rehearsal R3cital Series*, where contributing pianists perform for a small live and online audience. This series serves to (1) provide a realistic goal for the rehearsals, (2) contrast rehearsal and concert settings, and (3) simulate real concert conditions using the same multimodal recording setup. Two recitals have been held to date, featuring Pianist 1 performing works by Manuel Ponce and Modest Mussorgsky. [6]

This dataset is part of an ongoing research project and will continue to grow through additional performances, annotations, and analysis.

[2] This period corresponds roughly to the Baroque, Classical, Romantic, and early 20th Century periods of Western Classical music.
[3] https://musescore.com
[4] https://imslp.org/wiki/MainPage
[5] See Footnote 1 .
[6] https://r3citals.rach3project.com.

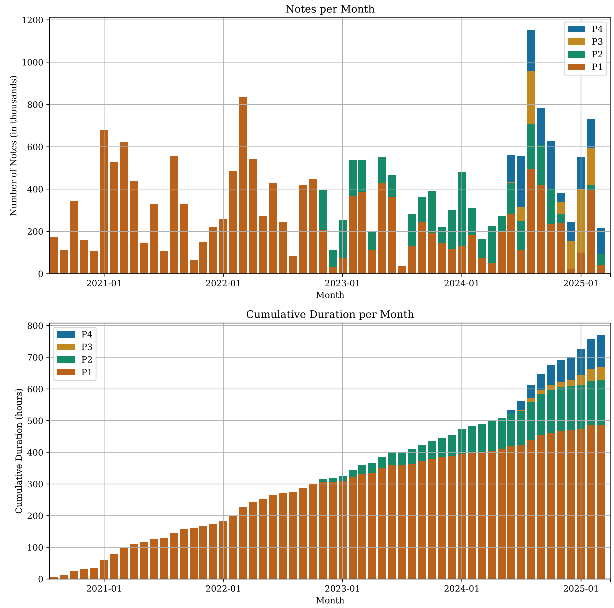


Figure 2: Cumulative distribution of notes and duration in the Rach3 MIDI dataset.

Weightage	Accuracy	Precision	Recall	F1
Macro	0.991	0.941	0.922	0.929
Weighted	0.991	0.989	0.991	0.989

Table 2: Evaluation of symbolic fingerprinting based piece identification.

4. REHEARSAL PIECE IDENTIFICATION

During approximately the first two years of recording rehearsals, multiple pieces were practiced and recorded into a single synchronized MIDI/audio/video take for every rehearsal session (i.e., a MIDI file and its corresponding synchronized audio and video files). Because the cameras were sometimes overheating during long recordings, this process was later modified so that each practiced piece was recorded into a separate MIDI/audio/video take. More recently, pianists started labeling these files according to the piece name. However, almost 60% of rehearsal pieces remained unlabeled, requiring a semi-automatic approach to piece identification.

For this purpose, we followed the symbolic fingerprinting method developed by Arzt et al. [25]. We created three-note tokens from MusicXML scores, generated hashes for these tokens, and stored them in a lookup table, mapping each hash to the corresponding scores. Tokens were then extracted from rehearsal recordings, and their hashes were matched against the lookup table, with the highest-matching score identified as the predicted piece.

We first ran this algorithm on 2155 labeled MIDI files from Pianist 1 and Pianist 2 containing a single piece, and whose respective scores were digitally and publicly available. The lookup table consisted of the hashes of 71 such digital scores. We assessed the algorithm's performance with this labeled data and provide the results in Table 2.

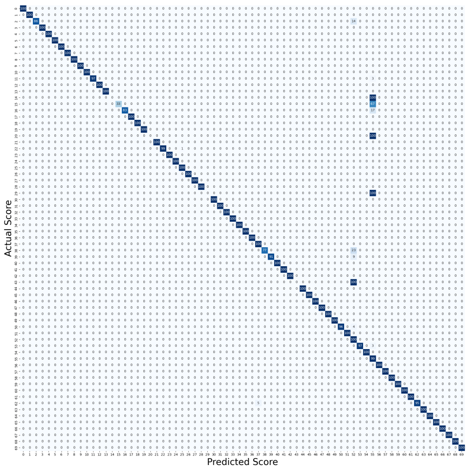


Figure 3: Confusion matrix of the piece identification method applied to 2155 labeled MIDI files from the dataset, yielding an accuracy of 0.99 across 71 scores.

Four pieces were 100% misidentified, all of which appeared in only one MIDI file: Pachulski's Prelude in C Minor Op. 8 No. 1, Grieg's 'Solveig's Song', Mendelssohn's Songs Without Words Op. 30 No. 1 and Chopin's Prelude Op. 28 No. 7. Among the other pieces, there were some misidentifications but most were identified with 100% accuracy. Common reasons for misidentification included: 1) Short rehearsal durations which did not provide enough hashes to comprehensively represent the piece being practiced. 2) Many repetitions of tiny fragments whose hashes could easily belong in other scores. This is especially the case for fragments of chromatic scales that are likely to appear in multiple scores. 3) Many pitch and timing errors, which sometimes occurred in early rehearsals.

In the next step, this algorithm was used to predict the pieces in the remaining unlabeled MIDI files. In the cases where there were multiple pieces within a single MIDI file, a separate pre-processing step was added to segment the MIDI file at points where there was a long silence (>4s), assuming that this is the point where the pianist switches from one piece to another during the rehearsal. The fingerprinting algorithm was then run on these files/segments to predict the piece being played. The pianist's rehearsal log was used to identify the pieces that were played on the given day, and the fingerprinting algorithm searches for hashes corresponding only to the scores of those pieces. Manual review of this process is ongoing.

5. CHALLENGES OF PERFORMANCE-SCORE ALIGNMENT FOR REHEARSAL DATA

Alignment is a crucial first step towards quantitative performance analysis. In symbolic alignment, note-wise alignment refers to the unique matching of individual notes, i.e., a score note may be matched to a single performance note or marked as a deletion, and a performance note may be matched to a single score note or marked as an insertion. Note alignment algorithms do best with a one-to-one correspondence between notes in the score and notes

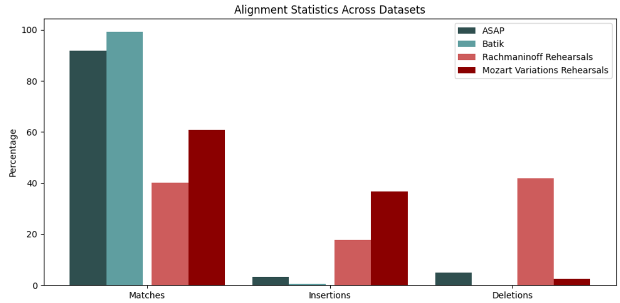


Figure 4: Comparison of the distribution of note matches, insertions and deletions for performances of two pieces in the Rach3 MIDI dataset and the (n)ASAP and Batik datasets

in the performance [3, 26, 27], which is not the case for rehearsal data.

Figure 4 shows the distribution of note matches, insertions and deletions for multiple performances of two pieces in the Rach3 MIDI dataset (Rachmaninoff's Piano Concerto, No. 3, first movement and Mozart's Twelve Variations on "Ah vous dirai-je, Maman", K 265). We compare the number of insertions and deletions for the Rachmaninoff and Mozart pieces to those of Liszt pieces in the (n)ASAP dataset [3] and Mozart Sonatas in the Batik dataset [5]. To make these comparisons, we ran performance-to-score alignments using the GlueNote [28], a state-of-the-art symbolic alignment method that uses learned representations and is claimed to be suitable for alignment in the presence of large mismatches. The figure shows more insertions and deletions in the rehearsal data than in the polished performance data of the (n)ASAP and Batik datasets. If the alignment methods were adequate, we would expect the proportions of insertions, deletions and matches of these two cases to be more similar. We do expect more errors in the rehearsal, but not to the extent shown in the plot. A potential factor in the error rate discrepancy is the extra repetitions occurring during rehearsal.

6. REHEARSAL STRUCTURE ANALYSIS

The goal of computational rehearsal structure analysis is to identify and group equivalent segment repetitions (see Figure 1), yielding insight into how rehearsals are organized. We define *equivalent segments* as sequences of performed notes that correspond to the same score passage, even if performed with different interpretations or mistakes and varying in length. Given a score, performance segments can be linked to corresponding score segments, with all performance segments corresponding to the same score segments treated as equivalent. In the absence of a score, it is necessary to identify and compare segments within a performance.

By considering earlier work on pattern-discovery and music structure segmentation, we conclude that Similarity Matrix approaches are better suited than Translational Equivalence Class (TEC) [29] approaches for this analysis. TEC methods treat music as a spatial arrangement of

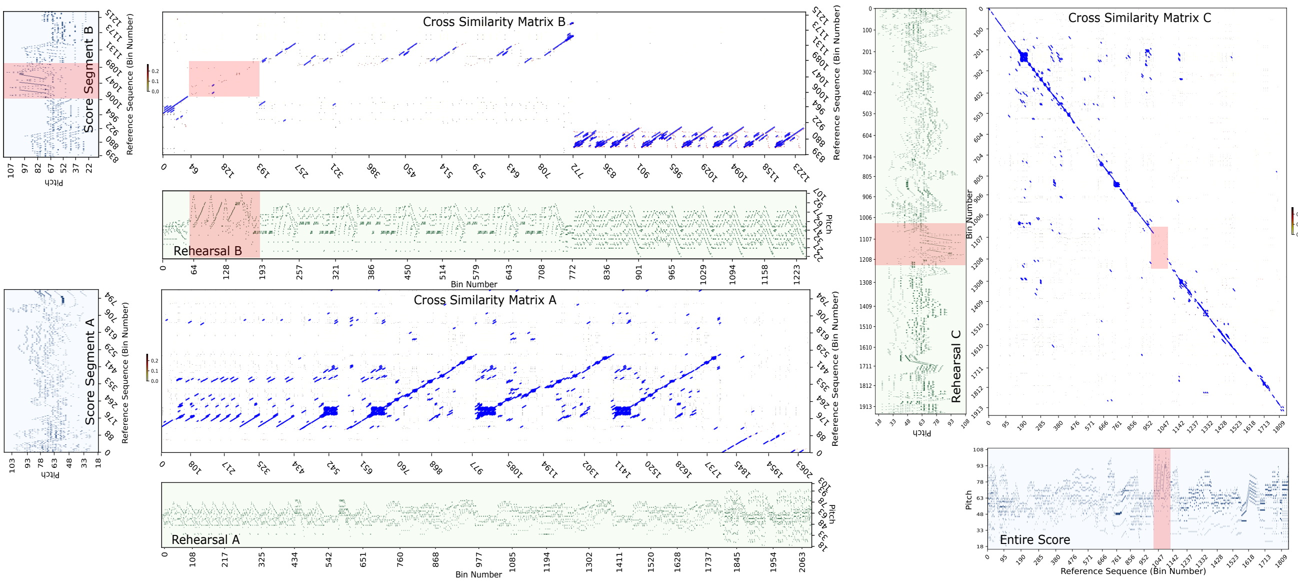


Figure 5: Cross Similarity Matrices with diagonals (blue) showing rehearsal structures for three sessions by Pianist 1 of Rapsodia Mexicana No. 2. Despite the marked discontinuity (red), rehearsal C shows a complete run-through performance through a diagonal from top-left to bottom-right. Rehearsal A demonstrates repetitions of the same score segment growing progressively longer. Rehearsal B shows a different mixed practice: playing a segment once, repeatedly practicing its ending, then returning to an earlier score location. Diagonal breaks in rehearsals B and C result from score ornamentations which result in mismatches between the rehearsal and score chord bins.

events and locate exact repetitions under translation (e.g., transposition, reversal) in multidimensional space. They explicitly search for sets of points that can be transformed into each other via translation and favor exactness of the pitch and time relationships, making them unsuitable given the variability present in rehearsal.

Similarity Matrix approaches accommodate inexact repetitions through flexible processing at multiple pipeline stages. These methods compute pairwise similarity between elements in feature sequences, either between performance and score (Cross Similarity Matrix, CSM) or within a single performance (Self Similarity Matrix, SSM). Related patterns emerge as high-similarity regions within these matrices, specifically as diagonals meeting criteria for minimum length, similarity threshold, and gap tolerance, which are then concatenated and filtered.

We propose an initial system for rehearsal structure analysis based on our definition of equivalent segments, using CSM and SSM to observe different instances of rehearsal structure in the Rach3 MIDI dataset (Figure 5). The pseudocode is available in the Appendix on the companion website.

A ‘chordification’ step handles the temporal variation that occurs in rehearsal contexts. Notes with close onsets (within $\Delta_t = 100$ ms) are grouped into a single ‘chord’ bin (a binary 128-dimensional vector with 1s at active pitch locations). In chord bins, a 1 is only present at a note’s onset location rather than at all locations where it is held, removing the effect of note offset differences. Furthermore, since chord bins are only created when there is at least one note onset, silences are removed from the sequence, eliminating tempo differences due to expressive choices or instability artifacts. The MIDI sequences to be compared (performance or score) are represented as $128 \times n$ matrices.

Instead of measuring similarity between chord bin pairs through Euclidean distances, we propose a more flexible approach. One of the $128 \times n$ chord bin sequences (performance for CSM, score for SSM) is converted to a $128 \times n$ pitch profile sequence, where the i-th pitch profile (p_i) is a probability distribution capturing pitch relationships in the local neighborhood of each active note in the i-th chord bin (b_i). Each p_i is obtained by applying a local smoothing 1D convolution window (w) to the pitch axis of each bin, allowing us to treat each entry $j \in \{0, \ldots, 127\}$ in p_i as the probability of observing pitch j in the context of b_i. Under the assumption that pitches in any bin b_i are binary independent events, the similarity between b_i and any pitch profile p_k is expressed as the likelihood of p_k representing the observed pitches in b_i, formulated as the following Bernoulli likelihood:

$$L(p_k|b_i) = \prod_{j=0}^{127} p_{k,j}^{b_{i,j}} \cdot (1 - p_{k,j})^{1-b_{i,j}} \tag{1}$$

where $p_{k,j}$ represents the probability of pitch j being active in profile k, and $b_{i,j} \in \{0, 1\}$ indicates whether pitch j is observed in chord bin i. Applying this computation for each chord bin i against all pitch profiles p_k (where $k = 0, \ldots, n-1$) constructs the similarity matrix by concatenating the likelihood results for each chord bin. For CSM, we compare performance chord bins with pitch profiles from the score (resulting in an $n_{\text{score}} \times n_{\text{perf}}$ matrix), whereas for SSM, both pitch profiles and chord bins come from the performance (resulting in an $n \times n$ matrix).

To find relevant regions in the similarity matrix, we traverse diagonals to identify those meeting prespecified minimum length, similarity threshold, and gap tolerance parameters. The diagonals are then post-processed by grouping them according to horizontal and vertical overlap ratios and merging groups based on diagonal intersections. The result is groups of diagonals, each reflecting a unique repeated segment.

Figure 5 shows an application of the rehearsal structure analysis described above to three stages of rehearsal. In the first (rehearsal A), the pianist practices from a specific starting point, and gradually extends the practiced segment to include the next section in the piece. Later (rehearsal B), during a different rehearsal session, the pianist repeats a segment multiple times before moving to a second segment elsewhere in the piece, which is also repeated. Finally, rehearsal C focuses on full run-through rehearsals where the goal is to play the piece from start to finish. These tend to happen later in the learning process.

Quantitative evaluation using common pattern discovery metrics [30] is not feasible due to incompatibility with our definition of equivalent segments; annotating a Rach3 MIDI evaluation set is planned for future work. Though simple, this approach is hard to tune, as optimal hyperparameters depend on performance details. In Figure 5, performance B illustrates how an unsuitable Δt for chord bins led to mismatched score and performance segments. Additional similarity matrices (see supplementary materials) show that note insertions cause diagonal offsets, creating extra sub-segment groups. Future work should focus on predicting hyperparameters from MIDI data, exploring alternative chord profiles, and improving diagonal grouping to handle insertion-induced offsets.

7. SCORE-INDEPENDENT AUTOMATIC PIANO MISTAKE IDENTIFICATION

Effective mistake identification systems can improve the processing of music rehearsal data. Information about predicted mistake locations and types can be incorporated into structural analysis or alignment pipelines and enable specialized analysis in these areas. Piano performance mistakes are typically categorized as pitch or rhythm deviations from the score, with performance-to-score alignment serving as the primary identification method. Given the challenges highlighted in Sections 5 and 6, we investigate whether the approach proposed in [31], which trains models for score-independent automatic identification of conspicuous piano performance mistakes, can label regions that might be particularly difficult to process due to mistakes, such as the note additions leading to the offset diagonals in Figure 5.

In [31] a Temporal Convolutional Network (TCN) was trained on a private dataset of mistake-annotated piano performances, including both sight-read and practiced performances. To compensate for limited training data, they pretrained a TCN autoencoder on a different private set of unannotated professional MIDI recordings before fine-tuning with the annotated set. However, this approach yielded only modest improvements, likely due to domain mismatch between professional-grade data and the test set.

Accordingly, we investigate whether we can replicate their experiment and train a score-independent automatic piano mistake detection model. We pre-train their same TCN autoencoder architecture with unlabeled files from the Rach3 MIDI dataset, followed by fine-tuning a final classification layer with labeled synthetic piano mistake data generated with the approach and toolkit in [32]. The synthetic mistakes toolkit applies alterations to mistake-free MIDI performances based on a proposed taxonomy of performance mistakes. It returns a modified version of an input performance with the applied mistakes, and the corresponding annotation file with mistake types and locations. We use the recommended input performance files indicated on the toolkit's webpage. [7] This collection includes actual performances (Vienna 4x22 [4], SMD [33], and 32 files from ASAP [34]) and music scores within the beginner to intermediate proficiency levels. The output mistake labels were summarized into a binary label marking the presence or absence of a mistake at discrete points in the resulting piano roll. We use these data to create train, validation and test splits and used the transcription precision/recall/F1 metrics for evaluation since the estimated and ground-truth annotations can be treated as note events at predefined pitches. Our best training configuration achieved 0.445 average F1-Measure, 0.400 average precision, and 0.528 average recall on the test set (all for synthetic data).

Although initial qualitative observations suggest that model predictions tend to cluster around mistake annotations, the locations are not exact. This imprecision would compromise our ability to rely on such mistake predictions to improve our rehearsal data processing pipelines. Further investigation is needed to determine which synthetic mistake types can be effectively learned, and to extend the toolkit to create mistakes that represent observations from the Rach3 MIDI Dataset, as current parameters are set more heuristically. Furthermore, it is possible to create a small collection of human-annotated mistake data from the Rach3 MIDI dataset to be used for testing.

8. CONCLUSION

This paper introduced the Rach3 MIDI dataset, the largest publicly available collection of piano rehearsal data, recorded over four years with four pianists. It forms part of an ongoing project with future releases planned, including audio and video recordings. Using the dataset, we explored critical computational challenges associated with piano rehearsal analysis, applying state-of-the-art methods in three areas: symbolic rehearsal-to-score alignment, rehearsal structure analysis, and automatic mistake identification. Our findings demonstrate that existing methods require substantial adaptation for rehearsal analysis. The Rach3 dataset provides both the foundation for computational rehearsal analysis and empirical evidence of methodological gaps that must be addressed.

[7] https://github.com/Alia-morsi/piano-synmist

9. ACKNOWLEDGMENTS

This work has been supported by the Austrian Science Fund (FWF), grant agreement PAT 8820923 ("*Rach3: A Computational Approach to Study Piano Rehearsals*"), by the European Research Council (ERC) under the EU's Horizon 2020 research & innovation programme, grant agreement No. 101019375 ("Whither Music?"), by IA y Música: Cátedra en Inteligencia Artificial y Música" (TSI-100929-2023-1), funded by the Secretaría de Estado de Digitalización e Inteligencia Artificial, and the European Union-Next Generation EU, under the program Cátedras ENIA 2022 para la creación de cátedras universidad-empresa en IA, and the Research Council of Norway through its Centres of Excellence scheme, project number 262762.

10. REFERENCES

[1] R. Chaffin and G. Imreh, "Practicing Perfection: Piano Performance as Expert Memory," *Psychological Science*, vol. 13, no. 4, pp. 342–349, Apr. 2005. [Online]. Available: https://www.taylorfrancis.com/books/9781135685461

[2] R. Chaffin, T. Lisboa, T. Logan, and K. T. Begosh, "Preparing for memorized cello performance: the role of performance cues," *Psychology of Music*, vol. 38, no. 1, pp. 3–30, Jan. 2010. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0305735608100377

[3] S. D. Peter, C. E. Cancino-Chacón, F. Foscarin, A. P. McLeod, F. Henkel, E. Karystinaios, and G. Widmer, "Automatic Note-Level Score-to-Performance Alignments in the ASAP Dataset," *Transactions of the International Society for Music Information Retrieval*, vol. 6, no. 1, pp. 27–42, Jun. 2023. [Online]. Available: http://transactions.ismir.net/articles/10.5334/tismir.149/

[4] W. Goebl, "The Vienna 4x22 Piano Corpus," 1999. [Online]. Available: https://doi.org/10.21939/4X22

[5] P. Hu and G. Widmer, "The batik-plays-mozart corpus: Linking performance to score to musicological annotations," in *Proceedings of the 24th International Society for Music Information Retrieval Conference, ISMIR 2023, Milan, Italy, November 5-9, 2023*, A. Sarti, F. Antonacci, M. Sandler, P. Bestagini, S. Dixon, B. Liang, G. Richard, and J. Pauwels, Eds., 2023, pp. 297–303. [Online]. Available: https://doi.org/10.5281/zenodo.10265283

[6] L. M. Gruson, "Rehearsal skill and musical competence: does practice make perfect?" in *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*. Oxford University Press, 01 2001. [Online]. Available: https://doi.org/10.1093/acprof:oso/9780198508465.003.0005

[7] K. Miklaszewski, "A case study of a pianist preparing a musical performance," *Psychology of Music*, vol. 17, no. 2, pp. 95–109, 1989. [Online]. Available: https://doi.org/10.1177/0305735689172001

[8] S. Reid, "Preparing for performance," in *Musical Performance*, 1st ed., J. Rink, Ed. Cambridge University Press, Dec. 2002, pp. 102–112. [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9780511811739A015/type/book_part

[9] P. Miksza, "A Review of Research on Practicing: Summary and Synthesis of the Extant Research with Implications for a New Theoretical Orientation," *Bulletin of the Council for Research in Music Education*, no. 190, pp. 51–92, Oct. 2011. [Online]. Available: https://scholarlypublishingcollective.org/bcrme/article/doi/10.5406/bulcouresmusedu.190.0051/255279/A-Review-of-Research-on-Practicing-Summary-and

[10] E. R. How, L. Tan, and P. Miksza, "A PRISMA review of research on music practice," *Musicae Scientiae*, vol. 26, no. 3, pp. 455–697, Sep. 2022.

[11] K. A. Ericsson, R. T. Krampe, and C. Tesch-Romer, "The Role of Deliberate Practice in the Acquisition of Expert Performance," *Psychological Review*, vol. 100, no. 3, pp. 364–403, 1993.

[12] S. Hallam, "Professional Musicians' Approaches to the Learning and Interpretation of Music," *Psychology of Music*, vol. 23, no. 2, pp. 111–128, Oct. 1995. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0305735695232001

[13] S. Hallam, I. Papageorgi, M. Varvarigou, and A. Creech, "Relationships between practice, motivation, and examination outcomes," *Psychology of Music*, vol. 49, no. 1, pp. 3–20, Jan. 2021. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0305735618816168

[14] J. Sokolovskis, D. Herremans, and E. Chew, "A novel interface for the graphical analysis of music practice behaviors," *Frontiers in Psychology*, vol. Volume 9 - 2018, 2018. [Online]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.02292

[15] R. Chaffin and G. Imreh, ""Pulling Teeth and Torture" : Musical Memory and Problem Solving," *Thinking & Reasoning*, vol. 3, no. 4, pp. 315–336, Nov. 1997. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/135467897394310

[16] ——, "A Comparison of Practice and Self-Report as Sources of Information About the Goals of Expert Practice," *Psychology of Music*, vol. 29, no. 1, pp. 39–69, Apr. 2001. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0305735601291004

[17] R. Chaffin and T. Lisboa, "Practicing perfection: How concert soloists prepare for performance," *Advances in Cognitive Psychology*, vol. 2, no. 2, pp. 113–130, Jan. 2006. [Online]. Available: http://www.ac-psych.org/en/download-pdf/volume/2/issue/2/id/13

[18] M. Wiseheart, A. A. D'Souza, and J. Chae, "Lack of spacing effects during piano learning," *PLOS ONE*, vol. 12, no. 8, pp. 1–13, 08 2017. [Online]. Available: https://doi.org/10.1371/journal.pone.0182986

[19] C. E. Carter and J. A. Grahn, "Optimizing music learning: Exploring how blocked and interleaved practice schedules affect advanced performance," *Frontiers in psychology*, vol. 7, p. 1251, 2016.

[20] M. A. Cara, "The effect of practice and musical structure on pianists' eye-hand span and visual monitoring," *Journal of eye movement research*, vol. 16, no. 2, pp. 10–16 910, 2023.

[21] R. M. Winters, S. Gururani, and A. Lerch, "Automatic practice logging: Introduction, dataset & preliminary study." in *Proceedings of the 17th International Society for Music Information Retrieval Conference*. ISMIR, Aug. 2016, pp. 598–604. [Online]. Available: https://doi.org/10.5281/zenodo.1416224

[22] A. Lerch, C. Arthur, A. Pati, and S. Gururani, "An Interdisciplinary Review of Music Performance Analysis," *Transactions of the International Society for Music Information Retrieval*, vol. 3, no. 1, pp. 221–245, Nov. 2020. [Online]. Available: http://transactions.ismir.net/articles/10.5334/tismir.53/

[23] S. Gururani, K. A. Pati, C.-W. Wu, and A. Lerch, "Analysis of Objective Descriptors for Music Performance Assessment," in *Proceedings of the International Conference on Music Perception and Cognition (ICMPC15/ESCOM10)*, Graz, Austria, 2018.

[24] C. E. Cancino-Chacón and I. Pilkov, "The Rach3 Dataset: Towards Data-Driven Analysis of Piano Performance Rehearsal," in *MultiMedia Modeling*, S. Rudinac, A. Hanjalic, C. Liem, M. Worring, B. Jónsson, B. Liu, and Y. Yamakata, Eds. Cham: Springer Nature Switzerland, 2024, pp. 28–41. [Online]. Available: https://doi.org/10.1007/978-3-031-56435-2_3

[25] A. Arzt, S. Böck, and G. Widmer, "Fast identification of piece and score position via symbolic fingerprinting," in *Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012, Mosteiro S.Bento Da Vitória, Porto, Portugal, October 8-12, 2012*, F. Gouyon, P. Herrera, L. G. Martins, and M. Müller, Eds. FEUP Edições, 2012, pp. 433–438. [Online]. Available: http://ismir2012.ismir.net/event/papers/433-ismir-2012.pdf

[26] S. D. Peter, "Online Symbolic Music Alignment With Offline Reinforcement Learning," in *Proceedings of the 24th International Society for Music Information Retrieval Conference, ISMIR 2023, Milan, Italy, November 5-9, 2023*, A. Sarti, F. Antonacci, M. Sandler, P. Bestagini, S. Dixon, B. Liang, G. Richard, and J. Pauwels, Eds., 2023, pp. 634–641. [Online]. Available: https://doi.org/10.5281/zenodo.10265367

[27] E. Nakamura, K. Yoshii, and H. Katayose, "Performance Error Detection and Post-Processing for Fast and Accurate Symbolic Music Alignment," in *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)*, Suzhou, China, 2017.

[28] S. Peter and G. Widmer, "The GlueNote: Learned Representations for Robust and Flexible Note Alignment," in *Proceedings of the 25th International Society for Music Information Retrieval Conference, ISMIR 2024, San Francisco, California, USA and Online, November 10-14, 2024*, B. Kaneshiro, G. J. Mysore, O. Nieto, C. Donahue, C. A. Huang, J. H. Lee, B. McFee, and M. C. McCallum, Eds., 2024, pp. 603–610. [Online]. Available: https://doi.org/10.5281/zenodo.14877409

[29] D. Meredith, K. Lemström, and G. Wiggins, "Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music," *Journal of New Music Research*, vol. 31, 04 2003.

[30] T. Collins, "Discovery of repeated themes and sections," *Retrieved 4th May, http://www. musicir. org/mirex/wiki/2013: Discovery of Repeated Themes & Sections*, 2013.

[31] A. Morsi, K. Tatsumi, A. Maezawa, T. Fujishima, and X. Serra, "Sounds out of Pläce? Score-Independent Detection of Conspicuous Mistakes in Piano Performances," in *Proceeding of the 24th International Society on Music Information Retrieval (ISMIR), November 5-9*, 2023.

[32] A. Morsi, H. Zhang, A. Maezawa, S. Dixon, and X. Serra, "Simulating piano performance mistakes for music learning," *Proceedings of the 21st Sound and Music Computing Conference (SMC 2024)*, July 2024.

[33] M. Müller, V. Konz, W. Bogler, and V. Arifi-Müller, "Saarland music data (SMD)," 2011.

[34] 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 21th International Society for Music Information Retrieval Conference, ISMIR 2020, Montreal, Canada, October 11-16, 2020*, 2020, pp. 534–541. [Online]. Available: http://archives.ismir.net/ismir2020/paper/000127.pdf