Basso Continuo Goes Digital: Collecting and Aligning a Symbolic Dataset of Continuo Performance

Adam Štefunko*
Charles University
Institute of Formal and Applied Linguistics
Prague, Czech Republic
stefunko@ufal.mff.cuni.cz

Suhit Chiruthapudi
Institute of Computational Perception
Johannes Kepler University
Linz, Austria
suhit.chiruthapudi@jku.at

Jan Hajič jr.
Charles University
Institute of Formal and Applied Linguistics
Prague, Czech Republic
hajicj@ufal.mff.cuni.cz

Carlos Eduardo Cancino-Chacón
Institute of Computational Perception
Johannes Kepler University
Linz, Austria
carlos_eduardo.cancino_chacon@jku.at

Abstract

Basso continuo is the baroque accompaniment practice of improvising harmony and upper voices upon a notated bass line. The performance – "realization" – of continuo is not just historical heritage, but a living cornerstone of the Historically Informed Performance movement. However, in music information retrieval it has received little attention and its "living" side has been overlooked entirely. We present a pilot dataset consisting of 6 hours of basso continuo performances in 175 MIDI recordings, which is the first of its kind. To connect contemporary practice to musicological knowledge, and to enable comparing the performances themselves, one must align performances to the notated bass lines. We analyze the challenges that continuo alignment presents, and evaluate baseline two-step alignment using state-of-the-art variants of hidden Markov models and dynamic time warping. Whereas the bass line is aligned well, assigning individual notes of the realization to the score will require further attention.

1 Introduction

Basso continuo is a baroque accompaniment practice in which a player performs a notated bass line and improvises upper voices upon it, with the bass line defining the harmonic constraints within which the accompaniment must fit. Often, these constraints are disambiguated with numeric and other figures, leading to the associated term "figured bass". This process, as well as the resulting musical part, is called continuo *realization* (Williams and Ledbetter, 2001). The basic realization consists of three voices above the bass line, but the artist is free to choose the textures of the accompaniment, as long as they adhere to harmonic constraints implied by the bass line. An excerpt is shown in Figure 1. Historical styles of basso continuo performance have been studied extensively by musicologists and performers themselves, but basso continuo is a living practice as much as it is a heritage of the past Mortensen (1996): "resurrecting" the skill of continuo has been one essential part of the Historically Informed Performance (HIP) movement (Christensen, 2002).

Continuo performers also study historical ways of *basso continuo* playing and develop their own personal understanding of the historical styles, which they apply on stage. However, their practices are mostly left unresearched: we do not really know what continuo players today are playing. We believe this is mostly due to the fact that the improvisatory nature, and thus the lack of musical

*Work done while author was jointly employed as an intern at the Institute of Computational Perception, Johannes Kepler University, Linz, Austria.

Proceedings of the 6th Conference on AI Music Creativity (AIMC 2025),
Brussels, Belgium, September 10th-12th

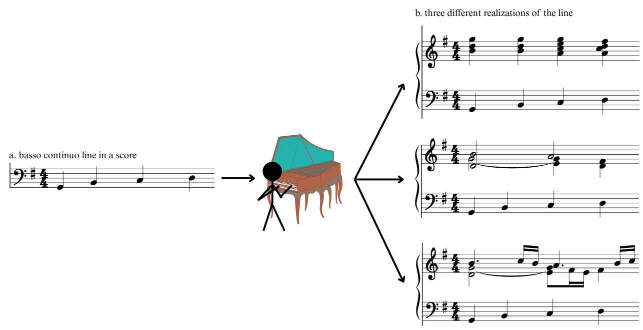

Figure 1: Different realizations of the same basso continuo line, visualised as music notation. Note that while music notation is the default way of communicating the tonal content of realizations visually, it leaves out many important properties of performance — microtiming, arpeggiations, articulation, etc.

notation. The only empirical observations one can possibly have is from performance: there is no such thing as a fully notated work of continuo.[2] This makes continuo performance less accessible to usual methods of musicological study. But, not including the current artists in continuo research misses the majority of what continuo is today: a practice.

We propose that the digital domain is inherently well-suited to the study of continuo. Because the bulk of the performer's task is to improvise tonal content, continuo practice is observable in the symbolic domain with MIDI technology, including nuance beyond what written scores can capture. Computational — probabilistic — models are then an appropriate approach to describe and abstract these practices (Paiement et al., 2009; Spiliopoulou and Storkey, 2011), corresponding to their ephemeral and variable nature. However, while methods exist that could process continuo performance and enable its study, no such recordings in fact exist.

We collect the first (pilot) dataset of MIDI recordings of continuo realizations, performed by 7 professional harpsichord[3] players and students. These 6 hours of recordings present the first empirical, first-hand evidence of how basso continuo is performed at the symbolic music level – that is, with an exact record of the tonal material.

However, to answer the question "How do we play continuo today?" and to make the dataset practical and actionable for further research, the recorded realizations must first be aligned to their scores — the notated bass lines. Continuo practice is defined in terms of the bass lines: baroque treatises write instructions of the type *"If the bass rises by a step, these are the valid options for accompaniment: ..."*,[4] and continuo players internalize these patterns and execute them for their specific bass lines. The scores are thus the fundamental link between contemporary continuo practice and its historical layer. Performance-to-score alignment is a long-standing task for MIR (Orio et al., 2003; Cancino-Chacón et al., 2018; Peter et al., 2023), but so far this task has primarily focused on later Western classical music from the common practice period, featuring fully notated scores that contain the exact notes used in performance, possibly with a few added ornaments and/or outright mistakes (Nakamura et al., 2017; Peter and Widmer, 2024). Continuo performance by definition includes more than is in the score: most of the notes played are added by the performer, sometimes in a simpler and sometimes in a more elaborate manner (see Figure 1). Continuo therefore presents a specific alignment challenge, between traditional performance-to-score alignment and estimating chord labels.

There are three main contributions of the paper:

[2]With the few exceptions in historical sources, such as Bach's realization for one Albinoni's violin sonata, or Tonelli's realizations for Corelli's violin sonatas, well known and studied in the continuo performance community. However, while these communicate the tonal content of the realizations, still they do not capture performance fully: how were Tonelli's rich chords arpeggiated?

[3]While we acknowledge the importance of non-keyboard continuo harmonic instruments such as lutes, bass viols, and more, collecting MIDI data for other than keyboard continuo instruments is impractical, as it would require highly specialized equipment.

[4]This mostly concerns the Italian and South German traditions: for example Gasparini's *L'Armonico Pratico al Cimbalo* or the *Fundamenta partiturae in compendio dato* of Matthias Gugl.

- We collect the first symbolic dataset of continuo realizations (Section 3).
- We define the task(s) of continuo alignment and provide manual ground truth for a part of the data (Section 4). The data and alignments are made available as the Aligned Continuo Realization Dataset (ACoRD)[5].
- We adapt existing performance-to-score alignment methods based on hidden Markov models (HMMs) and dynamic time warping (DTW) for continuo alignment and provide first strong baselines for the task (Sections 5 & 6). The code has been published as a repository[6].

2 Related Work

2.1 Basso Continuo

Basso continuo musicological research has for several decades focused on analyzing historical textual sources, mostly treatises by 17th and 18th century composers and theoreticians, and a rather small amount of existing notated historical basso continuo realizations. The work of Christensen (2002) provides a general overview of the basics of historical basso continuo playing. Giulia Nuti (2017) analyzes the specifics of the Italian basso continuo style in great detail. Mortensen (1996) argues that basso continuo playing is even nowadays a living performance practice. He writes that the contemporary playing differs from its historic counterpart, but does not provide a more detailed analysis of how basso continuo playing is practiced in our times.

Computational research related to basso continuo has focused primarily on the generation of basso continuo realization or numerical annotations from musical scores using mostly algorithmic procedures (Niitsuma and Saito, 2007; Niitsuma et al., 2011), decision trees (Wead and Knopke, 2007) and marginally also machine learning (Ju et al., 2020).This research area aligns with the broader field of AI music generation, where systems based on convolutional neural networks (Huang et al., 2019) and transformers (Thickstun et al., 2023) have demonstrated state-of-the-art performance. However, none of the existing systems specifically address the challenge of human-created performances of basso continuo improvisations. These systems cover only a predefined small portion of the possible improvisations. Processing an arbitrary keyboard player's performance, including possible mistakes, and providing useful feedback to the artist requires a model that covers a much greater proportion of the space of possible accompaniments.

No dataset of continuo realizations is publicly available. The harmonic language of *basso continuo* is well-defined and rooted in the Western classical music tradition (Christensen, 2002), and datasets exist that do exemplify the style of baroque composition and thus certain ideals of continuo realizations, such as Bach chorales (Conklin, 1966) or Giant-MIDI dataset (Kong et al., 2020), but these fail to capture the inherent variability of an improvised accompaniment.

2.2 Performance-to-Score Alignment

Performance-to-score alignment refers to aligning a performance with its corresponding musical score. Introduced by Dannenberg (1984) and Vercoe (1984), it has since become a foundational task in MIR. Music alignment is essential for quantitative performance analysis (Cancino-Chacón et al., 2018) and is the basis of many applications, including automatic accompaniment (Cancino-Chacón et al., 2023; Raphael and Gu, 2009), automatic page turning (Arzt et al., 2008; Henkel et al., 2021) and multimodal visualizations (Otsuka et al., 2011; Maezawa, 2024; Lartillot et al., 2020).

In the symbolic domain, performance-to-score alignment means matching notes in a symbolic performance (e.g., MIDI) to those in a symbolic score (e.g., MusicXML/MEI). Note-level symbolic alignment algorithms tend to be very accurate if the score and performance correspond well (Peter et al., 2023; Peter, 2023; Nakamura et al., 2017), as is the case in Western music of the common practice period. In note-wise symbolic alignment, a performed note is typically *matched* to a single score note, or marked as an *insertion*, or as part of an *ornament* (e.g., a trill) (Peter et al., 2023; Foscarin et al., 2022). A note in the score that is not performed is marked as a *deletion*.

Music alignment algorithms generally fall into two categories: probabilistic models and dynamic programming. Probabilistic state-space models treat alignment as latent state inference, where

[5] http://hdl.handle.net/11234/1-5963
[6] https://adamcho14.github.io/acord.github.io

Table 1: Information on scores with *basso continuo* in the ACoRD dataset

Piece ID	Name	Composer	Key	Instrumentation	# of Basso Continuo Notes
001	Pour la Bergere Lisète	J.-F. Dandrieu	F Major	voice & basso continuo	86
002	Partimento Book 2 No. 5	F. Fenaroli	B Major	basso continuo	116
003	Neues	G. Ph. Telemann	G Major	voice & basso continuo	82
004	Adagio from Sonata in D Minor	B. Marcello	D Minor	recorder & basso continuo	107
005	Partimento Book 2 No. 6	F. Fenaroli	B Minor	basso continuo	135

score positions are hidden variables (Cano et al., 1999; Cont, 2008; Duan and Pardo, 2011). The most common approach for symbolic alignment involves variants of HMMs, where the hidden states represent time positions in the score and the observations are the notes and onset times in the performance (Nakamura et al., 2015; Raphael and Gu, 2009; Cancino-Chacón et al., 2023). Dynamic programming methods, especially DTW, align sequences by minimizing cumulative cost and, while traditionally used for audio-based alignment, have recently been adapted for symbolic alignment (Peter et al., 2023; Peter, 2023).

3 ACoRD: a Pilot Continuo Realization Dataset

The Aligned Continuo Realization Dataset (ACoRD) consists of 175 pilot MIDI recordings of continuo realizations: 3 from musical pieces for a solo voice or instrument and basso continuo from the 18th century, and two *partimenti* (Williams and Cafiero, 2001) of Fedele Fenaroli, which are pedagogical tools for students of continuo, where they improvise realizations with no solo voices to accompany. All pieces were performed by 7 participants at different levels of proficiency: 4 professionals and 3 intermediate to advanced students of continuo.[7] Each participant recorded each piece 5 times, to capture within-artist variability in continuo realization. Two participants recorded their performances on Roland A-37 and the rest on M-Audio Keystation 61 MK3. The sound came from a sampled Christian Zell 1737 harpsichord running in the GrandOrgue open-source virtual simulator.[8] The recordings were made in Ardour,[9] an open-source digital audio workstation.

The overall playing time of the dataset is just above 6 hours. A total of **66,967 notes** in the performances is recorded, of which 28 % are notes from the score. Table 1 shows basic information about the recorded musical pieces.

The recordings in the dataset are not meant to represent textbook continuo realizations, but real-world performances that contain accidental flaws and intentional variations that arise as keyboard players realize continuo. Observed discrepancies between an ideally correct performance and the real performance include:

- Overly repeated or, on the other hand, skipped bass notes;
- Ornaments such as mordent or trill, or even added notes to the bass line;
- Sudden interruption and continuation at the same place after a brief break;
- Substitution errors in the bass part;
- Intentional transposition of the whole piece by one performer.

Further challenges are associated with inherent properties of continuo performance, such as overlegato articulation on the harpsichord, arpeggiation, or deliberate rhythmic instabilities. Some of these phenomena are illustrated in Figure 2.

4 Aligning Continuo: Definitions and Ground Truth

We have established in the Introduction that in order to interpret and further process the recorded continuo, we must align the realization to the score. The description of the data in Section 3 and Figure 2 indicated that this task is not trivial. What, however, does it mean to align continuo?

[7] One studies continuo typically after already having substantial expertise on the piano, and after some familiarisation with harpsichord. "Students" in this context are aged 18–25, on the threshold of performing professionally; and are not to be considered beginners.

[8] https://github.com/GrandOrgue/grandorgue

[9] https://ardour.org

Each note in the score implies harmonic constraints valid for its duration. The choice of texture, including ornamental notes that are allowed outside of the harmonic constraint, is left to the performer. The fundamental relationship between the score and the performance is that each note in the realization relates to a note in the score, and the task of **full realization alignment** is to find this assignment.

As opposed to aligning fully notated scores and performances, multiple performed notes are expected to align to each score note,without necessarily matching exactly any of their individual attributes – pitch or the implied harmonic constraint, or onset and offset. The characteristic features of basso continuo performance therefore implies a paradigm change compared to the full-score alignment task.

Fortunately, because the notated bass line itself must by definition also be part of the continuo realization, we can define a sub-task: aligning just the subset of performance notes that corresponds to the score notes. This **bass line alignment** can be defined as a one-to-one matching of notes with exactly matching attributes. For this symbolic-domain methods are well-known, and having such an alignment could then reduce the overarching alignment task to just matching all notes with onsets in between two adjacent aligned bass notes.

4.1 Manual Ground Truth Alignment Data

Because both alignment steps (bass line alignment and full alignment) are far from trivial, one must measure how well methods of performance-to-score alignment perform. Thus, manual ground truth is necessary.

In order to obtain ground truth data that can be used for evaluating both alignment steps, we have manually annotated 35 bass line alignments (one randomly chosen take of each piece of each player, with 3,593 matched score notes in total), and 15 full realization alignments (3 takes of each of the 5 pieces, with the 7 players represented as equally as possible; 6,741 matches in total). The manual alignment of full realizations took roughly one hour per file. All manual annotations were performed using the online tool *Parangonada* (Peter et al., 2023) and were supervised by one of the authors, a professional harpsichord player.

The bass alignment annotations were created considering a one-to-one mapping between the score notes and the performance bass notes, i.e., every score note is aligned to a maximum of one performance bass note. The realization alignment annotations were created considering a many-to-many mapping between the score notes and the performance notes. This will be discussed further in Section 4.2. For these annotations, we have chosen to manually correct some of the files that were already automatically aligned using the *DualDTWNoteMatcher* (Peter, 2023), an existing state-of-the-art offline algorithm, which will be described in Section 5.

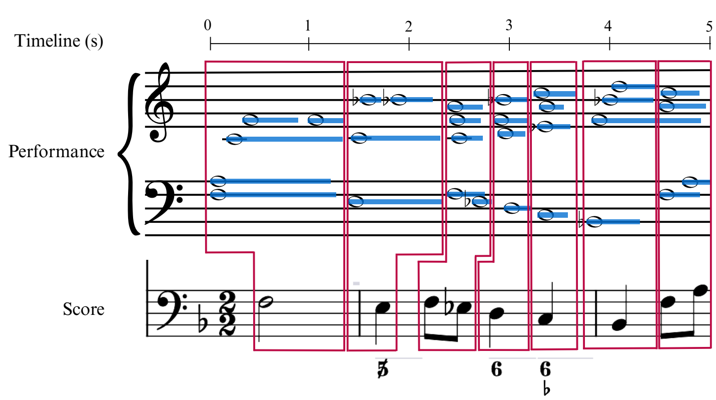

Figure 2: Illustrating the relationship between a continuo realization (top) and the notated bass line (bottom). Color codes corresponding segments of the continuo part.

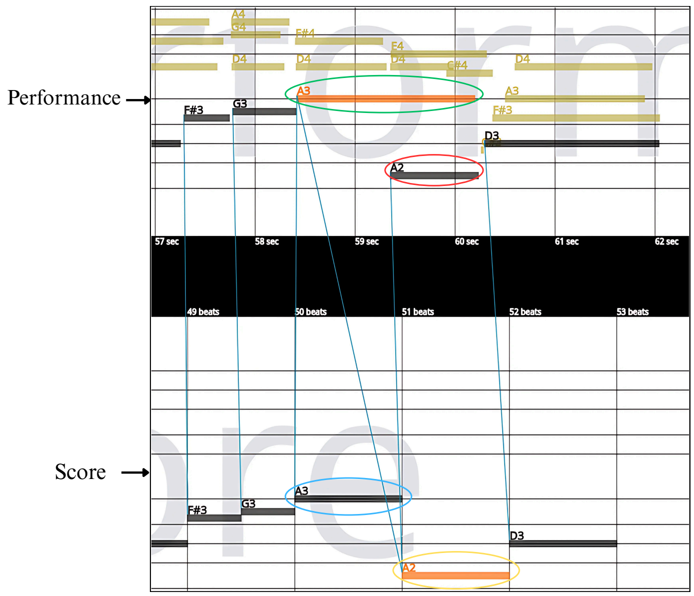

Figure 3: Illustration of a case study where a performance bass note (circled in green) corresponding to a score note (circled in blue) transforms into a realization note for the following performance bass note (circled in red) corresponding to the subsequent score note (circled in yellow).

4.2 Challenges of Continuo Alignment

The manual alignment annotation process was divided into two steps: performance bass note alignment annotation, and performance realization notes alignment annotation. A one-to-one mapping was chosen for the process of manually annotating the alignment between the basso continuo score and the performance's bass line. This decision follows the reasoning that the score notes follow a sequence that must be respected in the performance (Christensen, 2002), and therefore, there can only be one corresponding performance bass note for each score note. Any other notes that occur in the performance apart from the performance bass notes are thus considered the realizations of the bass note.

A many-to-many mapping was chosen for the process of annotating the alignment between the score notes and the performance realization notes (including the performance bass notes). This was done for the following reasons:

- To allow for the performance bass note as well as the other realization notes corresponding to that bass note to be aligned to one score note (many-to-one).
- To allow for cases where a realization note corresponding to a performance bass note continues to stretch in duration beyond the duration of the performance bass note and across the duration of the following performance bass note(s). In this case such a realization note is aligned to all the score notes corresponding to the performance bass notes that it has stretched over (one-to-many).
- To allow for situations where a performance bass note itself stretches in duration over the length of the next performance bass note(s), and therefore transforms into a realization of the following bass note(s). As can be seen in Figure 3, in such cases, the performance bass note (circled in green) is aligned with its own score note (circled in blue), and it is also aligned with the next score note (circled in yellow); i.e. the score note corresponding to the next performance bass note (circled in red).

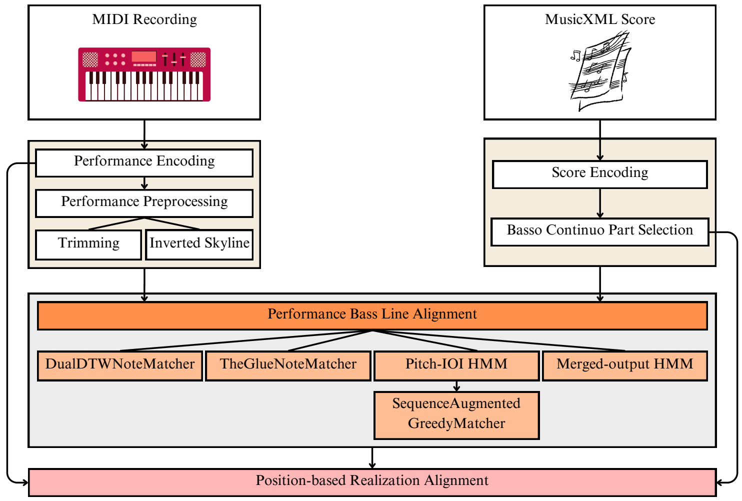

Figure 4: The modular alignment system pipeline. The system accepts scores in MusicXML format and performances as MIDI file.

5 Baseline Alignment System

The baseline alignment system over ACoRD, shown on Figure 4, enables operations with several preprocessing and alignment methods on the data. The system accepts MIDI files for performances, and MusicXML files for scores. These files are parsed using the `partitura` library (Cancino-Chacón et al., 2022). In case of multi-part scores (see column "Instrumentation" in Table 1), only the bass continuo line is extracted. Each performance and score pair forms an input to the alignment system.

This system operates in two steps. The first step is *bass line alignment*. We implement two optional continuo-specific preprocessing methods, and four state-of-the-art alignment systems are available.

These bass alignments are used in the second step — *full realization alignment*. For this step only a simple position-based alignment method is currently available, because it involves the many-to-many alignment paradigm.

5.1 Preprocessing

Score preprocessing consists only of basso continuo part selection from scores with multiple parts in them. The scores in our dataset are made so that this part is always the bottom one in a system, as is usually the case in standard scores of baroque music (with the accompanied solo lines in the upper parts, except for partimenti).

Performance preprocessing for bass line alignment aims at reducing the disproportionate number of insertions in the performance by guessing which notes can never be validly aligned with the score. We implement two methods which essentially preclassify anything that is "too high" as an insertion with respect to the bass line:

1. *Trimming*, which cuts out all performance notes whose pitch is higher than the highest note in the basso continuo score.
2. An *inverted skyline algorithm*, which retains only notes which are the lowest performance note at at least some point of their duration. This preprocessing method is based on the skyline algorithm for symbolic melody identification (Uitdenbogerd and Zobel, 1999; Simonetta et al., 2019).

We also work with raw encoded performances, to check what the influence of preprocessing is across different alignment methods. In any case no performance notes can be discarded for the full realization step.

5.2 Bass Line Alignment

We use four state-of-the-art alignment methods used for performance-to-score alignment, based mostly on HMMs and DTW. All the alignment methods used are publicly available as either as Python packages or C++ source code. We use the following systems:

- *DualDTWNoteMatcher* from the `parangonar` library (Peter, 2023)
- *TheGlueNoteMatcher* from the `parangonar` library (Peter and Widmer, 2024)
- *Pitch-IOI HMM* from the `matchmaker` library (Park et al., 2024)
- *Merged-output HMM* of Nakamura et al. (2017).

These systems have been used to align performances of full (piano) scores and, therefore, have strong constraints on pitch differences between the performance and the score. Therefore, they are not suited for the full realization alignment, but they are useful to align the realization bass line with the score (taken into account that the bass line should be present in the realization as it is written in the score at almost all times).

5.2.1 DualDTWNoteMatcher

A state-of-the-art offline alignment algorithm based on DTW (described in Peter (2023)) computes a forward and a backward DTW warping path based only on pitch information of performance and score notes. The system uses a pitch-based heuristics in cases where these two paths diverge. Afterwards, a second DTW is used to align the onsets of individual notes based on the mapping from the earlier step.

5.2.2 TheGlueNoteMatcher

TheGlueNote, introduced by Peter and Widmer (2024), is another DTW-based system. This system uses a transformer encoder network, which computes positional embeddings of two note sequences. The similarity matrix of these two sequences is then used for a weighted DTW-based warping path extraction.

5.2.3 Pitch-IOI HMM

The PitchIOI HMM is a probabilistic symbolic alignment method used for real time piano accompaniment (Cancino-Chacón et al., 2023). This method uses a switching Kalman filter, a hybrid model combining an HMM and a Kalman filter, with parameters conditioned on HMM states (Murphy, 1998). The observed variables are the performed MIDI pitch and the performed inter-onset intervals (IOIs, i.e., the time interval between consecutive onsets). The hidden variables are the score onset times, and the performance tempo (modeled by the Kalman filter part). The score position inference is done using the forward algorithm (Rabiner and Juang, 1986; Murphy, 1998). The output of this method only returns temporal alignment between performance time and score time, but does not provide a note-level alignment. In order to obtain note-level alignment, a greedy pitch-wise matching algorithm is used, which matches the performed MIDI pitch to the closest score note in the score position returned by the HMM.

5.2.4 Merged-output HMM

Nakamura et al. (2017) proposed an HMM-based alignment system that combines outputs from multiple HMM alignments. The system is using a prealigment by so-called temporal HMM. The information from prealignment is used to detect performance errors by applying Viterbi algorithm on clusters and notes in regions with extra notes. Finally, erroneous regions are realigned using merged-output HMM, which combines multiple musical streams – one for each hand, separated by a hand separation algorithm.

5.3 Position-based Full Realization Alignment

Assuming that a performance has a bass line already aligned with the score, we base the second alignment step on the temporal relationships between unaligned (realization) and already aligned (bass) notes in the performance.

First, we align performance realization notes with the performance bass notes (the performance notes aligned to the bass line in the previous step) according to several conditions. Let on_r and off_r be onset and offset times of a realization note, and on_b onset time of a bass note and off_b either offset time of the bass note or onset time of the following bass note, if the following bass note starts before the given bass note ends. A realization note gets aligned with a bass note if at least one of these four conditions, defining whether a realization note shares a play time with a bass note, applies:

(1) $\text{on}_b \leq \text{on}_r \leq \text{off}_r \leq \text{off}_b$

(2) $\text{on}_r < \text{on}_b < \text{off}_r \leq \text{off}_b$ and $\text{off}_r - \text{on}_b \geq \gamma(\text{off}_r - \text{on}_r)$

(3) $\text{on}_b \leq \text{on}_r < \text{off}_b < \text{off}_r$ and $\text{off}_b - \text{on}_r \geq \gamma(\text{off}_r - \text{on}_r)$

(4) $\text{on}_r < \text{on}_b < \text{off}_b < \text{off}_r$

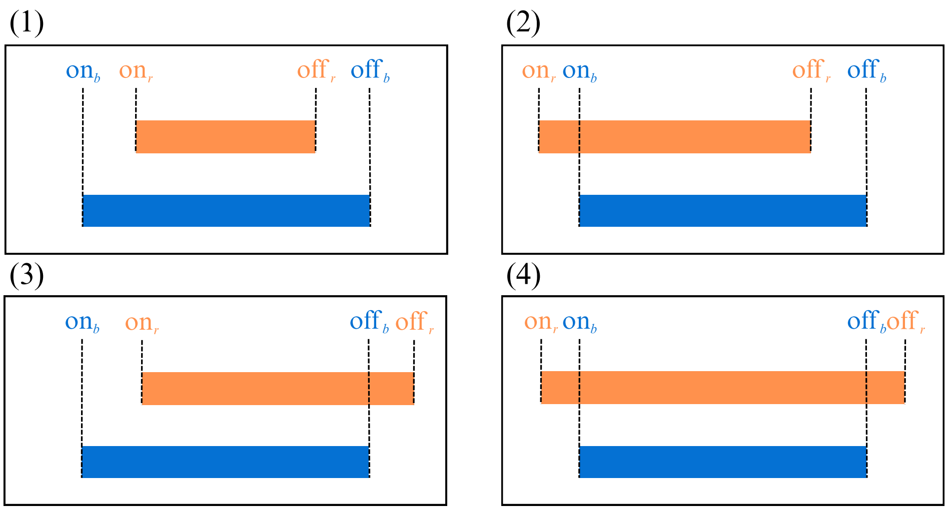

Figure 5: A performance realization note (orange) is aligned with an already aligned performance bass note (blue) if both notes are played simultaneously, as defined by conditions (1) – (4).

These inequalities, visualized in Figure 5, describe cases in which the performance realization note and bass note share a common duration time. This happens if one of the notes takes up the whole duration of the other, or if a large enough portion of the realization note's duration is played together with the bass note. The parameter $\gamma = 0.25$ governs how large the portion of the duration of the realization note has to be played simultaneously with the bass note. This condition is most likely to be triggered by a tied note that extends over multiple bass notes, which is prototypically done when preparing a dissonance, and as dissonance is more significant than its resolution, we would expect it to last for at least half the value of the following bass note. Setting $\gamma = 0.25$ thus leaves a margin for e.g., imperfect articulation or timing in performance. Other values of this parameter may prove to lead to better performance, but a system that is too sensitive to γ is likely overfitting a test set and not robust for real-world settings.

This does not necessarily align all the realization notes with the bass notes, Because this is required by the task of realization, alignment as described in Section 4, a heuristic is applied to all realization notes that have not yet been aligned: each unaligned note is connected to the closest bass note.

Finally, all realization notes are matched with the score notes corresponding to their performance bass notes through the bass-line-to-score alignment. Note that this cannot be done directly because we need to compare onset times like-to-like: performance onset times are in wallclock time (expressed as MIDI ticks), while score onset times are in beats.

6 Alignment Experiments

Our experiments were conducted in two steps. First, we tested how well different alignment methods and preprocessing techniques could align the bass line corresponding to the basso continuo part in the score. In the second step, we evaluated the alignment of the full basso continuo realization. All measurements were performed on the ground truth data described in Section 4.

6.1 Evaluation Metrics

In note-level symbolic alignment (Foscarin et al., 2022; Peter et al., 2023; Nakamura et al., 2017), a pair of a performance note and a score note is considered a *match* if the two notes are aligned (by an alignment algorithm or an annotator). A score note that has no counterpart in the performance is marked as a *deletion* and a performance note not found in the score is labeled as an *insertion*. An alignment can be then represented as a list of dictionaries where each dictionary has a label (match, insertion or deletion) and an index of a performance and/or score note.

Symbolic alignment methods for Western classical music of the common practice period are typically evaluated by computing precision, recall, and F1-score aggregated over all labels, or by focusing only on matches (Peter et al., 2023; Peter, 2023; Peter and Widmer, 2024). However, in basso continuo, the distribution of matches, insertions, and deletions differs from that observed in piano music of the common practice period. Figure 6 shows the distribution of matches, insertions, and deletions in the manually annotated bass line alignments from the ACoRD dataset, as well as from three widely used symbolic datasets of piano performances aligned to their scores: ASAP (Peter et al., 2023), Batik (Hu and Widmer, 2023), and Vienna 4x22 (Goebl, 2016). As shown in the figure, insertions account for nearly three-quarters of the data in the basso continuo alignments, compared to less than 7% in the other datasets. Therefore, instead of reporting aggregated metrics across all labels, we report metrics separately for each label when evaluating bass line alignment. Note that for the realization alignment task, there are no insertions, as all notes in the realization are aligned to the corresponding bass note(s).

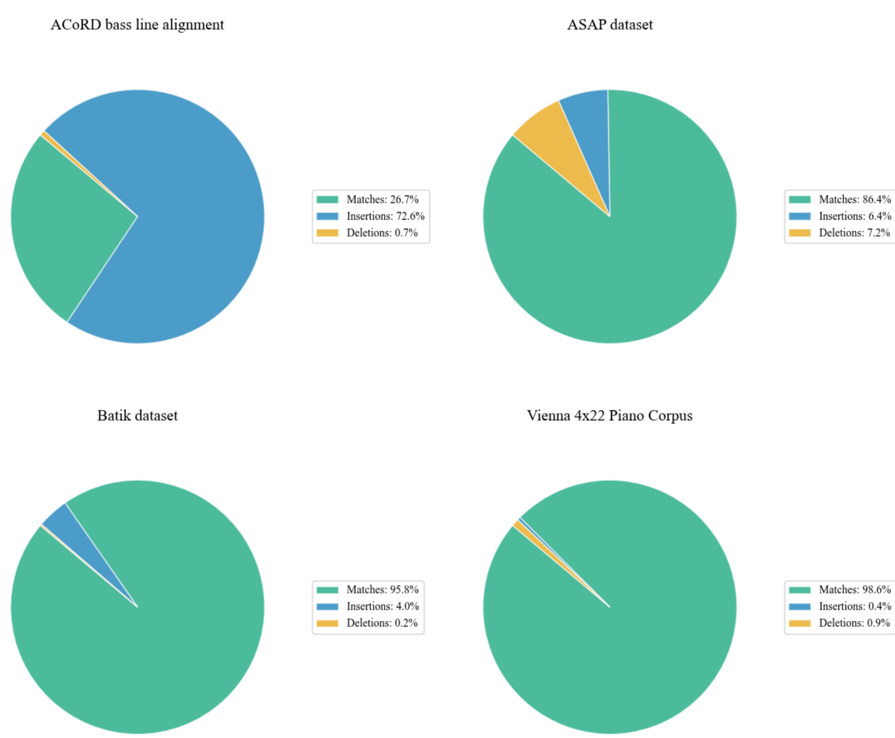

Figure 6: Comparison of label distribution across three different datasets: a. Basso Continuo Realization Dataset, b. ASAP dataset, c. Batik-plays-Mozart corpus, d. Vienna 4x22 Piano Corpus.

6.2 Bass Line Alignment

For the bass alignment task, we tested the performance of the four alignment algorithms described in Section 5.2. We also observed how three different performance preprocessing techniques affect the performance of these alignment methods: no preprocessing, trimming, and inverted skyline (see Section 5.1). Therefore, we have run 12 experiments with different combinations of a preprocessing technique and an alignment system. Each combination produced alignment files, which have been stored as *Parangonada* CSV files (Peter et al., 2023). We then compared these alignments with manually aligned ground truth data and evaluated the alignments according to the methodology described in Section 6.1.

6.3 Realization Alignment

For the realization alignment task, we use the output of the bass alignment with the position-based full realization alignment described in Section 5.3. We compared how the 12 different outputs from the bass line alignment (4 alignment methods $\times$ 3 preprocessing methods) affect the overall realization alignment. Additionally, we test how the position-based full realization alignment would work in the best-case "oracle" scenario, having access to the true bass line alignment from the ground truth.

6.4 Automatic Alignment Results

Results for the bass line alignment experiments are shown in Table 2. To test whether there is a statistically significant difference in the performance of alignment methods (DualDTW, GlueNote, PitchIOI HMM and Merged-output HMM) and preprocessing techniques (no preprocessing, trimming, inverted skyline), we conducted an Aligned Rank Transform (ART) ANOVA (Wobbrock et al., 2011) on the F1-scores for the 35 manually annotated files, for each label (matches, insertions and deletions). ART ANOVA is a non-parametric version of the two-way ANOVA test, well suited for non-normally distributed data, as is the case of the F1-score in our results.

The ART ANOVA tests for bass line alignment show that across all three categories (F1-scores for matches, insertions, and deletions), the interaction between preprocessing technique and alignment method was consistently significant (all $p < 0.001$), indicating that the accuracy of each alignment method depends strongly on the preprocessing applied. For matches, there was also a significant main effect of alignment method ($p = 0.03$) and a marginal effect of preprocessing ($p = 0.07$). Insertions showed no significant main effects for either preprocessing ($p = 0.62$) or method ($p = 0.85$). In contrast, deletions presented a significant main effect of preprocessing ($p = 0.002$), but marginal effect of alignment method ($p = 0.08$). Note however, deletion label data do not provide much value because the number of deletions is small, compared to the other labels (see Figure 6).

The results in Table 2 suggest that, for bass line alignment, the DualDTWNoteMatcher (Peter et al., 2023) is the best performing alignment method. It shows consistently good performance even without preprocessing, with only a slight improvement when trimming is applied, particularly for deletions. In contrast, the other methods benefit substantially from the use of trimming.

The results for full realization alignment are shown in Table 3. In this case, we only report results for matches, since, by construction, there are no insertions, and the deletions would be the same as in the case of bass line alignment. We conducted an ART ANOVA on the 15 fully annotated realization alignments to test whether there is a statistically significant difference in the F1-score of the alignment methods (the 4 methods described above plus the "oracle" bass line alignment) and preprocessing techniques. The results show again a significant effect for the interaction of alignment method and preprocessing ($p < 0.001$), but neither alignment method ($p = 0.38$) nor preprocessing ($p = 0.48$) showed a significant effect.

The results in Table 3 also suggest that the DualDTWNoteMatcher is the best performing automatic alignment method, but it still lags behind from the oracle manual alignment. As shown in Figure 6, given that there are approximately three performed non-bass notes per score note, correctly identifying nine out of ten notes implies an error in every third or fourth bass note on average—or about once per measure. In practice, however, these mistakes are likely to be distributed less evenly. This level of accuracy is not yet sufficient for practical applications. A more detailed analysis of the errors would be necessary to make further progress, but that is beyond the current scope. We simply provide these baseline results as a foundation for future work to build upon.

Table 2: Performance of bass alignment methods across three evaluation labels (*match*, *insertion*, and *deletion*) under different preprocessing techniques

Alignment Method	No Preprocessing			Trimming			Inverted Skyline		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Match									
DualDTWNoteMatcher	0.985	0.974	0.979	0.986	0.980	0.983	0.963	0.677	0.793
TheGlueNoteMatcher	0.668	0.456	0.535	0.918	0.869	0.891	0.921	0.620	0.737
Pitch-IOI HMM	0.639	0.644	0.641	0.628	0.633	0.631	0.385	0.289	0.330
Merged-output HMM	0.441	0.345	0.379	0.875	0.871	0.872	0.752	0.615	0.673
Insertion									
DualDTWNoteMatcher	0.993	0.997	0.995	0.995	0.996	0.995	0.886	0.994	0.937
TheGlueNoteMatcher	0.868	0.982	0.921	0.969	0.986	0.977	0.876	0.993	0.930
Pitch-IOI HMM	0.961	0.957	0.959	0.959	0.955	0.957	0.884	0.978	0.928
Merged-output HMM	0.830	0.937	0.879	0.969	0.972	0.971	0.873	0.944	0.907
Deletion									
DualDTWNoteMatcher	0.576	0.646	0.583	0.604	0.645	0.605	0.064	0.693	0.110
TheGlueNoteMatcher	0.067	0.482	0.102	0.368	0.609	0.401	0.053	0.591	0.093
Pitch-IOI HMM	0.123	0.081	0.094	0.109	0.072	0.083	0.031	0.215	0.049
Merged-output HMM	0.033	0.343	0.050	0.127	0.166	0.109	0.030	0.189	0.047

Table 3: Performance of realization alignment methods under different preprocessing techniques

Alignment Method	No Preprocessing			Trimming			Inverted Skyline		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Match									
Manual bass alignment	0.966	0.925	0.945	–	–	–	–	–	–
DualDTWNoteMatcher	0.947	0.907	0.926	0.947	0.907	0.926	0.746	0.658	0.698
TheGlueNoteMatcher	0.473	0.428	0.449	0.883	0.840	0.861	0.701	0.617	0.656
Pitch-IOI HMM	0.636	0.607	0.621	0.626	0.598	0.612	0.307	0.272	0.288
Merged-output HMM	0.320	0.289	0.303	0.899	0.860	0.879	0.785	0.707	0.743

7 Discussion and Conclusions

What artists in fact play when they perform *basso continuo* today has been a topic overshadowed as a subject of study by focus on the historical foundations and sources of continuo and questions of historically informed performance practice (such as whether to accompany certain repertoire on the harpsichord or the organ). This was also due to lack of continuo realization data that can be analyzed in appropriate representations, since basso continuo is in its essence improvisation on a given notated part and, thus, it is not primarily recorded or notated.

The digital domain is well-suited to broaden the scope of musicological and artistic research to include continuo performances of current artists. To initiate this process, we have collected the ACoRD pilot dataset of 6 hours of continuo realizations, consisting of 175 MIDI recordings of performances of 5 baroque pieces with basso continuo, performed by 7 harpsichordists.

To enable further research over this dataset, we have provided ground truth for continuo alignment, and we have adapted currently existing performance-to-score alignment methods based on dynamic time warping and hidden Markov model for a baseline system capable of basso continuo alignment. This required a change of paradigm from note-to-note alignment, as it is usually required in full-score alignments of Western classical music, to many-to-many notes alignment. We have tested four state-of-the-art alignment systems together with three preprocessing methods. While the *DualDTWNoteMatcher* alignment method performs well for bass line alignment (with an F1-score above 0.99 for insertions and around 0.99 for matches, though performance in the small deletions category remains unsatisfactory), and also achieves over 0.9 F1-score for full realization alignment, a more detailed analysis of specific issues in continuo alignment is needed to bring this fundamental task to a level suitable for artist-facing applications.

Such observations from continuo research can also serve as a precursor to computational research of other semi-improvisatory genres that use partially notated music. Improvising accompaniment upon incomplete notation (Butt, 2002) by creating musical textures in harmony according to a given style is the main task of the majority of keyboard players across multiple genres (Dobbins, 1980; Sarath, 2013), beyond the Historically Informed Performance movement, such as jazz (Dobbins, 1984; Konowitz, 1969) or pop music (Fulara, 2013; Marino, 2021).

Basso continuo is a fascinating intersection of the contemporary and the historical, the written and the improvised, the theory and its embodied practice. It has been barely explored in the digital domain despite the inherent advantages that this domain provides for its empirical study. Further down this line of work are also applied goals: professional harpsichordists, but especially the growing number of continuo learners would benefit from a model of continuo that provides personalized feedback in individual practice. But to this end, more must be done. We hope that making available the ACoRD dataset and the baseline alignment system is a good first step in digital preservation, modeling, and analysis of basso continuo, and we look forward to how others may take this opportunity to direct more attention at continuo in the music computing world.

Ethical Statement

We collected informed consent from the participants to use their recordings and share them under the CC-BY-NC-SA 4.0 license. Data was anonymized.

Acknowledgments and Disclosure of Funding

This work has been partially supported by the Austrian Science Fund (FWF), grant agreement PAT 8820923 ("*Rach3: A Computational Approach to Study Piano Rehearsals*"). Further partial support was given by the project "Human-centred AI for a Sustainable and Adaptive Society" (reg. no.: CZ.02.01.01/00/23_025/0008691), co-funded by the European Union, and by the SVV project number 260 821. The computing infrastructure was provided by the LINDAT/CLARIAH-CZ Research Infrastructure,[10] supported by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LM2023062).

References

Arzt, A., Widmer, G., and Dixon, S. (2008). Automatic Page Turning for Musicians via Real-Time Machine Listening. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*.

Butt, J. (2002). *Playing with history: the historical approach to musical performance*. Cambridge University Press.

Cancino-Chacón, C., Peter, S., Hu, P., Karystinaios, E., Henkel, F., Foscarin, F., Varga, N., and Widmer, G. (2023). The ACCompanion: Combining Reactivity, Robustness, and Musical Expressivity in an Automatic Piano Accompanist. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence IJCAI-23*, Macao, S. A. R.

Cancino-Chacón, C., Peter, S. D., Karystinaios, E., Foscarin, F., and Grachten, M. (2022). Partitura: A Python Package for Symbolic Music Processing. In *Proceedings of the Music Encoding Conference (MEC 2022)*, Halifax, Canada.

Cancino-Chacón, C. E., Grachten, M., Goebl, W., and Widmer, G. (2018). Computational Models of Expressive Music Performance: A Comprehensive and Critical Review. *Frontiers in Digital Humanities*, 5:25.

Cano, P., Loscos, A., and Bonada, J. (1999). Score-Performance Matching using HMMs. In *Proceedings of the International Computer Music Conference (ICMC'99)*, Beijing, China.

Christensen, J. B. (2002). *18th Century Continuo Playing: A Historical Guide to the Basics*. Bärenreiter, Kassel.

Conklin, D. (1966). Bach Chorales. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5GC7P.

[10] https://lindat.cz

Cont, A. (2008). Antescofo: Anticipatory Synchronization and Control of Interactive Parameters in Computer Music. In *Proceedings of the International Computer Music Conference (ICMC '08)*, Belfast, Ireland.

Dannenberg, R. B. (1984). An On-Line Algorithm for Real-Time Accompaniment. In *Proceedings of the International Computer Music Conference (ICMC'84)*, volume 84, pages 193–198, Paris, France.

Dobbins, B. (1980). Improvisation: An essential element of musical proficiency. *Music Educators Journal*, 66(5):36–41.

Dobbins, B. (1984). *The contemporary jazz pianist: a comprehensive approach to keyboard improvisation*, volume 1. C. Colin.

Duan, Z. and Pardo, B. (2011). A state space model for online polyphonic audio-score alignment. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 197–200, Prague, Czech Republic. IEEE.

Foscarin, F., Karystinaios, E., Peter, S. D., Cancino-Chacón, C., Grachten, M., and Widmer, G. (2022). The match file format: Encoding Alignments between Scores and Performances. In *Proceedings of the Music Encodinc Conference (MEC 2022)*. arXiv. arXiv:2206.01104.

Fulara, A. (2013). The model of counterpoint improvisation and the methods of improvisation in popular music. *AVANT. Pismo Awangardy Filozoficzno-Naukowej*, (1):417–454.

Goebl, W. (2016). The Vienna 4x22 Piano Corpus. DOI: 10.21939/4X22 URL: http://dx.doi.org/10.21939/4X22.

Henkel, F., Schwaiger, S., and Widmer, G. (2021). Fully Automatic Page Turning on Real Scores. In *Extended Abstracs for the Late-Breaking Demo Session of the 22nd International Society for Music Information Retrieval Conference (ISMIR 2021)*, Online. arXiv:2111.06643 [cs].

Hu, P. and Widmer, G. (2023). The Batik-plays-Mozart Corpus: Linking Performance to Score to Musicological Annotations. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR 2023)*, Milan, Italy.

Huang, C.-Z. A., Cooijmans, T., Roberts, A., Courville, A., and Eck, D. (2019). Counterpoint by convolution. *arXiv preprint arXiv:1903.07227*.

Ju, Y., Margot, S., McKay, C., Dahn, L., and Fujinaga, I. (2020). Automatic figured bass annotation using the new Bach Chorales Figured Bass Dataset. In *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR 2020)*, Online.

Kong, Q., Li, B., Chen, J., and Wang, Y. (2020). Giantmidi-piano: A large-scale midi dataset for classical piano music. *arXiv preprint arXiv:2010.07061*.

Konowitz, B. L. (1969). *Jazz improvisation at the piano–A textbook for teachers*. Columbia University.

Lartillot, O., Cancino-Chacón, C., and Brazier, C. (2020). Real-Time Visualisation of Fugue Played by a String Quartet. In *Proceedings of the Sound and Music Computing Conference (SMC 2020)*, Online.

Maezawa, A. (2024). I Got Rhythm, so Follow Me More: Modeling Score-Dependent Timing Synchronization in a Piano Duet. In *Proceedings of the Sound and Music Computing Conference (SMC 2024)*, Porto, Portugal.

Marino, S. (2021). Improvisation in pop-rock music. In *The Routledge Handbook of Philosophy and Improvisation in the Arts*, pages 431–445. Routledge.

Mortensen, L. U. (1996). Unerringly tasteful'?: Harpsichord continuo in Corelli's op. 5 sonatas. *Early Music*, 24(4).

Murphy, K. P. (1998). Switching Kalman Filters. Technical Report 98-10, Compaq Cambridge Research Lab.

Nakamura, E., Cuvillier, P., Cont, A., Ono, N., and Sagayama, S. (2015). Autoregressive Hidden Semi-Markov Model of Symbolic Music Performance for Score Following. In *16th International Society for Music Information Retrieval Conference (ISMIR 2015)*, Málaga, Spain.

Nakamura, E., Yoshii, K., and Katayose, H. (2017). 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.

Niitsuma, M., Matsubara, M., Oono, M., and Saito, H. (2011). Development of a method for automatic basso continuo playing. *Information Processing & Management*, 47(3):440–451.

Niitsuma, M. and Saito, H. (2007). Automatic thorough-bass realization in baroque music. In *Proceedings of the International Computer Music Conference, ICMC 2007*, pages 425–428. International Computer Music Conference, ICMC 2007 ; Conference date: 27-08-2007 Through 31-08-2007.

Nuti, G. (2017). *The performance of Italian basso continuo: style in keyboard accompaniment in the seventeenth and eighteenth centuries*. Routledge.

Orio, N., Lemouton, S., and Schwarz, D. (2003). Score following: State of the art and new developments. *New Interfaces for Musical Expression (NIME)*.

Otsuka, T., Nakadai, K., Takahashi, T., Ogata, T., and Okuno, H. (2011). Real-Time Audio-to-Score Alignment Using Particle Filter for Coplayer Music Robots. *EURASIP Journal on Advances in Signal Processing*, 2011(1):384651.

Paiement, J.-F., Bengio, S., and Eck, D. (2009). Probabilistic models for melodic prediction. *Artificial Intelligence*, 173(14):1266–1274.

Park, J., Cancino-Chacón, C., Kwon, T., and Nam, J. (2024). Matchmaker: A Python library for Real-time Music Alignment. In *Proceedings of the Late Breaking/Demo Session at the 25th International Society for Music Information Retrieval Conference*, San Francisco, USA.

Peter, S. D. (2023). Online Symbolic Music Alignment with Offline Reinforcement Learning. In *Proceedings of the 24th International Society for Music Information Retrieval Conference (ISMIR 2023)*, Milan, Italy. arXiv:2401.00466 [cs].

Peter, S. D., Cancino-Chacón, C. E., Foscarin, F., McLeod, A. P., Henkel, F., Karystinaios, E., and Widmer, G. (2023). Automatic Note-Level Score-to-Performance Alignments in the ASAP Dataset. *Transactions of the International Society for Music Information Retrieval (TISMIR)*.

Peter, S. D. and Widmer, G. (2024). Thegluenote: Learned representations for robust and flexible note alignment. In *International Society for Music Information Retrieval Conference (ISMIR 2024)*, San Francisco, USA.

Rabiner, L. and Juang, B. (1986). An introduction to hidden markov models. *IEEE ASSP Magazine*, 3(1):4–16.

Raphael, C. and Gu, Y. (2009). Orchestral Accompaniment for a Reproducing Piano. In *Proceedings of the International Computer Music Conference (ICMC 2009)*, Montreal, Canada.

Sarath, E. (2013). *Music theory through improvisation: A new approach to musicianship training*. Routledge.

Simonetta, F., Cancino-Chacón, C., Ntalampiras, S., and Widmer, G. (2019). A convolutional approach to melody line identification in symbolic scores. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 924–931, Delft, Netherlands.

Spiliopoulou, A. and Storkey, A. (2011). Comparing probabilistic models for melodic sequences. In Gunopulos, D., Hofmann, T., Malerba, D., and Vazirgiannis, M., editors, *Machine Learning and Knowledge Discovery in Databases*, pages 289–304, Berlin, Heidelberg. Springer Berlin Heidelberg.

Thickstun, J., Hall, D., Donahue, C., and Liang, P. (2023). Anticipatory Music Transformer. *arXiv preprint arXiv:2306.08620*.

Uitdenbogerd, A. and Zobel, J. (1999). Melodic matching techniques for large music databases. In *Proceedings of the ACM international conference on Multimedia*, pages 57–66.

Vercoe, B. (1984). The synthetic performer in the context of live performance. In *Proceedings of International Computer Music Conference*, pages 199–200, Paris, France.

Wead, A. and Knopke, I. (2007). A computer-based implementation of basso continuo rules for figured bass realizations. In *ICMC*.

Williams, P. and Cafiero, R. (2001). *Partimento*, volume 1. Oxford University Press.

Williams, P. and Ledbetter, D. (2001). *Continuo*, volume 1. Oxford University Press.

Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 143–146.