PIANOFINGERPRINTS: REDISCOVERING CHOPIN VIA COMPOSITION-AWARE CONTRASTIVE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Music discovery from symbolic archives requires relationships among works that persist across heterogeneous collections and MIDI realizations. We propose PianoFingerprints, a composition-anchored multilevel atlas combining co-registered three-channel time–pitch MIDI tensors, ViT-Base, and source-priority multi-positive contrastive learning. A segment–realization–work hierarchy separates reproducible work structure from archive and realization variation. Across 47,489 realizations from 11 archives, KDF is held out from training for cross-archive evaluation and fitted separately for single-archive discovery. Held-out KDF contains 232 nonempty realizations of 139 normalized works. The performance-archive geometry agrees with work geometry independently aggregated from held-out KDF ($r_s = 0.767$), showing that relative work similarities recur across archives. Two nonoverlapping halves of the KDF realizations also yield similar work geometries ($r_s = 0.870$), indicating stability across different realizations of the same works. In the single-archive atlas, mean within-genre similarity exceeds mean cross-genre similarity (0.199 versus 0.045). There is no evidence that work distinctiveness has a statistically significant global linear association with composition year ($r = -0.148$, $p = 0.075$). For 66 cross-genre neighbour pairs reproduced in both archive views, the closest ten-second passages have a mean symbolic similarity of 0.548. Shuffling specific work identities within each genre, while retaining the same pair of genres for every relation, yields a permutation mean of 0.403 ($\Delta = 0.145$, $p_{\text{perm}} = 1.0 \times 10^{-4}$). Together, these results position PianoFingerprints as a corpus-conditioned discovery instrument for Chopin archives, using cross-archive replication to identify reproducible structure and local passages to support interpretation.

1 INTRODUCTION

Symbolic piano archives combine MIDI entered from scores, captured from instrument performances, and generated through automatic transcription of audio (Hawthorne et al., 2019; Zhang et al., 2022; Kong et al., 2022; Bradshaw & Colton, 2025; Borovik, 2026). We call each stored MIDI instantiation of a work a *MIDI realization*. Realizations of the same work share musical content while differing in timing, velocity, articulation, editing, segmentation, and transcription error. These differences make independently assembled archives useful for music discovery because relationships that recur across sources are supported by multiple realizations of the repertoire.

Archive-scale analysis therefore requires a representation that reconnects realizations of the same work and retains continuous relationships among different works. It must aggregate local passages through realization and work levels. Performance research treats timing and dynamics as central musical variables (Gabrielsson, 2003; Cancino-Chacón et al., 2018; Oore et al., 2020), and aligned symbolic datasets preserve these variables across piano realizations (Foscarin et al., 2020; Hawthorne et al., 2019; Kim et al., 2025). Stable work-level coordinates make these heterogeneous observations comparable.

Transformers have modeled long-range symbolic structure, controllable generation, and pretrained musical understanding (Huang et al., 2019; Wang & Xia, 2021; Yu et al., 2022; Zeng et al., 2021). Contrastive learning has produced reusable music representations and cross-modal symbolic retrieval systems (Khosla et al., 2020; Spijkervet & Burgoyne, 2021; Wu et al., 2023; 2025; Bradshaw et al.,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

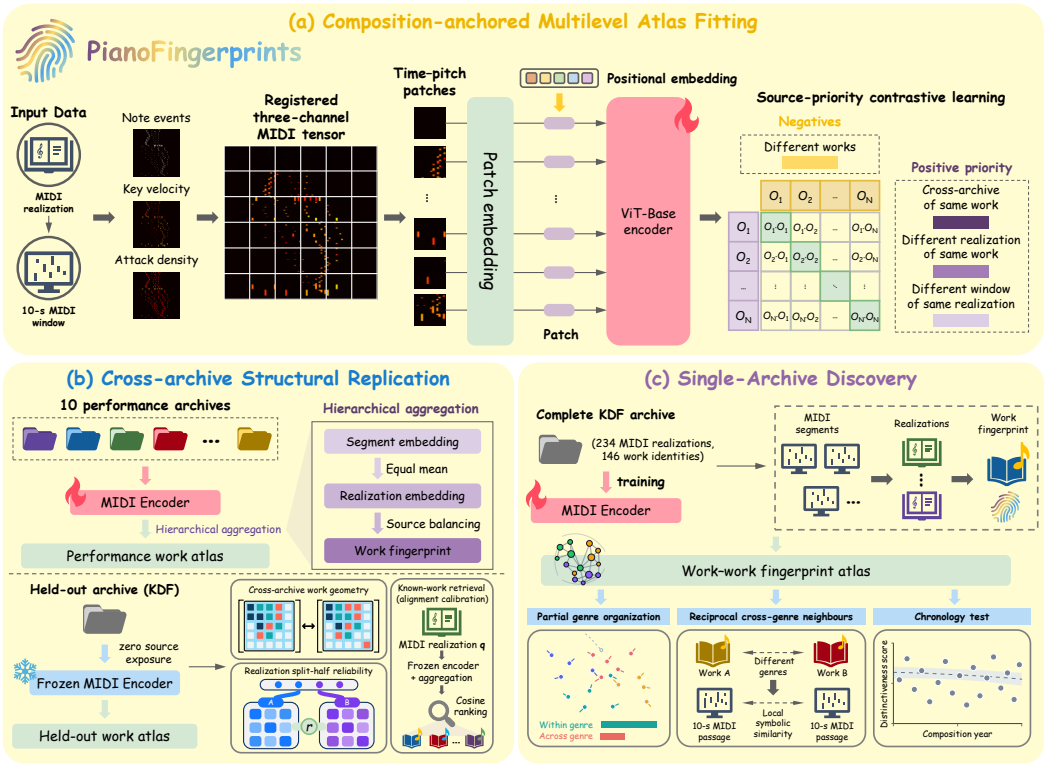


Figure 1: PianoFingerprints fits a composition-anchored multilevel atlas and separates structural replication from corpus discovery. (a) Each 10-second MIDI realization window is represented by co-registered activity, normalized velocity, and attack density, divided into time–pitch patches, and encoded by ViT-Base. Source-priority sampling favors a same-work realization from another archive, then another realization from the same archive, and finally another window from the same realization; all same-work observations already in the batch are positives and all different-work observations are negatives. Source priority governs positive selection, not the marginal distribution of anchors. (b) For cross-archive structural replication, the encoder is fitted on ten performance-MIDI archives (47,256 realizations, 222 works) and frozen before evaluation on held-out KDF (232 realizations, 139 works). Segment embeddings are averaged within realizations and source-balanced across the performance archives to form work fingerprints. Independently aggregated work geometry and realization split-half reliability are the structural endpoints, while known-work retrieval is an alignment calibration. (c) For single-archive discovery, the same method definition is fitted to the complete KDF archive and interrogated for partial genre organization, reciprocal cross-genre neighbours with local passage support, and chronological structure.

2025). These systems mainly address generation, tagging, or cross-modal retrieval. Corpus discovery additionally requires positives defined by shared composition identity, co-registered performance variables, and explicit aggregation from passages to works.

We propose PianoFingerprints, a composition-anchored multilevel atlas for discovering relationships among works across heterogeneous MIDI archives. We first ask whether work relationships constructed from the Kunst der Fuge (KDF) archive (Kunst der Fuge, 2022) recur in ten independently sourced performance-MIDI archives. We then fit KDF directly to examine genre organization, reciprocal cross-genre neighbours and their local-passage support, and the association between work distinctiveness and composition year.

Our main contributions are:

- **Composition-anchored multilevel atlas.** We co-register note activity, velocity, and attack density, then aggregate segment embeddings through realization, archive, and work levels to construct comparable work fingerprints.

- **Source-priority multi-positive learning.** We prioritize same-work positives across archives, realizations, and windows while using all same-work and different-work batch observations in the contrastive objective. The hierarchy reduces automatically to realization and window tiers for a single archive.
- **Cross-archive structural replication.** We compare independently aggregated work–work geometries from ten performance-MIDI archives and held-out KDF, together with realization split-half reliability, matched objectives, and modern symbolic encoders.
- **Locally supported Chopin discovery.** We identify cross-archive-consistent cross-genre neighbours and their closest symbolic passages, characterize partial genre organization, and find no statistically significant global linear association with composition year.

2 Related Work

Symbolic representations and piano realizations. Symbolic music can be encoded as event sequences, time–pitch matrices, or graphs (Le et al., 2025; Zhang et al., 2023). Engineered descriptors instead summarize pitch, interval, rhythm, dynamics, density, and texture (Eerola & Toiviainen, 2004; McKay et al., 2018). Event-sequence and graph-based models retain fine-grained temporal organization beyond fixed descriptor inventories (Walder, 2016; Zhang et al., 2023). Voice separation, alignment, and transcription studies further show that polyphonic and timing ambiguities persist after music has been converted to symbolic form (McLeod & Steedman, 2016; Peter & Widmer, 2024; Benetos et al., 2019; Hawthorne et al., 2018). PianoFingerprints co-registers note activity, velocity, and attack density on the same time–pitch grid, giving corresponding tensor locations the same temporal and piano-key coordinates.

Piano-performance research has modeled timing, dynamics, and interpretation with rule-based, probabilistic, and recurrent methods (Widmer & Goebl, 2004; Gu & Raphael, 2012; Oore et al., 2020; Cancino-Chacón et al., 2018). Computational analyses of music emotion and cognitive accounts of music perception provide complementary perspectives on how measurable musical attributes relate to affect and listener response (Panda et al., 2023; Vuust et al., 2022). Large piano resources such as GiantMIDI-Piano, ATEPP, MAESTRO, ASAP, and Aria-MIDI collect symbolic realizations produced by different capture and transcription pipelines (Kong et al., 2022; Zhang et al., 2022; Hawthorne et al., 2019; Foscarin et al., 2020; Bradshaw & Colton, 2025). This heterogeneity motivates the PianoFingerprints hierarchy: local segments are first aggregated within realizations and archives before source-balanced work fingerprints are compared.

Transformers and contrastive music learning. Transformers provide a common encoder family for both grid and event representations. Patch-based vision transformers retain global self-attention over a tokenized two-dimensional grid (Dosovitskiy et al., 2021), while symbolic-music transformers model long-range timing, controllable generation, pretrained understanding, and domain-informed geometry (Huang et al., 2019; Tian et al., 2025; Wang & Xia, 2021; Yu et al., 2022; Guo et al., 2023). For a discovery atlas, however, the encoder architecture determines representational capacity but not which musical relations its geometry should preserve. Contrastive learning supplies this relation through explicitly defined positive and negative sets (Chen et al., 2020; Khosla et al., 2020; Radford et al., 2021), and music applications have used such objectives for transferable audio features and cross-modal symbolic retrieval (Spijkervet & Burgoyne, 2021; Guinot et al., 2024; Wu et al., 2023; 2025). PianoFingerprints combines a ViT-Base encoder with positives defined by shared composition identity and prioritized across archives, realizations, and windows, while observations from different works form negatives.

Large-scale music pretraining learns complementary invariances from audio and symbolic corpora. CLMR evaluates contrastive audio transfer through downstream tagging (Spijkervet & Burgoyne, 2021), Audio-MAE reconstructs masked spectrogram patches (Huang et al., 2022), and MERT combines acoustic and musically informed teachers (Li et al., 2024). Symbolic systems instead model event sequences, autoregressive performance MIDI, or music–language alignment, as in MusicBERT, Aria, and CLaMP/CLaMP 2 (Zeng et al., 2021; Bradshaw et al., 2025; Wu et al., 2023; 2025). These models target transfer, generation, or cross-modal alignment rather than the construction of an archive-conditioned work–work coordinate. We therefore use MusicBERT, Aria, and CLaMP 2 as modern representation references, while PianoFingerprints couples composition-identity supervision to explicit segment–realization–archive–work aggregation.

Corpus organization. Computational musicology combines quantitative measurement with historically informed interpretation (Clarke & Cook, 2004; Mor et al., 2020). Corpus-level conclusions depend on sampling, metadata, representation, and the chosen unit of analysis. Guidelines for multimodal music corpora additionally emphasize clear source definitions, citability, and long-term access (Gotham et al., 2025). PianoFingerprints makes the analytical hierarchy explicit by separating passage-level evidence, realization-level reliability, and work-level genre or chronology analyses. Cross-archive work-matrix agreement tests whether the inferred organization recurs across independently assembled sources, while low-dimensional projections remain descriptive visualizations of the original embedding geometry.

3 METHOD

3.1 OVERVIEW

PianoFingerprints maps local MIDI passages into a composition-anchored coordinate and aggregates them into work fingerprints (Figure 1). Each 10-second window contains co-registered activity, normalized velocity, and attack-density planes on an identical time–pitch grid. ViT-Base encodes the resulting patches as unit-normalized segment embeddings. Same-work observations form positives across archives, realizations, or windows, while different-work observations form negatives.

The single-archive and multi-archive regimes share the encoder and batch-all multi-positive objective. In KDF, designated positives prioritize another realization and then another window. Multi-archive fitting adds a cross-source realization as the first tier. Equal-weight aggregation then maps segments to realizations, archive-specific work centroids, and source-balanced work fingerprints. The external model is fitted on ten performance archives and evaluated on held-out KDF, while the KDF fit constructs the discovery atlas used for genre, passage, and chronology analyses.

3.2 REPRESENTATION LEVELS

Let c index a composition identity, and let $X_{c,a,r,s}$ denote segment s from MIDI realization r of the work identified by c in archive a. The encoder f_θ maps it to $z_{c,a,r,s} \in \mathbb{R}^{256}$ with $\|z_{c,a,r,s}\|_2 = 1$. Writing $\mathrm{nrm}(v) = v/\|v\|_2$, $\mathcal{R}_{c,a}$ for the retained realizations of that work in archive a, and $\mathcal{A}_c$ for the archives containing it, we aggregate without refitting:

$$u_{c,a,r} = \mathrm{nrm}\left(\frac{1}{S_{c,a,r}} \sum_{s=1}^{S_{c,a,r}} z_{c,a,r,s}\right), \tag{1}$$

$$v_{c,a} = \mathrm{nrm}\left(\frac{1}{|\mathcal{R}_{c,a}|} \sum_{r \in \mathcal{R}_{c,a}} u_{c,a,r}\right), \qquad e_c = \mathrm{nrm}\left(\frac{1}{|\mathcal{A}_c|} \sum_{a \in \mathcal{A}_c} v_{c,a}\right), \tag{2}$$

where $S_{c,a,r}$ is the number of retained windows. The hierarchy first gives equal weight to realizations and then equal weight to archives, preventing long files or large sources from dominating an inference-time work fingerprint. In a single archive, $e_c = v_{c,a}$, so the final source-balancing stage disappears automatically.

3.3 CO-REGISTERED TIME–PITCH INPUT

MIDI files are sampled on a 100-Hz grid and divided into 10-second windows. A window becomes $X \in [0,1]^{3\times 88\times 1000}$ over the standard 88 piano pitches. The first channel records key-down note activity from onset to offset. The second stores note-on velocity divided by 127; within sustain-control intervals, a cumulative maximum extends the velocity roll over time. The third writes a local attack-count value at each attacked pitch. For note n, let $\rho_n = |\{m : |t_m - t_n| < w/2\}|$ denote the number of attacks in a centered $w = 0.2$-s neighborhood. The density plane is

$$d(t,p) = \frac{\min\Big(10 \sum_{n:p_n=p} \mathbf{1}(|t - t_n| < w/2)\rho_n, 127\Big)}{127}, \tag{3}$$

where t_n and p_n are the onset and pitch of note n. The multiplier rescales sparse MIDI attack counts before clipping. Appendix B visualizes the aligned planes.

3.4 Encoder and learning objective

The 88×1000 support is partitioned into nonoverlapping 11×50 patches, yielding $8 \times 20 = 160$ tokens. For $X^{(q)} \in \mathbb{R}^{3\times11\times50}$, the flattened 1,650-vector is projected by $W_{\text{patch}} \in \mathbb{R}^{768\times1650}$ and combined with a learned positional embedding $p_q \in \mathbb{R}^{768}$:

$$x_q = W_{\text{patch}}\,\text{vec}(X^{(q)}) + p_q, \qquad q = 1, \ldots, 160. \tag{4}$$

A ViT-Base-style encoder applies global average pooling and a projection head to produce the normalized 256-dimensional embedding.

Each batch contains 32 anchors and one designated positive per anchor. The designated positive is drawn by the explicit archive–realization–window hierarchy in Appendix Eq. A.1; when only one archive exists, its cross-archive tier is empty and the rule reduces to another realization, then another window. Let $\mathcal{B}$ be the resulting 64 observations and $P(i) = \{p \in \mathcal{B} \setminus \{i\} : c_p = c_i\}$ the complete set of same-work observations available to anchor i. The designated pairing guarantees $P(i) \neq \varnothing$, but the loss uses *all* same-work batch observations as positives and every different-work observation as a negative. With cosine similarity $\text{sim}(x, y) = x^\top y$ for unit embeddings and temperature $\tau = 0.07$, PianoFingerprints uses the supervised contrastive objective (Khosla et al., 2020)

$$\mathcal{L} = -\frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}}\frac{1}{|P(i)|}\sum_{p\in P(i)}\log\frac{\exp(\text{sim}(z_i, z_p)/\tau)}{\sum_{j\in\mathcal{B}\setminus\{i\}}\exp(\text{sim}(z_i, z_j)/\tau)}. \tag{5}$$

The loss is identical in both regimes; only anchor traversal and designated-positive construction change. Multi-archive anchors are realizations shuffled without replacement at natural file frequency, with one of four uniformly spaced windows drawn per visit. The single-archive discovery fit instead uses work-distinct batches that traverse all 234 discovery-view files and shuffled within-file window decks. Thus source priority governs positive selection, not source-balanced optimization. Controlled comparisons share the 86.65M-parameter ViT-Base, co-registered input, batch size, 8,000 updates, schedule, precision, and initialization protocol. Standard SupCon (Khosla et al., 2020) removes only source priority. The hierarchical sampled contrastive variant uses an NCE-style estimator (Gutmann & Hyvärinen, 2010), while paired InfoNCE (van den Oord et al., 2018), Composition CE (Goodfellow et al., 2016), and triplet learning (Schroff et al., 2015) replace the estimator or objective. Composition CE applies cross-entropy classification to the composition-identity labels. Appendices D and E report the full sampling policy.

3.5 Discovery operators

We analyze the learned coordinate at segment, realization, and work resolutions. At the realization level, we compare distributions with MMD-, Wasserstein-, Jensen–Shannon-, Kullback–Leibler-, cosine-, and Chamfer-derived similarities. At retrieval time, segment-to-segment similarities are pooled first within candidate realizations and then equally across the realizations of a work. At the work level, K-means and agglomerative clustering probe unsupervised grouping, genre-conditioned cosine summaries measure alignment with conventional metadata, and three chronological coordinates quantify distance from an early-period centroid, distinctiveness from the remaining corpus, and consistency within genre. Inference-time channel masking and response maps localize model dependence without changing the fitted encoder. Quantitative comparisons use the original embedding space, while projections and response maps visualize local organization and model dependence.

4 Experimental Design

4.1 Single-archive discovery corpus

For clarity, *KDF discovery view* denotes the 234 tensorized MIDI files and 146 project-level work labels used for the single-archive atlas, whereas *held-out KDF view* denotes the 232 nonempty realizations and 139 normalized works used for cross-archive validation. The two views apply task-specific parsing and identity normalization to the same KDF collection (Appendix Table C.1). The discovery view contains 23,976 windows. Seventy-seven labels have one realization, 53 have two, 13 have three, and three have four. Windows use a 2.5-s stride, producing 75% overlap. Seventeen genre labels occur overall, with strong imbalance in both works and windows (Appendix C).

The files were collected from the Chopin archive hosted by Kunst der Fuge (Kunst der Fuge, 2022). A prior comparison characterizes the broader collection as sequencer-entered MIDI (Kong et al., 2022). Project-level work labels supplied with the collection define the fixed discovery view used in this study.

4.2 Cross-archive structural replication

The cross-archive collection contains 47,489 retained records from 11 archive labels and 222 normalized identities. The other ten archives contribute 47,256 realizations for fitting. The held-out KDF view contributes 232 nonempty realizations and 139 normalized works. Excluding composite or ambiguous identities leaves 211 KDF files and 129 shared single-work identities for the primary geometry audit. Performance centroids are source balanced at aggregation, while training anchors retain natural source frequencies. Work labels align matrix rows but do not specify the $\binom{129}{2}$ tested relations. Tables D.1 and C.1 report the archive inventory and analysis-specific sample counts.

References include symbolic pitch–rhythm–dynamics descriptors (McKay et al., 2018), plus matched Standard SupCon (Khosla et al., 2020), an NCE-style hierarchical sampled contrastive variant (Gutmann & Hyvärinen, 2010), paired InfoNCE (van den Oord et al., 2018), triplet learning (Schroff et al., 2015), and Composition CE (Goodfellow et al., 2016). Modern references comprise Aria (Bradshaw et al., 2025), CLaMP 2 (Wu et al., 2025), and MusicBERT (Zeng et al., 2021), with MusicBERT evaluated before and after task adaptation. Controlled ViT comparisons share the architecture, input, split, and training budget. Each trainable cross-archive method contributes one completed fit, while the remaining fixed representations contribute point estimates.

4.3 Discovery-specific validity tests

The single-archive KDF atlas is fitted with all 23,976 overlapping 10-second windows extracted at a 2.5-second stride. The four-window subset described below is used only for post-hoc local-passage corroboration and does not enter single-archive fitting. We define a directed neighbour before inspecting passages: a work must be among the other's top five neighbours in both the KDF-only and performance-only geometries of the reported cross-archive fit. Reciprocal cross-genre relations form the discovery set. For this local test, four uniformly spaced 10-second windows from every eligible KDF realization of an endpoint work are summarized by a fixed 166-dimensional pitch-class, register, temporal-activity, density, velocity, and global-statistics descriptor assembled from established symbolic feature families (Eerola & Toiviainen, 2004; McKay et al., 2018). Each relation receives the largest descriptor cosine across all cross-work window pairs, and the graph statistic averages these relation-level maxima. We compare it with 9,999 genre-preserving work-identity permutations that retain every network position and endpoint genre. We repeat the complete graph-selection procedure for every matched learned objective before applying the same local endpoint.

Local-neighbour validation is reported in Appendix D.3. Single-archive attention views in the supplementary material visualize response concentrations from the corresponding fitted models.

4.4 Internal audit and secondary analyses

Within the KDF discovery view, the internal audit uses all 69 works with at least two realizations. Its 157 query files are ranked against the 146 project-level candidate works after excluding the query file from its prototype. Every file contributes to encoder fitting, and Appendix F reports these fitted-atlas construction metrics. We also recompute the original inference-time channel-masking analysis with the reported single-archive ViT-Base encoder. A separate refitting analysis compares the available channel combinations.

At the work level, we examine K-means and agglomerative clustering for $k = 2, \ldots, 20$ using silhouette, adjusted Rand index (ARI), and normalized mutual information (NMI) against conventional genre labels. We also compute early-centroid distance, corpus distinctiveness, and within-genre consistency as chronological coordinates. Attention views show response concentrations for selected works and passages. Appendix Figures F.1 and F.2 distinguish inference-time channel sensitivity from independently refitted input combinations. Complete definitions and supporting results are in Appendices H–I.

Results first report cross-archive work-geometry replication and then the KDF genre, local-neighbour, and chronology analyses. Retrieval calibration and channel sensitivity provide construction diagnostics.

5 RESULTS

5.1 CHOPIN WORK STRUCTURE IS REPRODUCED ACROSS ARCHIVES

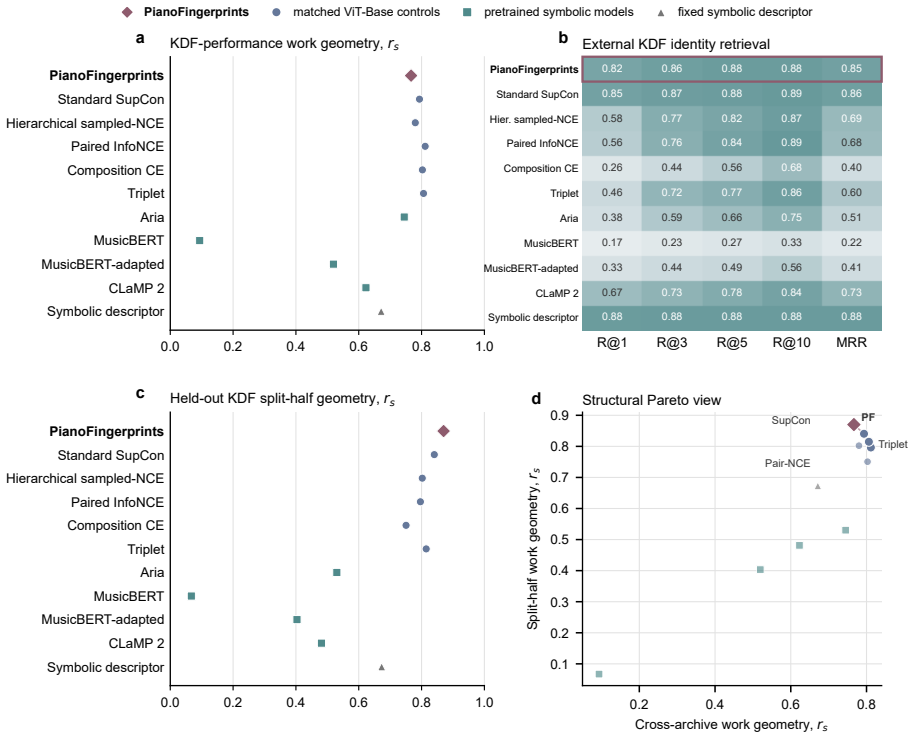


Figure 2: Cross-archive structural replication and alignment calibration. Aria, CLaMP 2, and the symbolic descriptor are included for comparison; MusicBERT is shown before and after task adaptation. (a) Spearman agreement between the upper triangles of independently aggregated performance-atlas and KDF work-similarity matrices over 129 shared works, quantifying whether relative work similarities recur across archives. (b) Work-macro Recall@1, Recall@3, Recall@5, Recall@10, and MRR for unseen-KDF identity retrieval. (c) Mean Spearman agreement between two work-similarity matrices formed from nonoverlapping realization halves in 50 repeats over 64 eligible unseen-KDF works. (d) Joint view of the two structural correlations. The connecting line traces the outer boundary of the observed methods toward the upper-right corner, where both cross-archive replication and split-half reliability are stronger.

The performance-atlas and held-out-KDF work-similarity matrices agree at $r_s = 0.767$ over 129 shared works. This coefficient ranks all 8,256 off-diagonal work-pair cosines in the two independently aggregated matrices. Matched ViT objectives reach $r_s = 0.780$–0.811. PianoFingerprints exceeds Aria (0.745), CLaMP 2 (0.623), MusicBERT and MusicBERT-adapted (0.094 and 0.519), and the symbolic descriptor (0.671). Figure 2 summarizes the four evaluation views, and Appendix Table D.2 reports every point estimate.

Identity retrieval measures whether held-out KDF realizations enter the correct known-work region. PianoFingerprints reaches Recall@1/Recall@3/Recall@5/Recall@10/MRR of 0.824/0.860/0.880/0.880/0.849. Standard SupCon and the symbolic descriptor give slightly higher full-realization Recall@1 than PianoFingerprints (0.846 and 0.880 versus 0.824). PianoFingerprints, however, has higher split-half geometry than both (0.870 versus 0.841 and 0.673), higher

split-half Recall@1 (0.917 versus 0.883 and 0.906), and higher cross-archive geometry than the descriptor (0.767 versus 0.671). It also leads both comparators in four-window 10-second segment retrieval (Appendix Table D.4). Full-realization retrieval and structural reliability therefore describe complementary properties of the atlas.

Realization split-half reliability evaluates reconstruction within held-out KDF. Across 50 nonoverlapping splits of 64 eligible works, PianoFingerprints obtains Recall@1 0.917, work-matrix agreement $r_s = 0.870$, chance-adjusted neighbour overlap@5 0.582, and same-work–nearest-impostor margin 0.064. Standard SupCon reaches 0.883, 0.841, 0.577, and 0.063, respectively. PianoFingerprints has the highest Recall@1 and geometry reliability among the learned rows.

Appendix F reports identity alignment within the fitted single-archive atlas. Recall@1/Recall@3/Recall@5/Recall@10/MRR are $0.790/0.962/0.987/1.000/0.873$, with every query contributing to encoder fitting.

5.2 The KDF atlas reveals partial genre organization

Genre coherence is uneven in the KDF atlas. Mean within-genre work similarity is 0.199, compared with 0.045 across genres. Figure 3a shows high positive central cosines for ballades, nocturnes, scherzos, and sonatas, although the latter two contain few works. The distributions for *études*, mazurkas, preludes, and waltzes lie nearer zero and span positive and negative values, combining weak shared direction with substantial within-label variation. Polonaises are intermediate, and the centroid matrix in Figure 3b shows graded relations rather than uniformly compact genre groups.

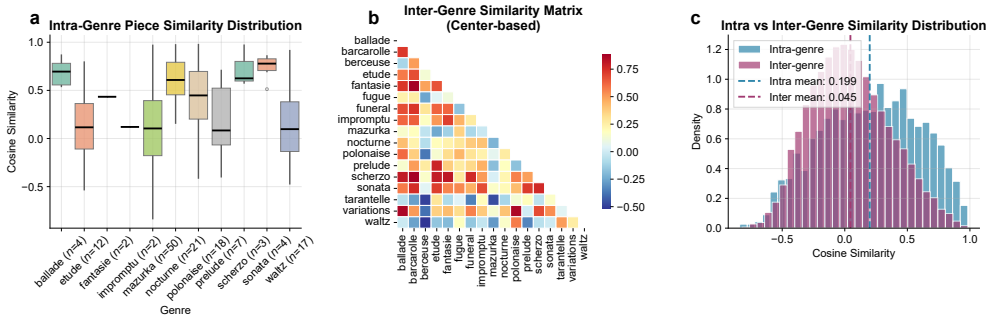


Figure 3: Work-level genre relations in the fitted KDF coordinate. (a) Within-genre work-pair cosine distributions; n denotes the number of works carrying each label. (b) Cosine similarities among genre centroids. (c) Pooled within-genre and sampled cross-genre work-pair distributions; dashed lines mark the displayed means. Panels summarize fitted-corpus geometry with repeated works and unequal genre support.

Cross-archive-consistent neighbours retain local support. We next asked whether cross-archive work neighbours are supported by specific local passages. Different-genre works that were reciprocal top-five neighbours in both atlas views yielded 66 undirected relations among 75 works. For each relation, every retained window from one work was compared with every retained window from the other using a fixed 166-dimensional symbolic descriptor that was not used to fit PianoFingerprints. The largest window-pair cosine defined the relation's best-passage score, whose mean was 0.548.

The genre-preserving work-identity permutation shuffled identities only within each genre. Every relation retained its network position and its pair of genre labels, so it remained cross-genre while each endpoint work was replaced by a randomly selected work from the same genre. We recomputed every best-passage score and their mean after each shuffle. Across 9,999 permutations, the permutation mean was 0.403 and the 95% permutation interval was $[0.354, 0.451]$. No permuted mean reached 0.548 ($\Delta = 0.145$, plus-one $p_{\text{perm}} = 1.0 \times 10^{-4}$). Thus, the particular work pairs reproduced as neighbours across archives share more similar local symbolic passages than arbitrary work pairs drawn from the same two genres. Appendix Figure D.2 gives relation-level scores, the permutation distribution, and the 20 highest-scoring named pairs.

5.3 CHRONOLOGY DOES NOT DEFINE A SINGLE GLOBAL AXIS

There is no evidence of a statistically significant global linear association between work distinctiveness and composition year ($r = -0.148$, nominal $p = 0.075$, $n = 146$). The KDF atlas therefore does not support a single monotonic account of Chopin's compositional development. Figure 4a,d instead shows irregular work-to-work displacement and genre traces evolving in different directions, motivating period-specific, nonlinear, or genre-conditioned paths that may disappear in a corpus-wide regression. Testing this hypothesis requires joint modelling of genre, life period, and nonlinear time.

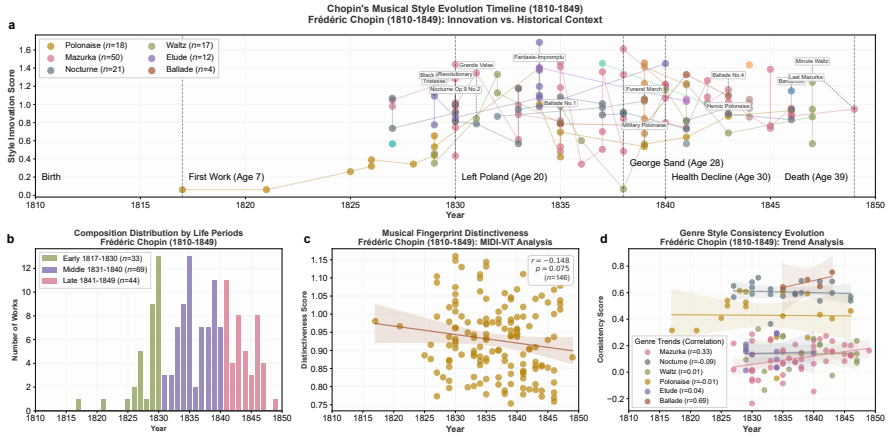


Figure 4: Chronological coordinates derived from fitted KDF work embeddings. (a) Early-centroid distance with historical markers. (b) Represented works by year and life period. (c) Work distinctiveness versus year. (d) Within-genre consistency for the six largest genres. There is no evidence of a statistically significant global linear association between distinctiveness and year ($r = -0.148$, nominal $p = 0.075$, $n = 146$).

6 DISCUSSION

Rediscovery as multilevel corpus reconstruction. The single-archive atlas organizes passages, realizations, and works in a shared coordinate. The KDF holdout then tests whether its 129×129 work geometry recurs across independently assembled sources. Several objectives attain high cross-archive agreement, while PianoFingerprints has the strongest learned split-half geometry in the reported fits. This multilevel protocol connects structural replication to genre, cross-genre neighbours, local passage support, and chronology.

Atlas extensibility. Single-archive discovery and cross-archive replication show how a fitted atlas can connect passages, MIDI realizations, and work-level structure. Adding source records, performers, editions, and transcription processes could partition fingerprint variation into compositional, interpretive, and encoding components. Applying the hierarchy to other composers or mixed corpora would enable comparative fingerprint atlases across repertoires, performance traditions, and collection pipelines.

7 CONCLUSION

PianoFingerprints transforms heterogeneous symbolic realizations into a multilevel discovery atlas whose work relationships can be reconstructed across independent archives. The KDF atlas reveals graded genre organization, cross-genre neighbours grounded in locally similar passages, and no evidence for one global linear chronology. Alignment calibration and structural replication make these discoveries testable beyond their source archive, while passage analysis exposes local evidence behind global relations. This establishes the feasibility of a reproducible symbolic-music discovery workflow. Richer metadata and additional composers could extend it toward comparative fingerprints of compositional organization, repertoire, and performance.

AI USE STATEMENT

Generative AI tools were used for English-language polishing. The authors reviewed the edited text and take responsibility for the final content.

ETHICS STATEMENT

This study analyzes publicly accessible symbolic music archives and involves no interaction with human participants.

REPRODUCIBILITY STATEMENT

The Method and Experimental Design sections specify the representation, learning objective, aggregation hierarchy, fitting procedure, and evaluation protocols. Supplementary Sections C–I provide archive sources, analysis-specific sample accounting, architecture and fitting settings, objective comparisons, ablations, retrieval definitions, complete supporting results, and statistical procedures.

REFERENCES

E. Benetos, S. Dixon, Z. Duan, and S. Ewert. Automatic music transcription: An overview. *IEEE Signal Processing Magazine*, 36(1):20–30, 2019. doi: 10.1109/MSP.2018.2869928. URL https://doi.org/10.1109/MSP.2018.2869928.

I. Borovik. PianoCoRe: Combined and refined piano MIDI dataset. *Transactions of the International Society for Music Information Retrieval*, 9(1):144–163, 2026. doi: 10.5334/tismir.333. URL https://doi.org/10.5334/tismir.333.

I. Borovik, D. Gavrilev, and V. Viro. SyMuPe: Affective and controllable symbolic music performance. In *Proc. 33rd ACM Int. Conf. Multimedia*, pp. 10699–10708, 2025. doi: 10.1145/3746027.3755871. URL https://doi.org/10.1145/3746027.3755871.

L. Bradshaw and S. Colton. Aria-MIDI: A dataset of piano MIDI files for symbolic music modeling. In *Proc. Int. Conf. Learning Representations (ICLR)*, 2025. URL https://openreview.net/forum?id=X5hrhgndxW.

L. Bradshaw, H. Fan, A. Spangher, S. Biderman, and S. Colton. Scaling self-supervised representation learning for symbolic piano performance. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, 2025. URL https://arxiv.org/abs/2506.23869.

C. E. Cancino-Chacón, M. Grachten, W. Goebl, and G. Widmer. Computational models of expressive music performance: A comprehensive and critical review. *Frontiers in Digital Humanities*, 5:25, 2018. doi: 10.3389/fdigh.2018.00025. URL https://doi.org/10.3389/fdigh.2018.00025.

T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *Proc. Int. Conf. Machine Learning*, pp. 1597–1607, 2020. URL https://proceedings.mlr.press/v119/chen20j.html.

E. Clarke and N. Cook (eds.). *Empirical Musicology: Aims, Methods, Prospects*. Oxford University Press, Oxford, U.K., 2004. URL https://global.oup.com/academic/product/empirical-musicology-9780195167498.

A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16×16 words: Transformers for image recognition at scale. In *Proc. Int. Conf. Learning Representations (ICLR)*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

T. Eerola and P. Toiviainen. MIR in MATLAB: The MIDI toolbox. In *Proc. 5th Int. Conf. Music Information Retrieval (ISMIR)*, pp. 22–27, 2004. URL https://archives.ismir.net/ismir2004/paper/000193.pdf.

F. Foscarin, A. McLeod, P. Rigaux, F. Jacquemard, and M. Sakai. ASAP: A dataset of aligned scores and performances for piano transcription. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 534–541, 2020. URL https://archives.ismir.net/ismir2020/paper/000127.pdf.

A. Gabrielsson. Music performance research at the millennium. *Psychology of Music*, 31(3):221–272, 2003. doi: 10.1177/03057356030313002. URL https://doi.org/10.1177/03057356030313002.

W. Goebl. The vienna 4x22 piano corpus, 1999. URL https://doi.org/10.21939/4X22. Dataset, version 1.0.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. URL https://www.deeplearningbook.org/.

M. Gotham, B. Bemman, and I. Vatolkin. Towards an "everything corpus": A framework and guidelines for the curation of more comprehensive multimodal music data. *Transactions of the International Society for Music Information Retrieval*, 8(1):70–92, 2025. doi: 10.5334/tismir.228. URL https://doi.org/10.5334/tismir.228.

Y. Gu and C. Raphael. Modeling piano interpretation using switching Kalman filter. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 145–150, 2012. URL https://archives.ismir.net/ismir2012/paper/000145.pdf.

J. P. M. Guinot, E. Quinton, and G. Fazekas. Semi-supervised contrastive learning of musical representations. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 571–579, 2024. doi: 10.5281/zenodo.14877401. URL https://doi.org/10.5281/zenodo.14877401.

Z. Guo, J. Kang, and D. Herremans. A domain-knowledge-inspired music embedding space and a novel attention mechanism for symbolic music modeling. In *Proc. AAAI Conf. Artificial Intelligence*, volume 37, pp. 5070–5077, 2023. doi: 10.1609/aaai.v37i4.25635. URL https://doi.org/10.1609/aaai.v37i4.25635.

M. Gutmann and A. Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proc. 13th Int. Conf. Artificial Intelligence and Statistics*, volume 9, pp. 297–304. PMLR, 2010. URL https://proceedings.mlr.press/v9/gutmann10a.html.

C. Hawthorne, E. Elsen, J. Song, A. Roberts, I. Simon, C. Raffel, J. Engel, S. Oore, and D. Eck. Onsets and frames: Dual-objective piano transcription. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 50–57, 2018. URL https://archives.ismir.net/ismir2018/paper/000019.pdf.

C. Hawthorne, A. Stasyuk, A. Roberts, I. Simon, C.-Z. A. Huang, S. Dieleman, E. Elsen, J. Engel, and D. Eck. Enabling factorized piano music modeling and generation with the MAESTRO dataset. In *Proc. Int. Conf. Learning Representations (ICLR)*, 2019. URL https://openreview.net/forum?id=r1lYRjC9F7.

C.-Z. A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, I. Simon, C. Hawthorne, A. M. Dai, M. D. Hoffman, M. Dinculescu, and D. Eck. Music transformer: Generating music with long-term structure. In *Proc. Int. Conf. Learning Representations (ICLR)*, 2019. URL https://openreview.net/forum?id=rJe4ShAcF7.

P.-Y. Huang, H. Xu, J. Li, A. Baevski, M. Auli, W. Galuba, F. Metze, and C. Feichtenhofer. Masked autoencoders that listen. In *Advances in Neural Information Processing Systems*, volume 35, pp. 28708–28720, 2022. doi: 10.52202/068431-2081. URL https://arxiv.org/abs/2207.06405.

P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan. Supervised contrastive learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://arxiv.org/abs/2004.11362.

Y. Kim, J. Park, J. Bae, K. Kim, T. Kwon, A. Lerch, and J. Nam. PianoVAM: A multimodal piano performance dataset. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 528–535, 2025. URL https://archives.ismir.net/ismir2025/paper/000061.pdf.

Q. Kong, B. Li, J. Chen, and Y. Wang. GiantMIDI-Piano: A large-scale MIDI dataset for classical piano music. *Transactions of the International Society for Music Information Retrieval*, 5(1): 87–98, 2022. doi: 10.5334/tismir.80. URL https://doi.org/10.5334/tismir.80.

Kunst der Fuge. Frédéric chopin: Classical MIDI archive, 2022. URL https://www.kunstderfuge.com/chopin.htm. Archive website; accessed September 1, 2026.

D.-V.-T. Le, L. Bigo, D. Herremans, and M. Keller. Natural language processing methods for symbolic music generation and information retrieval: A survey. *ACM Computing Surveys*, 57(7): 1–40, 2025. doi: 10.1145/3714457. URL https://doi.org/10.1145/3714457.

Y. Li, R. Yuan, G. Zhang, Y. Ma, X. Chen, H. Yin, C. Xiao, C. Lin, A. Ragni, E. Benetos, N. Gyenge, R. Dannenberg, R. Liu, W. Chen, G. Xia, Y. Shi, W. Huang, Z. Wang, Y. Guo, and J. Fu. MERT: Acoustic music understanding model with large-scale self-supervised training. In *Proc. Int. Conf. Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=w3YZ9MSlBu.

C. McKay, J. Cumming, and I. Fujinaga. jSymbolic 2.2: Extracting features from symbolic music for use in musicological and MIR research. In *Proc. 19th Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 348–354, 2018. URL https://archives.ismir.net/ismir2018/paper/000026.pdf.

A. McLeod and M. Steedman. HMM-based voice separation of MIDI performance. *Journal of New Music Research*, 45(1):17–26, 2016. doi: 10.1080/09298215.2015.1136650. URL https://doi.org/10.1080/09298215.2015.1136650.

B. Mor, S. Garhwal, and A. Kumar. A systematic literature review on computational musicology. *Archives of Computational Methods in Engineering*, 27(3):923–937, 2020. doi: 10.1007/s11831-019-09337-9. URL https://doi.org/10.1007/s11831-019-09337-9.

M. Müller, V. Konz, W. Bogler, and V. Arifi-Müller. Saarland music data (SMD). In *Late-Breaking and Demo Session of the Int. Soc. Music Information Retrieval Conf. (ISMIR)*, 2011. URL https://resources.mpi-inf.mpg.de/SMD/2011_MuellerKonzBoglerArifi_SaarlandMusicData_ISMIR-LateBreaking.pdf.

S. Oore, I. Simon, S. Dieleman, D. Eck, and K. Simonyan. This time with feeling: Learning expressive musical performance. *Neural Computing and Applications*, 32(4):955–967, 2020. doi: 10.1007/s00521-018-3758-9. URL https://doi.org/10.1007/s00521-018-3758-9.

R. Panda, R. Malheiro, and R. P. Paiva. Audio features for music emotion recognition: A survey. *IEEE Transactions on Affective Computing*, 14(1):68–88, 2023. doi: 10.1109/TAFFC.2020.3032373. URL https://doi.org/10.1109/TAFFC.2020.3032373.

S. D. Peter and G. Widmer. TheGlueNote: Learned representations for robust and flexible note alignment. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 603–610, 2024. doi: 10.5281/zenodo.14877409. URL https://doi.org/10.5281/zenodo.14877409.

A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning transferable visual models from natural language supervision. In *Proc. Int. Conf. Machine Learning*, pp. 8748–8763, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015. doi: 10.1109/CVPR.2015.7298682. URL https://openaccess.thecvf.com/content_cvpr_2015/html/Schroff_FaceNet_A_Unified_2015_CVPR_paper.html.

J. Spijkervet and J. A. Burgoyne. Contrastive learning of musical representations. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 673–681, 2021. URL https://archives.ismir.net/ismir2021/paper/000084.pdf.

S. Tian, C. Zhang, W. Yuan, W. Tan, and W. Zhu. XMusic: Towards a generalized and controllable symbolic music generation framework. *IEEE Transactions on Multimedia*, 27:6857–6871, 2025. doi: 10.1109/TMM.2025.3590912. URL https://doi.org/10.1109/TMM.2025.3590912.

A. van den Oord, Y. Li, and O. Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. URL https://arxiv.org/abs/1807.03748.

P. Vuust, O. A. Heggli, K. J. Friston, and M. L. Kringelbach. Music in the brain. *Nature Reviews Neuroscience*, 23(5):287–305, 2022. doi: 10.1038/s41583-022-00578-5. URL https://doi.org/10.1038/s41583-022-00578-5.

C. Walder. Modelling symbolic music: Beyond the piano roll. In *Proc. Asian Conf. Machine Learning*, pp. 174–189, 2016. URL https://proceedings.mlr.press/v63/walder88.html.

Z. Wang and G. Xia. MuseBERT: Pre-training of music representation for music understanding and controllable generation. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 722–729, 2021. URL https://archives.ismir.net/ismir2021/paper/000090.pdf.

G. Widmer and W. Goebl. Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, 33(3):203–216, 2004. doi: 10.1080/0929821042000317804. URL https://doi.org/10.1080/0929821042000317804.

S. Wu, D. Yu, X. Tan, and M. Sun. CLaMP: Contrastive language–music pre-training for cross-modal symbolic music information retrieval. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 157–165, 2023. URL https://archives.ismir.net/ismir2023/paper/000017.pdf.

S. Wu, Y. Wang, R. Yuan, Z. Guo, X. Tan, G. Zhang, M. Zhou, J. Chen, X. Mu, Y. Gao, Y. Dong, J. Liu, X. Li, F. Yu, and M. Sun. CLaMP 2: Multimodal music information retrieval across 101 languages using large language models. In *Findings of the Association for Computational Linguistics: NAACL 2025*, 2025. URL https://aclanthology.org/2025.findings-naacl.27/.

B. Yu, P. Lu, R. Wang, W. Hu, X. Tan, W. Ye, S. Zhang, T. Qin, and T.-Y. Liu. Museformer: Transformer with fine- and coarse-grained attention for music generation. *Advances in Neural Information Processing Systems*, 35:1376–1388, 2022. URL https://arxiv.org/abs/2210.10349.

M. Zeng, X. Tan, R. Wang, Z. Ju, T. Qin, and T.-Y. Liu. MusicBERT: Symbolic music understanding with large-scale pre-training. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pp. 791–800, 2021. doi: 10.18653/v1/2021.findings-acl.70. URL https://aclanthology.org/2021.findings-acl.70/.

H. Zhang, J. Tang, S. R. M. Rafee, S. Dixon, G. Fazekas, and G. A. Wiggins. ATEPP: A dataset of automatically transcribed expressive piano performance. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 446–453, 2022. URL https://archives.ismir.net/ismir2022/paper/000053.pdf.

H. Zhang, E. Karystinaios, S. Dixon, G. Widmer, and C. E. Cancino-Chacón. Symbolic music representations for classification tasks: A systematic evaluation. In *Proc. Int. Soc. Music Information Retrieval Conf. (ISMIR)*, pp. 848–858, 2023. URL https://archives.ismir.net/ismir2023/paper/000101.pdf.

A TERMINOLOGY AND NOTATION

We use *work* for the musical entity represented in the atlas and *composition identity* for the label that identifies that work and defines the contrastive relation. Accordingly, *composition-aware* and *composition-anchored* indicate that positive and negative relations are determined by shared or different composition identities; they do not introduce another aggregation level. *Co-registered* means that the activity, velocity, and attack-density channels share identical time and piano-key coordinates. An *archive* is a named data source, a *MIDI realization* is one stored file version of a work within an archive, and a *segment* is a retained 10-second window. Across the assembled sources, realizations include sequencer- or score-entered MIDI, MIDI captured from piano performances, and automatic transcriptions of performance audio. *Corpus-conditioned* means that the representation was fitted on the complete collected corpus before a downstream analysis was performed. Throughout, *PianoFingerprints* denotes the ViT-Base encoder trained with batch-all multi-positive supervised contrastive learning and the sampler defined below. Hierarchical sampled contrastive learning, source-agnostic Standard SupCon, paired InfoNCE, Composition CE, and triplet learning provide comparison objectives. Composition CE applies cross-entropy classification to the composition-identity labels. Table A.1 defines the symbols used below.

Table A.1: Symbols used in the representation and objective.

Symbol	Definition
c, a, r, s	Composition-identity index; archive index; MIDI-realization index; retained-segment index.
$X_{c,a,r,s}$	Co-registered three-channel tensor for segment s of realization r of the work identified by c in archive a.
f_θ; $z_{c,a,r,s}$	Segment encoder; its unit-normalized 256-dimensional output.
$S_{c,a,r}$; $\mathcal{R}_{c,a}$; $\mathcal{A}_c$	Number of retained segments in a realization; set of realizations of work c in archive a; set of archives containing work c.
$u_{c,a,r}$; $v_{c,a}$; e_c	Unit-normalized realization fingerprint; archive-specific work fingerprint; source-balanced work fingerprint.
$X^{(q)}$; x_q; p_q	Input patch q; its projected token; its positional embedding.
i; $P(i)$; $\mathcal{T}_k(i)$	Anchor observation; its same-work positives in the batch; its tier-k designated-positive candidates.
$\mathrm{sim}(x, y)$; τ	Cosine similarity between unit vectors; contrastive temperature.

For an anchor observation $i = (c_i, a_i, r_i, s_i)$, the designated-positive candidate tiers are

$$\begin{aligned}
\mathcal{T}_1(i) &= \{(a, r) : a \in \mathcal{A}_{c_i} \setminus \{a_i\},\ r \in \mathcal{R}_{c_i,a}\},\\
\mathcal{T}_2(i) &= \{(a_i, r) : r \in \mathcal{R}_{c_i,a_i} \setminus \{r_i\}\},\\
\mathcal{T}_3(i) &= \{s : 1 \le s \le S_{c_i,a_i,r_i},\ s \ne s_i\}, \qquad k_i = \min\{k : \mathcal{T}_k(i) \ne \varnothing\}.
\end{aligned} \tag{A.1}$$

If $k_i \le 2$, we draw $(a_i^+, r_i^+) \sim \mathrm{Unif}(\mathcal{T}_{k_i}(i))$ and then draw s_i^+ uniformly from that realization's retained windows. If $k_i = 3$, we set $(a_i^+, r_i^+) = (a_i, r_i)$ and draw $s_i^+ \sim \mathrm{Unif}(\mathcal{T}_3(i))$. The designated positive is $j_i^+ = (c_i, a_i^+, r_i^+, s_i^+)$. Thus a multi-archive anchor first receives a same-work positive from another archive, then another realization in the same archive, and finally another window of the same realization. In the single-archive KDF fit, $\mathcal{T}_1(i) = \varnothing$ for every anchor, so the rule reduces automatically to the latter two tiers. This hierarchy guarantees one positive for each anchor; Eq. 5 subsequently treats every same-work observation already present in the batch as positive.

B CO-REGISTERED INPUT VISUALIZATION

Figure B.1 shows the three input planes co-registered on an identical 88×1000 time–pitch grid. Activity occupies the key-down onset-to-offset interval. Velocity uses the same interval before sustain-control extension, whereas density is attached to a pitch-specific neighborhood around each note attack and is weighted by the local all-pitch attack count.

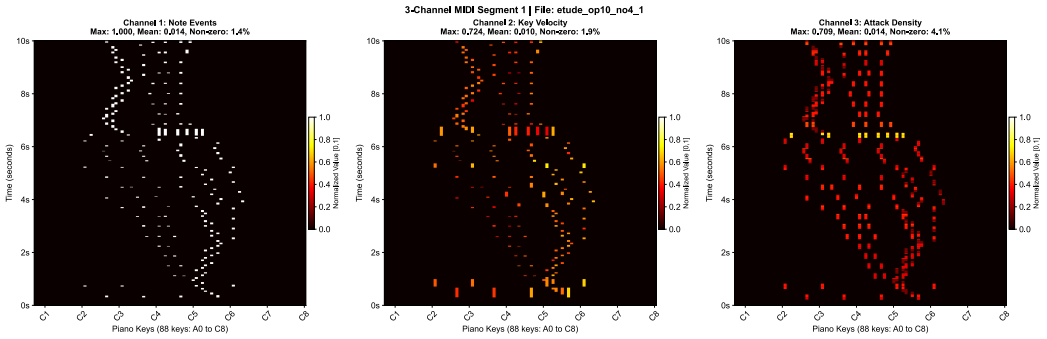


Figure B.1: Co-registered three-channel input for a 10-second segment of Chopin's *Étude* Op. 10, No. 4. Activity (left) spans each note from onset to offset. Normalized MIDI velocity (center) supplies a file-specific control trace and follows the sustain-control extension. Attack density (right) writes the local all-pitch attack count into a 0.2-s neighborhood at each attacked pitch.

C KDF CORPUS VIEWS AND EXPERIMENTAL SUBSETS

The KDF discovery view contains 234 MIDI realizations, 146 work labels, and 23,976 retained windows. Seventy-seven works have one realization, 53 have two, 13 have three, and three have four. Figure C.1 summarizes the realization and genre distributions. The six largest labels by work count are mazurka (50), nocturne (21), polonaise (18), waltz (17), *étude* (12), and ballade (4). Their window counts are 4,332, 3,925, 4,480, 2,720, 1,286, and 1,761. Work-level aggregation prevents long realizations from automatically contributing more points to clustering and chronology; segment-level analyses retain unequal label support and therefore use balanced metrics.

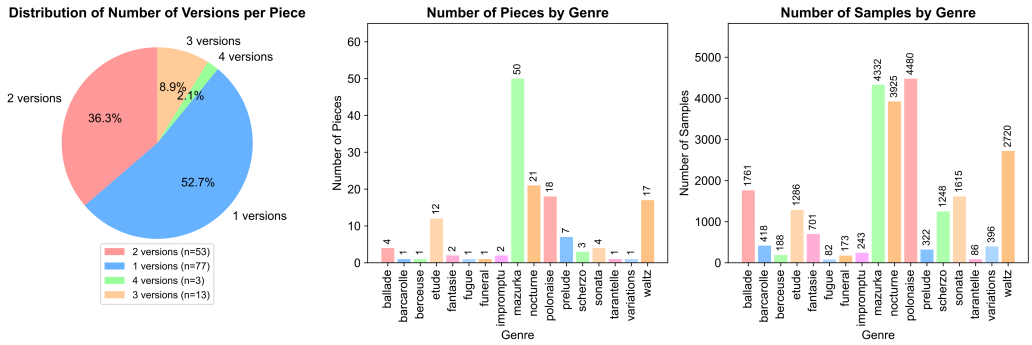


Figure C.1: Collected-corpus inventory. Left: realization counts per work. Middle: work counts across the 17 retained genre labels. Right: 10-second window counts by genre.

Table C.1: KDF views and experimental subsets used in the manuscript. Each row reports the observations retained for its stated analysis.

Experimental view	MIDI realizations	Works	Role in the study
Single-archive discovery	234	146	Fits the KDF atlas using 23,976 overlapping 10-second windows.
Held-out KDF view	232	139	Evaluates a model fitted on the other ten archives without KDF exposure.
Shared-work structural subset	211	129	Compares work geometry and evaluates segment-to-work retrieval across archives.
Realization split-half subset	146	64	Measures reliability between two nonoverlapping realization groups for each work.
Internal multiversion subset	157 queries	69	Audits alignment to another realization of the same work after removing the query file from every prototype.

The single-archive view preserves the 146 project-level labels associated with the 234 usable KDF tensors. Cross-archive evaluation normalizes these labels against the external archive inventories, yielding 232 nonempty realizations of 139 works. The structural comparison further retains 211 realizations from 129 unambiguous single-work identities shared across the two archive partitions. Split-half reliability and internal alignment use the smaller multirealization subsets listed in Table C.1. Archive sources are identified by their original publications or websites in Section D.

D CROSS-ARCHIVE STRUCTURAL REPLICATION

The single-archive atlas uses all KDF-Chopin files for discovery. For cross-archive replication, the same ViT-Base encoder is fitted on ten independently assembled performance-MIDI archives and KDF-Chopin is held out in its entirety. KDF has zero optimization exposure. The primary endpoint compares independently aggregated KDF and performance work–work geometries over work identities shared across the two partitions.

D.1 UNIFIED COLLECTION AND SOURCE ACCOUNTING

The collection contains 47,489 retained realizations from 11 source labels and 222 normalized work identities. Table D.1 reports the complete source accounting, and Figure D.1 reports every retained source–work cell without subsampling. KDF-Chopin is held out in its entirety. The other ten archives form the training partition, containing 47,256 realizations and the union of all 222 normalized work identities, and are fitted with the batch-all objective defined in Section 3.

Table D.1: Archive inventory for cross-archive replication. "Works" gives the number of normalized identities represented in each source; union totals appear in the final rows. Each retained realization is counted under one primary source. KDF is excluded from fitting.

Archive label	Role	Retained	Nonempty	Works	Median/work
PianoCoRe/Aria-MIDI	Train	39,990	39,990	209	88.0
PianoCoRe/PERiScoPe	Train	5,188	5,188	189	15.0
PianoCoRe/ATEPP	Train	1,301	1,301	139	7.0
PianoCoRe/ASAP	Train	249	249	34	4.5
MAESTRO v3	Train	197	197	56	2.0
ASAP	Train	191	191	31	4.0
PianoCoRe/GiantMIDI	Train	79	79	78	1.0
Vienna 4x22	Train	44	44	2	22.0
SMD	Train	13	13	13	1.0
PianoVAM	Train	4	4	4	1.0
Kunst der Fuge (KDF-Chopin)	Held out	233	232	139	2.0
All source labels	—	47,489	47,488	222[a]	—
Training partition	Train	47,256	47,256	222[a]	—

[a] Union of normalized identities across sources.

The source labels in Table D.1 preserve their acquisition routes. The five PianoCoRe branches come from the normalized PianoCoRe release (Borovik, 2026): Aria-MIDI (Bradshaw & Colton, 2025), PERiScoPe (Borovik et al., 2025), ATEPP (Zhang et al., 2022), ASAP (Foscarin et al., 2020), and GiantMIDI-Piano (Kong et al., 2022). The directly ingested sources are MAESTRO v3 (Hawthorne et al., 2019), ASAP (Foscarin et al., 2020), Vienna 4x22 (Goebl, 1999), Saarland Music Data (SMD) (Müller et al., 2011), PianoVAM (Kim et al., 2025), and the Kunst der Fuge Chopin archive (Kunst der Fuge, 2022).

Aria-MIDI contributes 39,990 realizations, corresponding to 84.21% of the complete registry and 84.62% of the ten-source training partition. All quality-filtered files are retained without per-work or per-source caps. Multi-archive anchors therefore follow the natural empirical frequency of retained realizations. Source priority selects the same-work positive while leaving the anchor marginal unchanged, and the positive realization is drawn uniformly from the pooled candidates in the highest available tier. Source balancing is applied during inference-time work aggregation. The external fit consequently reflects the natural source imbalance of the assembled performance archives.

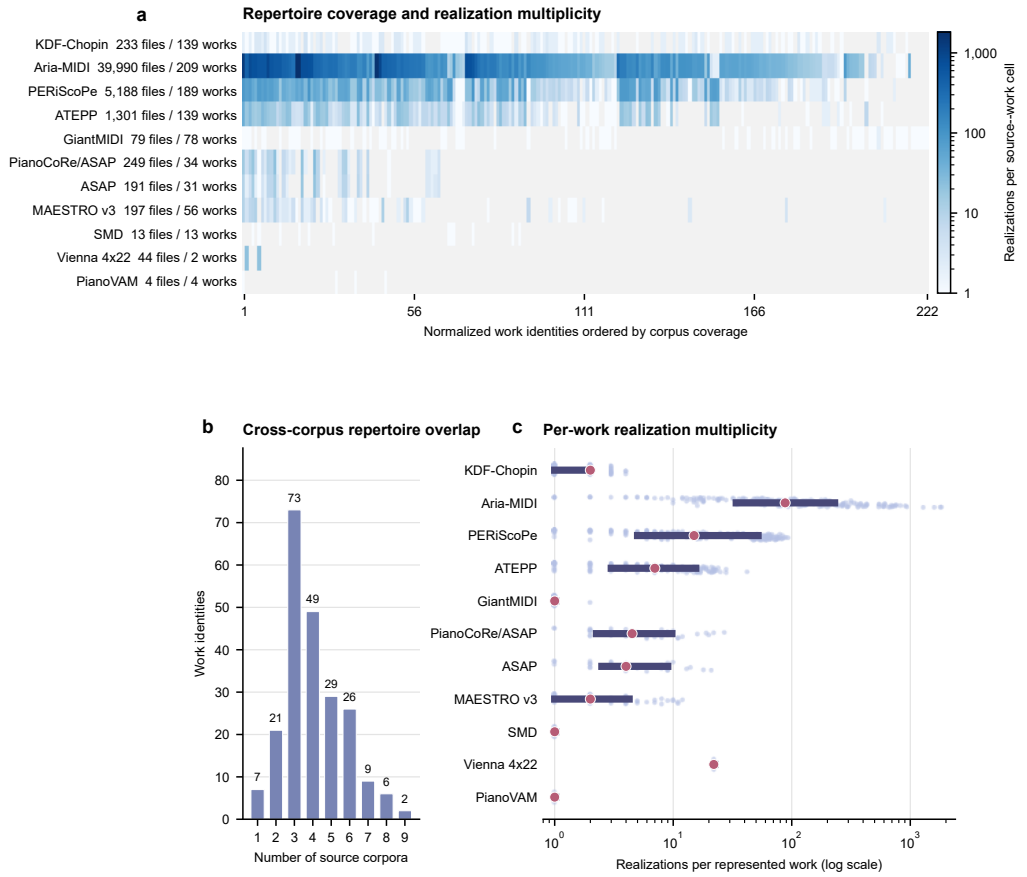


Figure D.1: Coverage of the unified performance-MIDI collection. (a) Retained realizations per source–work cell on a logarithmic color scale; gray denotes absence. (b) Number of work identities represented in exactly k of the 11 archives. The observed coverage ranges from one to nine archives, so no work identity occurs in ten or all eleven archives. (c) Complete per-work realization-count distributions for each source. All 47,489 realizations are counted once under their primary source, and no work or source is sampled for display.

D.2 COMPUTE-MATCHED OBJECTIVES AND FIXED REFERENCES

The first six rows of Table D.2 are controlled PianoFingerprints variants. They share the co-registered input, ViT-Base backbone, 256-dimensional output, fitting data, batch construction, and 8,000-update budget described in Appendix E. PianoFingerprints uses batch-all multi-positive SupCon with source-priority positive selection. Standard SupCon (Khosla et al., 2020) removes source priority. The hierarchical sampled contrastive variant uses an NCE-style estimator (Gutmann & Hyvärinen, 2010), while paired InfoNCE (van den Oord et al., 2018) changes the contrastive estimator; Composition CE (Goodfellow et al., 2016) and triplet learning (Schroff et al., 2015) replace the objective. Aria (Bradshaw et al., 2025), MusicBERT (Zeng et al., 2021), and CLaMP 2 (Wu et al., 2025) provide pretrained references, with MusicBERT also evaluated after task adaptation. The training-free descriptor follows established pitch, rhythm, dynamics, density, and texture families (Eerola & Toiviainen, 2004; McKay et al., 2018). Each trainable method contributes one completed fit; fixed representations contribute deterministic point estimates.

Table D.2 reports the available comparisons among the fitted models. For view $h \in \{\mathrm{KDF}, \mathrm{Perf}\}$, let $e_c^{(h)}$ be the independently aggregated fingerprint of common work c and $M_{cc'}^{(h)} = (e_c^{(h)})^\top e_{c'}^{(h)}$. With $\mathrm{vech}_\triangle$ denoting the strict upper triangle, cross-archive geometry is

$$r_s = \mathrm{corr}_{\mathrm{Spearman}}\left[\mathrm{vech}_\triangle(M^{(\mathrm{KDF})}), \mathrm{vech}_\triangle(M^{(\mathrm{Perf})})\right], \tag{D.1}$$

computed over the same 129 conservative works. Work labels establish row correspondence, and the embeddings determine the $\binom{129}{2} = 8{,}256$ pairwise similarities. Retrieval ranks performance-derived work prototypes for KDF realization queries. If $\mathcal{W}_{\mathrm{eval}}$ is the evaluated work set, $\mathcal{Q}_c$ is the query set for work c, and $\mathrm{rank}(q)$ is the one-indexed rank of its correct work, the reported work-macro recall is

$$\mathrm{R@}\,k = \frac{1}{|\mathcal{W}_{\mathrm{eval}}|} \sum_{c \in \mathcal{W}_{\mathrm{eval}}} \frac{1}{|\mathcal{Q}_c|} \sum_{q \in \mathcal{Q}_c} \mathbf{1}[\mathrm{rank}(q) \le k]\,. \tag{D.2}$$

Split-half reliability partitions KDF realizations within each eligible work and compares two independently aggregated halves by the same matrix correlation and macro retrieval definitions. Trainable entries are single-fit point estimates, whereas the other fixed representations are deterministic extractions.

Table D.2: Complete-KDF holdout structural replication and realization split-half results. Geometry is Spearman r_s; retrieval entries are work-macro recall. The first six rows are controlled PianoFingerprints variants. MusicBERT is reported before and after task adaptation. For each metric, the largest distinct point estimate is bold and the second largest is underlined.

Method	Cross-archive r_s	External R@1	External R@3	External R@10	Split-half r_s	Split-half R@1
Controlled PianoFingerprints variants						
PianoFingerprints	0.767	0.824	0.860	0.880	**0.870**	**0.917**
Standard SupCon	0.793	<u>0.846</u>	<u>0.870</u>	<u>0.891</u>	<u>0.841</u>	0.883
Hierarchical sampled contrastive	0.780	0.577	0.772	0.873	0.802	0.707
Paired InfoNCE	**0.811**	0.559	0.757	**0.895**	0.796	0.657
Composition CE	0.803	0.264	0.441	0.683	0.751	0.325
Triplet	<u>0.806</u>	0.456	0.718	0.859	0.815	0.595
Modern pretrained symbolic references						
Aria	0.745	0.377	0.591	0.753	0.530	0.094
MusicBERT	0.094	0.168	0.228	0.335	0.067	0.090
MusicBERT-adapted	0.519	0.328	0.443	0.563	0.403	0.177
CLaMP 2	0.623	0.674	0.731	0.839	0.481	0.168
Training-free symbolic reference						
Symbolic descriptors	0.671	**0.880**	**0.884**	0.884	0.673	<u>0.906</u>

The first column shows that a substantial portion of Chopin work geometry recurs across independently collected archives. Several controlled objectives have larger cross-archive correlations than PianoFingerprints, whereas PianoFingerprints is the strongest learned row on both split-half statistics. The fixed descriptor exceeds PianoFingerprints on external R@1. The results distinguish cross-archive structural replication and realization reliability from closed-set identity retrieval.

D.3 LOCAL-PASSAGE SUPPORT FOR CROSS-GENRE NEIGHBOURS

For each retained method from Table D.2, we independently select work pairs that are reciprocal top-five neighbours in both the held-out KDF geometry and the performance-archive geometry and that have different genre labels. Every eligible KDF realization contributes four uniformly spaced ten-second windows. A fixed 166-dimensional descriptor, assembled from established symbolic feature families (Eerola & Toiviainen, 2004; McKay et al., 2018), summarizes pitch class, register, temporal activity, attack density, velocity, and global statistics for each window. For a relation (i, j), its local score is

$$s_{ij} = \max_{u \in \mathcal{W}_i,\, v \in \mathcal{W}_j} \cos(d_u, d_v), \tag{D.3}$$

where $\mathcal{W}_i$ and $\mathcal{W}_j$ contain the retained windows of the two works and d is the standardized, unit-normalized descriptor. Thus all stored cross-work window combinations are compared, while each work relation contributes one best-matching-passage score. For method m with relation set E_m, the graph-level statistic is $T_m = |E_m|^{-1} \sum_{(i,j) \in E_m} s_{ij}$.

Each of 9,999 permutations independently shuffles work identities within genre. The network position and both endpoint genres remain fixed, but the specific work at each endpoint is replaced. Recomputing the relation scores and their mean produces a genre-preserving reference distribution for every method. The one-sided Monte Carlo value uses plus-one correction,

$$p_{\mathrm{perm},m} = \frac{1 + \sum_{b=1}^{9,999} \mathbf{1}(T_{b,m} \geq T_m)}{10,000}. \tag{D.4}$$

Table D.3 compares the six controlled PianoFingerprints variants with the Aria and MusicBERT references. PianoFingerprints identifies the largest cross-genre relation set (66 relations across 75 works) and the highest local mean (0.548). Its permutation mean is 0.403 ($\Delta = 0.145$, $p_{\mathrm{perm}} = 1.0 \times 10^{-4}$). Thus, its cross-archive neighbours retain unusually similar local symbolic passages. Figure D.2 resolves this result relation by relation and lists the 20 highest-scoring candidate pairs.

Table D.3: Local-passage support for reciprocal cross-genre neighbours selected independently by the controlled PianoFingerprints variants and two pretrained references. "Local mean" averages the best cross-work passage cosine over the selected relations. The permutation distribution shuffles work identities within genre while retaining graph positions and both endpoint genres. Reported p_{perm} values are one-sided Monte Carlo values with plus-one correction. The best value is bold and the second-best value is underlined for relation count, endpoint coverage, local mean, and enrichment.

Method	Relations	Works	Local mean	Permutation mean [95% interval]	Δ	p_{perm}
Controlled PianoFingerprints variants						
PianoFingerprints	**66**	<u>75</u>	**0.548**	0.403 [0.354, 0.452]	**0.145**	1.0×10^{-4}
Standard SupCon	<u>58</u>	**81**	<u>0.542</u>	0.411 [0.361, 0.460]	<u>0.130</u>	1.0×10^{-4}
Hierarchical sampled contrastive	49	59	0.520	0.396 [0.336, 0.456]	0.123	1.0×10^{-4}
Paired InfoNCE	50	69	0.505	0.407 [0.353, 0.461]	0.098	4.0×10^{-4}
Composition CE	46	58	0.489	0.399 [0.336, 0.460]	0.091	2.3×10^{-3}
Triplet	45	64	0.491	0.411 [0.355, 0.467]	0.080	2.4×10^{-3}
Modern pretrained symbolic references						
Aria	8	15	0.537	0.414 [0.276, 0.552]	0.124	0.039
MusicBERT	6	12	0.479	0.375 [0.199, 0.557]	0.104	0.128

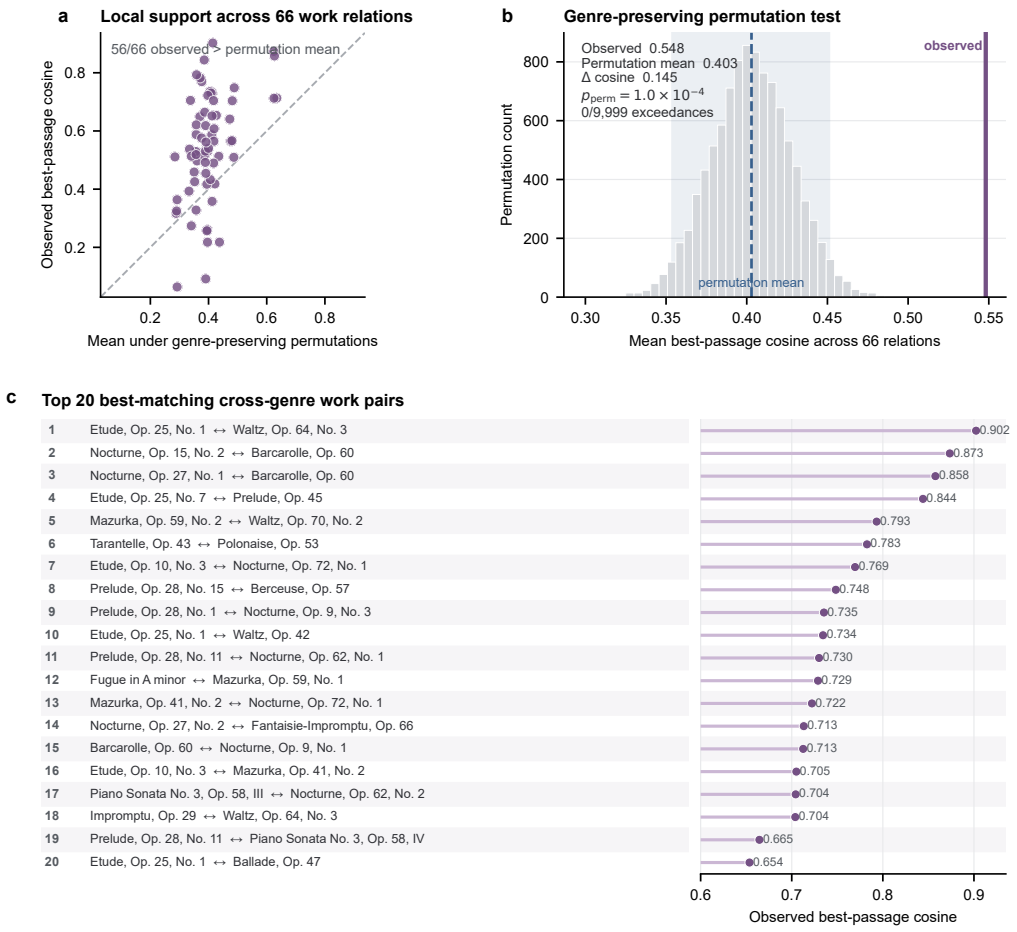


Figure D.2: Local symbolic support for PianoFingerprints cross-archive cross-genre neighbours. Each point or statistic is based on the 66 undirected reciprocal cross-genre relations shared by the performance-archive and KDF atlas views, involving 75 endpoint works. Each eligible KDF realization contributes four uniformly spaced ten-second windows, and every relation is scored by the largest fixed-descriptor cosine across all cross-work window pairs. (a) Observed best-passage score for each relation against its mean across 9,999 permutations that reassign work identities within genre. The dashed diagonal denotes equality; 56 of 66 relations lie above it. (b) Permutation distribution of the graph-level mean score across all 66 relations. The purple line marks the observed mean 0.548, the dashed blue line marks the permutation mean 0.403, and shading marks the 95% interval $[0.354, 0.451]$. No permutation equals or exceeds the observation ($0/9{,}999$; $p_{\mathrm{perm}} = 1.0 \times 10^{-4}$ after plus-one correction). (c) The 20 relations with the largest observed best-passage cosine, ranked and labelled with concise catalogue-style work names as candidates for expert inspection. The descriptor is computed after neighbour selection and is separate from PianoFingerprints.

D.4 STRUCTURAL RELIABILITY AND RETRIEVAL CALIBRATION

The cross-archive matrix endpoint is complemented by realization reliability within held-out KDF. PianoFingerprints obtains split-half work-geometry $r_s = 0.870$, chance-adjusted neighbour overlap@5 0.582, and same-work impostor margin 0.064. Standard SupCon gives 0.841, 0.577, and 0.063, respectively. PianoFingerprints is higher on all three reported reliability measures in these fitted models.

Segment-to-work retrieval under four-window and dense queries. A query contains one 10-second KDF segment and targets its normalized work identity. The four-window audit uses four uniformly spaced windows per eligible MIDI (844 queries from 211 files and 129 works). The dense audit uses every nonempty, non-overlapping window from the same files (5,328 queries), averaging windows within each MIDI before taking the work macro-average so that longer realizations receive equal file weight. Both protocols use the same external performance-MIDI gallery, with four uniformly spaced windows per realization and 196 candidate work identities. We compare the final source-priority multi-positive PianoFingerprints objective with sampling/loss controls, the symbolic descriptor, Aria, CLaMP 2, and MusicBERT before and after task adaptation (Bradshaw et al., 2025; Zeng et al., 2021; Wu et al., 2025).

Table D.4: Ten-second segment-to-work retrieval for all methods. The four-window protocol reports work-macro metrics over 844 queries; the dense protocol reports file-balanced work-macro metrics over all 5,328 eligible non-overlapping queries. The query files (211), query works (129), gallery, and 196 candidate works are identical. Aria and CLaMP 2 are included for comparison, and MusicBERT is shown before and after task adaptation. The best value in each column is bold and the second-best value is underlined.

	Four uniformly spaced windows ($n = 844$)			All non-overlapping windows ($n = 5{,}328$)		
Method	R@1	R@5	MRR	R@1	R@5	MRR
Proposed and controlled ViT objectives / sampling						
PianoFingerprints	**0.571**	**0.774**	**0.660**	<u>0.304</u>	<u>0.487</u>	<u>0.398</u>
Standard SupCon	<u>0.549</u>	<u>0.760</u>	<u>0.648</u>	0.256	0.455	0.357
Hierarchical sampled contrastive	0.363	0.626	0.485	0.153	0.324	0.248
Paired InfoNCE	0.362	0.624	0.485	0.147	0.321	0.243
Composition CE	0.126	0.316	0.231	0.059	0.178	0.133
Triplet	0.240	0.494	0.363	0.113	0.277	0.203
Training-free symbolic reference						
Symbolic descriptors	0.077	0.223	0.162	0.041	0.155	0.114
Modern pretrained symbolic encoders						
Aria	0.214	0.482	0.347	0.190	0.441	0.313
MusicBERT	0.071	0.229	0.162	0.039	0.158	0.111
MusicBERT-adapted	0.199	0.472	0.328	0.125	0.311	0.227
CLaMP 2	0.525	0.712	0.614	**0.319**	**0.503**	**0.412**
Uniform random ranking	0.005	0.026	0.030	0.005	0.026	0.030

PianoFingerprints is the strongest from-scratch ViT objective on all six segment-retrieval metrics and the strongest reported method for the four-window metrics. CLaMP 2 is the strongest dense-query reference. Its large-scale symbolic pretraining and lossless MIDI-to-MTF event serialization are plausible contributors. The published CLaMP 2 and Aria data descriptions leave the overlap status of individual Chopin realizations unresolved, so Table D.4 reports them as off-the-shelf pretrained references. MusicBERT-adapted measures transfer followed by task adaptation.

D.5 MODERN SYMBOLIC REFERENCES AND ATLAS EXPANSION

Table D.2 includes modern pretrained references under the same KDF holdout. Aria and CLaMP 2 use their released representations, and MusicBERT is shown before and after task adaptation. The unified registry adds 83 normalized identities beyond the 139 KDF identities. After removing 16 complete-set or ambiguous labels, 67 additional single-work identities remain. Expansion therefore adds new works and independently sourced realizations, while the KDF holdout measures recurrence of the original archive's work structure in a performance-trained atlas.

E ARCHITECTURE AND FITTING DETAILS

Table E.1 summarizes the architecture, sampling hierarchy, and fitting configuration used by the final models.

Table E.1: Final architecture and compute-matched fitting configuration.

Stage	Input or setting	Output or value
Input tensor	10-s MIDI window	$B \times 3 \times 88 \times 1000$
Patch embedding	11×50 patches	$B \times 160 \times 768$
Transformer	12 blocks, 12 heads	160 patch tokens, width 768
Pool/head	Global average + head	$B \times 256$, L_2 norm.
Positive sampling	Source-priority hierarchy	Source, realization, window
Positive set	All same-work batch items	Multi-positive SupCon
Negative set	All different-work batch items	No sampled denominator
Batch/loss	32 anchors, 64 observations	$\tau = 0.07$
Optimizer	AdamW, lr 3×10^{-4}	Weight decay 0.05
Schedule	Cosine, min. 10^{-6}	8,000 optimizer updates
Precision/reporting	bfloat16	One fit per trainable method
Stabilization	Gradient clipping	Maximum norm 1.0

Each optimizer batch starts from 32 anchor–designated-positive pairs, yielding 64 observations. The designated-positive hierarchy controls which relations are likely to enter the batch, whereas the loss uses every same-work observation already present as a positive and every different-work observation as a negative. This batch-all definition applies to the single-archive discovery atlas and the multi-archive replication atlas. In KDF alone, the unavailable source tier is skipped; work-distinct anchor batches visit every one of the 234 stored files once per nominal epoch, and shuffled no-replacement decks within files expose all 23,976 windows over training. In the multi-archive regime, anchors are the 47,256 retained training realizations shuffled at their natural file frequencies; one of four uniformly spaced windows is chosen per anchor visit, and cross-archive same-work positives are preferred.

The final runs retain the full ViT-Base scale because the co-registered tensor admits a regular time–pitch patching scheme. The 11×50 patch covers 11 adjacent piano keys and 0.5 seconds, producing 160 tokens per segment. This resolution balances local polyphonic detail against the quadratic token cost of global self-attention and matches the single-archive atlas configuration.

E.1 Recorded single-archive training dynamics

Figure E.1 documents optimization of the original single-archive fitted atlas. It reports training-set accuracy, loss, and pair similarities recorded throughout fitting. The final model follows the fixed training schedule reported above.

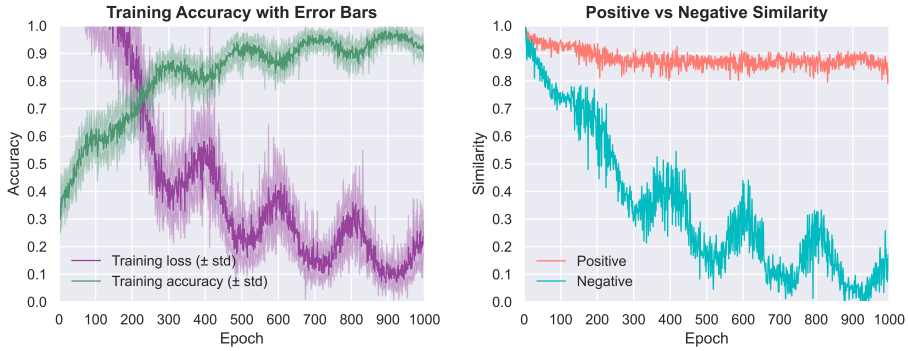


Figure E.1: Recorded single-archive training dynamics over 1000 nominal epochs. Left: training accuracy and loss trajectories. Right: positive- and negative-pair similarities.

F SINGLE-ARCHIVE CHANNEL SENSITIVITY AND INTERNAL RETRIEVAL

F.1 CHANNEL SENSITIVITY OF THE FITTED ATLAS

The original channel-masking analysis is recomputed with the reported single-archive ViT-Base encoder. One co-registered plane is zeroed at inference time and genre-proxy performance is recomputed over the same sampled segments. The encoder remains fixed, and the three segment-level folds contain overlapping windows. The resulting changes measure model sensitivity to activity, velocity, and attack density.

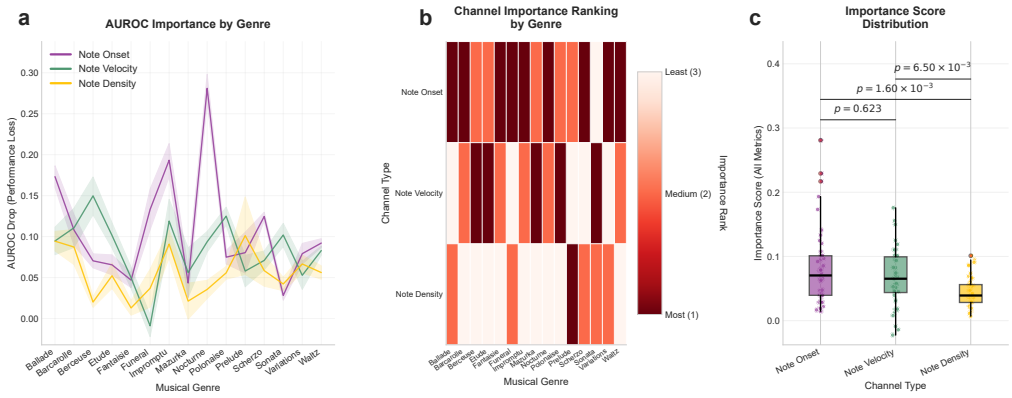


Figure F.1: Channel sensitivity of the final single-archive model. (a) Per-label AUROC decrease after masking activity, velocity, or attack density. (b) Within-label channel rank. (c) Importance averaged over balanced accuracy, F1, and AUROC under segment-level three-fold resampling. The encoder remains fixed during inference-time masking.

F.2 CHANNEL-INPUT COMPARISON UNDER HIERARCHICAL SAMPLED CONTRASTIVE LEARNING

This auxiliary KDF experiment compares four separately fitted ViT-Base input configurations: activity–velocity–attack density, activity–velocity, activity–attack density, and velocity–attack density. The encoder architecture, preprocessing, hierarchical sampled contrastive objective, initialization protocol, and update budget are held fixed across the four configurations. The final PianoFingerprints model instead uses the batch-all multi-positive SupCon objective defined in Section 3 and Appendix E.

Figure F.2 reports the four configurations. The three-channel model reaches final training nearest-positive accuracy 0.609 and internal Recall@1 0.433. The activity–velocity, activity–attack-density, and velocity–attack-density models reach Recall@1 0.350, 0.242, and 0.191, respectively. This comparison isolates channel composition under a common training objective, complementing the inference-time masking analysis of the final PianoFingerprints model.

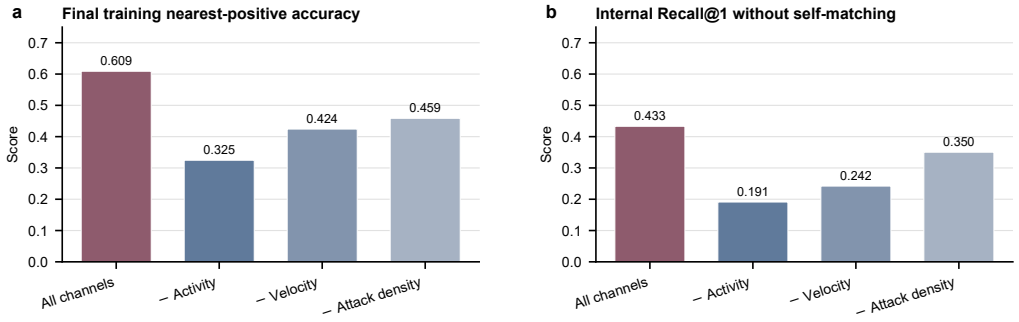


Figure F.2: Channel-input comparison under hierarchical sampled contrastive learning. Four ViT-Base models are fitted separately using activity–velocity–attack density, activity–velocity, activity–attack density, or velocity–attack density. (a) Final training nearest-positive accuracy. (b) Internal full-realization Recall@1 without self-matching. The four configurations share the same initialization protocol, objective, preprocessing, and update budget.

F.3 FULL-REALIZATION RETRIEVAL WITHOUT SELF-MATCHING

The internal audit uses the 69 KDF works with at least two MIDI realizations. Each of the 157 eligible realizations serves once as a query. Before ranking, that exact MIDI file is removed from the work prototypes, while the other realization or realizations of its work remain in the gallery. The query therefore cannot retrieve itself and must align with a fingerprint constructed from another realization of the same work. Every KDF file contributes to encoder fitting, so the statistics measure identity alignment within the fitted single-archive atlas. Each objective is represented by one fitted ViT-Base model. Table F.1 gives the numerical results, and Figure F.3 shows the metrics and cumulative rank profiles. Cross-archive evaluation uses the separate KDF-held-out fits reported in Appendix D.

Table F.1: Internal full-realization alignment without self-matching for one fitted ViT-Base model per objective. For each query, its MIDI file is removed before work prototypes are constructed; the correct prototype therefore contains only other realizations of that work. The best value in each column is bold and the second-best value is underlined.

Method	R@1	R@3	R@5	R@10	MRR
PianoFingerprints	**0.790**	**0.962**	**0.987**	**1.000**	**0.873**
Hierarchical sampled contrastive	0.433	0.694	0.841	0.943	0.603
Paired InfoNCE	<u>0.631</u>	<u>0.879</u>	<u>0.943</u>	<u>0.981</u>	<u>0.760</u>
Composition CE	0.153	0.274	0.363	0.503	0.268
Triplet	0.153	0.268	0.338	0.484	0.264

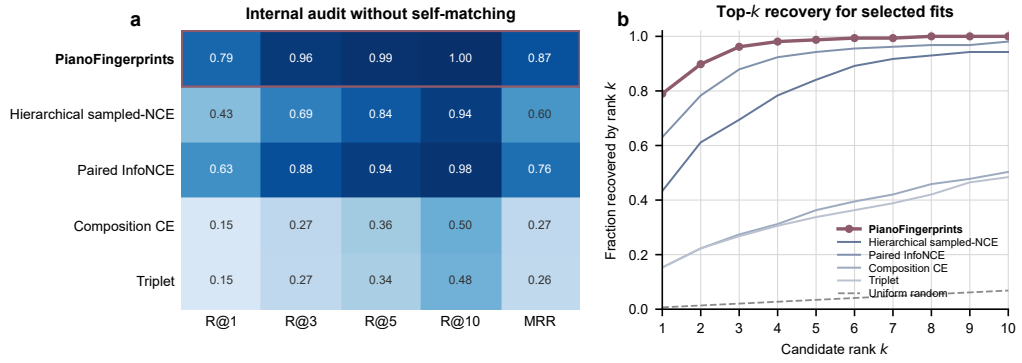


Figure F.3: Full-realization single-archive alignment without self-matching. (a) Recall@1, Recall@3, Recall@5, Recall@10, and MRR for one fitted model per objective. (b) Fraction of the 157 query MIDI files recovered through ranks 1–10; the dashed line is uniform random ranking among 146 candidate works. Each query file is removed before the work prototypes are constructed, so the correct prototype contains only other realizations of the same work.

G SINGLE-ARCHIVE CONSTRUCTION CHECKS AND PROJECTIONS

The following pairwise analysis is retained from the original single-archive PianoFingerprints atlas. It contains 110 same-work realization pairs and $\binom{146}{2} = 10{,}585$ different-work pairs. Table G.1 reports the corresponding similarity summaries. The observations are dependent because works recur across pairs; the distributions and standardized separations are therefore more informative than a nominal independent-pair test.

Table G.1: Same- and different-work similarity summaries (mean $\pm$ standard deviation).

Formulation	Same work	Different work
MMD-derived	0.9568 ± 0.0283	0.3050 ± 0.1014
Wasserstein-derived	0.9908 ± 0.0027	0.9407 ± 0.0130
Jensen–Shannon-derived	0.9607 ± 0.0230	0.5459 ± 0.0943
Kullback–Leibler-derived	0.3192 ± 0.0971	0.0145 ± 0.0597
Cosine	0.9911 ± 0.0060	0.5361 ± 0.1675
Chamfer-derived	0.8016 ± 0.0211	0.4821 ± 0.0732

The six formulations operate on different numerical ranges, so their raw mean gaps are scale dependent. Across these distributional summaries, the fitted representation places multiple realizations of a work closer than realizations from different works. Works recur across pairs because the analysis exhaustively compares the fitted corpus.

G.1 COMPLETE PAIRWISE DISTRIBUTIONS

Figure G.1 retains the spread and outliers that are suppressed by the numerical means. Wasserstein-derived similarities are especially compressed near one, while KL-derived similarities occupy a much lower range; cosine and MMD-derived variants yield the most visually direct same/different separation in this fitted corpus. Figure G.2 complements these distributions with local projections for two small-work experiments.

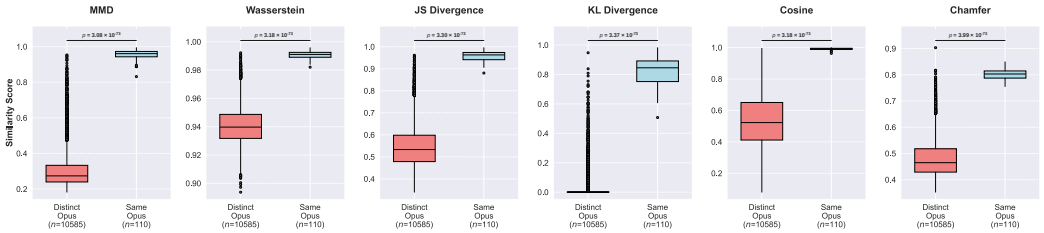


Figure G.1: Full pairwise distributions for the six similarity formulations. Each score is oriented so that larger values indicate greater similarity. Displayed pair-level p values are nominal because works recur across the plotted pairs.

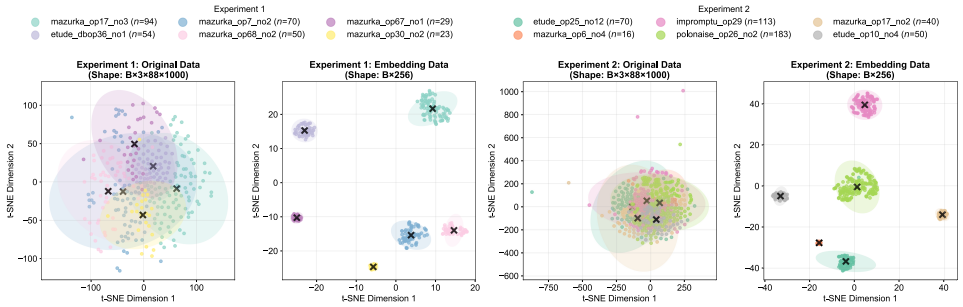


Figure G.2: Local t-SNE projections for two small-work experiments. Original co-registered features and learned segment embeddings are shown side by side. Crosses mark projected centers and translucent ellipses show displayed two-dimensional ranges. Reported geometry and neighbour metrics are computed in the original embedding space.

H SINGLE-ARCHIVE GENRE AND CHRONOLOGY ANALYSES

Work-level clustering was examined for $k = 2, \ldots, 20$. Figure H.1 summarizes the search curves, projections, and assignments. The largest K-means silhouette in this range occurs at the upper search boundary. K-means gives silhouette/ARI/NMI of $0.204/0.112/0.455$, and agglomerative clustering gives $0.186/0.106/0.448$.

The endpoint maximum shows that compactness continues to increase as the searched partitions become finer, leaving the natural cluster count unresolved. The external-label scores are modest and indicate organization both within and across conventional genre names.

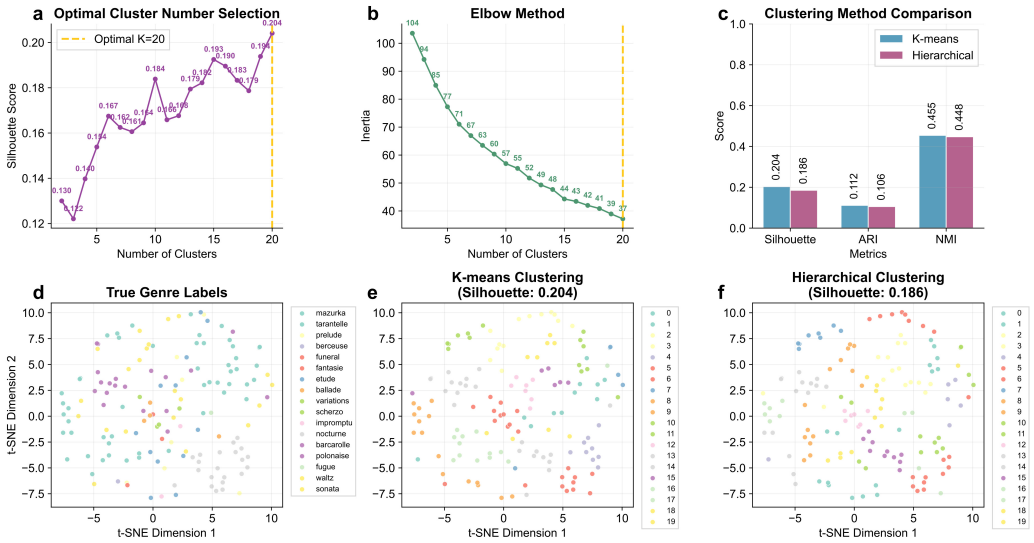


Figure H.1: Work-level clustering over 146 embeddings. Panels show silhouette and inertia over the searched range, descriptive indices, a genre-colored t-SNE view, and the two cluster assignments.

H.1 WITHIN-LABEL SUMMARIES AND CHRONOLOGICAL OPERATORS

The fitted-atlas genre-relation summary is shown in main Figure 3; Table H.1 provides its per-label numerical detail. These pairwise summaries are descriptive because works recur across many pairs and labels have unequal support.

Table H.1: Within-label cosine-similarity summaries. Labels with very few work pairs are descriptive examples only.

Label	Mean	SD	Median	Min.	Max.	Pairs
Ballade	0.696	0.131	0.695	0.534	0.874	6
Étude	0.144	0.321	0.116	−0.533	0.801	66
Fantasia	0.434	0.000	0.434	0.434	0.434	1
Impromptu	0.120	0.000	0.120	0.120	0.120	1
Mazurka	0.105	0.374	0.105	−0.840	0.975	1,225
Nocturne	0.605	0.212	0.608	0.152	0.890	210
Polonaise	0.425	0.317	0.448	−0.418	0.983	153
Prelude	0.171	0.337	0.084	−0.405	0.714	21
Scherzo	0.724	0.180	0.625	0.569	0.977	3
Sonata	0.743	0.118	0.778	0.513	0.805	6
Waltz	0.139	0.335	0.097	−0.478	0.920	136

Let $\mathcal{W}$ be the N dated works, g_i the conventional genre label of work i, $\mathcal{E} = \{j \in \mathcal{W} : 1817 \leq y_j \leq 1825\}$ the early-period subset, and $\mathcal{G}_i = \{j \in \mathcal{W} : g_j = g_i\}$ its same-genre set. For unit work fingerprints e_i, define the unit early centroid $c_{\mathcal{E}} = \mathrm{nrm}(|\mathcal{E}|^{-1} \sum_{j \in \mathcal{E}} e_j)$. The retained chronological coordinates are

$$E_i = 1 - e_i^\top c_{\mathcal{E}}, \qquad \text{early-centroid distance,} \tag{H.1}$$

$$D_i = \frac{1}{N-1} \sum_{j \in \mathcal{W} \setminus \{i\}} \left(1 - e_i^\top e_j\right), \qquad \text{corpus distinctiveness,} \tag{H.2}$$

$$C_i = \frac{1}{|\mathcal{G}_i| - 1} \sum_{j \in \mathcal{G}_i \setminus \{i\}} e_i^\top e_j, \qquad \text{within-genre consistency.} \tag{H.3}$$

The consistency coordinate is evaluated when $|\mathcal{G}_i| \geq 2$. All three coordinates are defined relative to the fitted corpus and its catalogue metadata.

Main Figure 4 displays these operators. There is no evidence of a statistically significant global linear association between distinctiveness and composition year ($r = -0.148$, nominal $p = 0.075$, $n = 146$).

I SINGLE-ARCHIVE ATTENTION DIAGNOSTICS

The attention views locate response concentrations on the co-registered input support. Figure I.1 shows representative segment-level maps. Alignment with themes, meter, and phrase boundaries can be evaluated when corresponding score annotations are available.

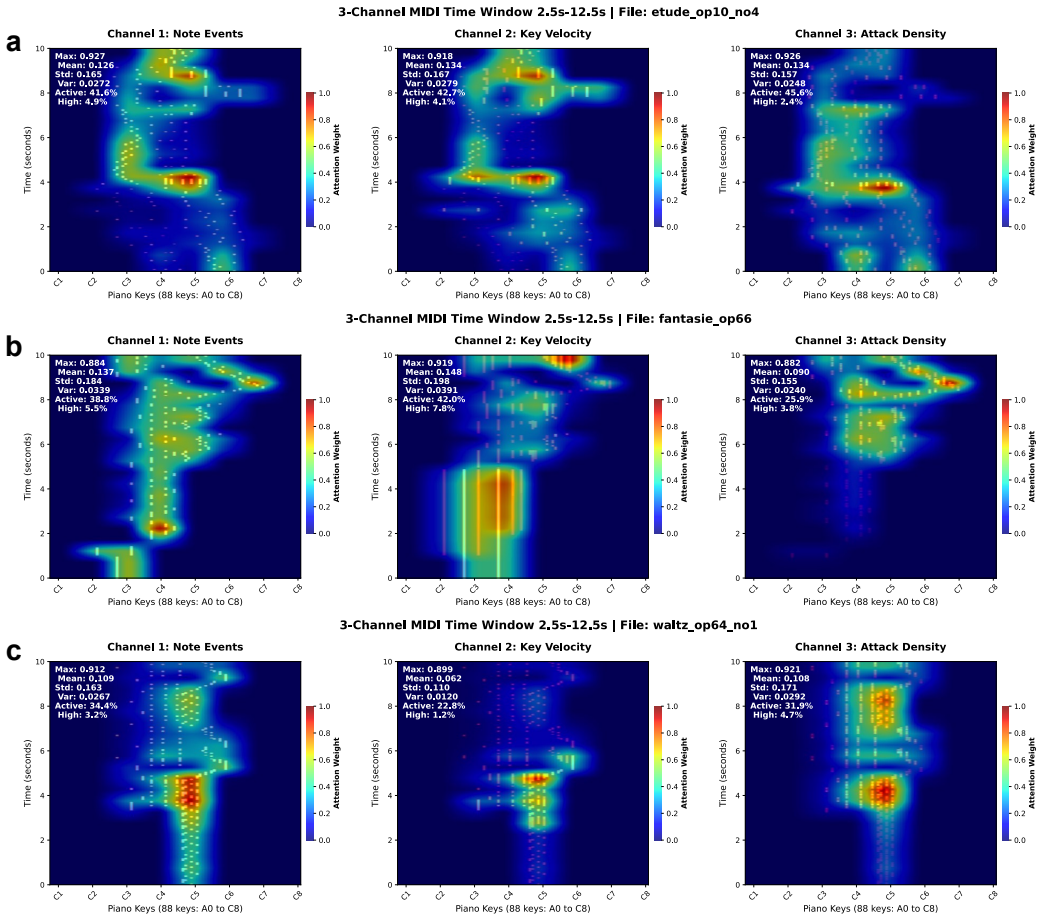


Figure I.1: Qualitative segment-level attention for three displayed works. Each row shows activity, velocity, and density responses on a common 10-second support.

I.1 EXPANDED CROSS-REALIZATION VIEW

Figure I.2 preserves the original full-piece display at a scale large enough to inspect individual realization rows. Each channel map concatenates responses produced for 10-second encoder windows. Repeated horizontal or vertical bands nominate candidate passages for subsequent alignment and can reflect recurring texture, registration, or preprocessing patterns.

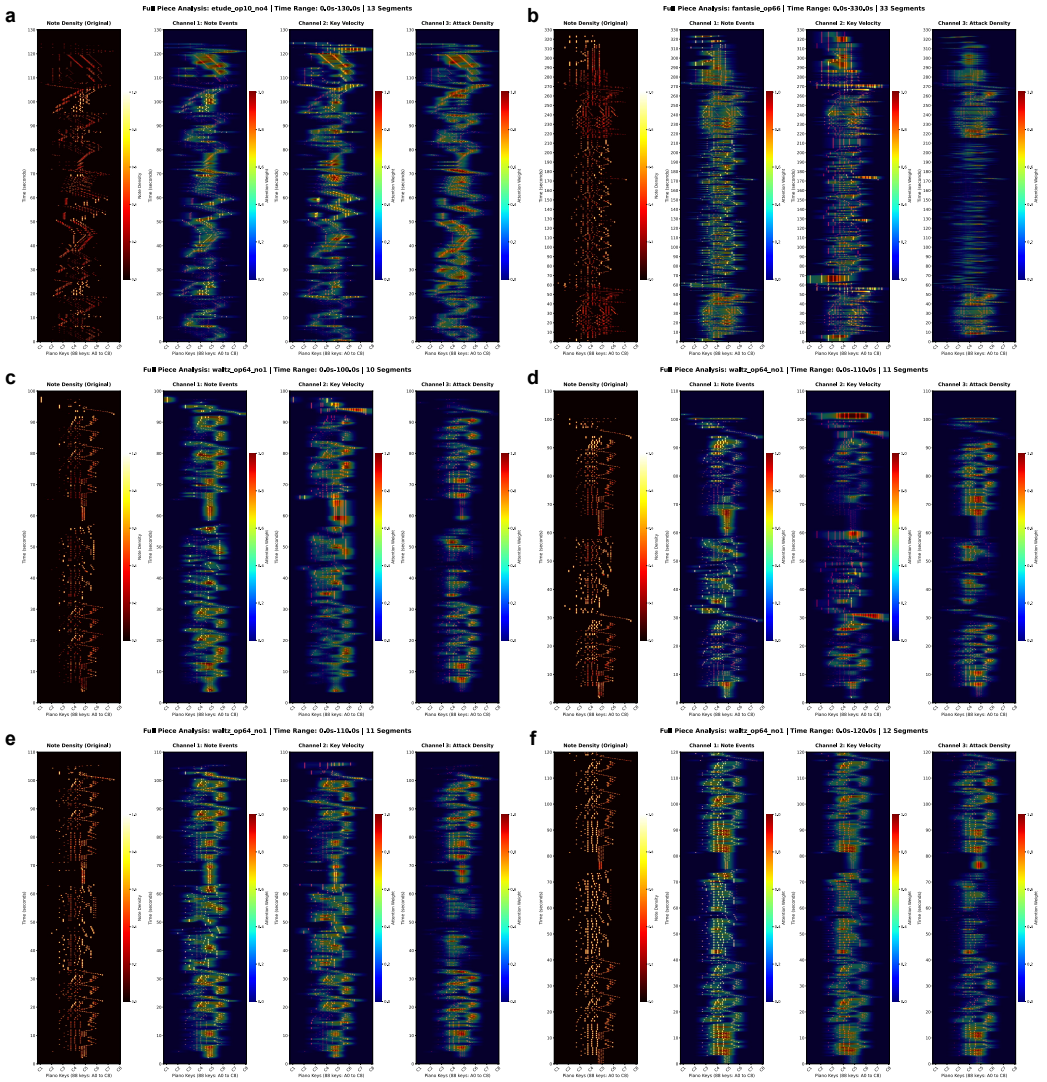


Figure I.2: Expanded full-piece response display. The rows include one realization each of Etude Op. 10, No. 4 and Fantaisie-Impromptu Op. 66, followed by four distinct MIDI realizations of Waltz Op. 64, No. 1. Recurring and differing bands nominate passages for cross-realization alignment.