LONGITUDINAL ANALYSIS OF EXPRESSIVE STABILIZATION IN REPEATED PIANO PERFORMANCE

Sein Lee[1] **You Jin Kim**[2] **Moo Kyoung Song**[3] **Juhan Nam**[1]
[1] Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
[2] Kyung Hee University, Seoul, South Korea
[3] Yonsei University, Seoul, South Korea
{seinlee,juhan.nam}@kaist.ac.kr, yjinkim@khu.ac.kr, msong999@yonsei.ac.kr

ABSTRACT

This study examines the process of interpretive change and stabilization in musical performance using music information retrieval (MIR) techniques. Whereas much existing performance analysis has focused on cross-sectional comparisons of stylistic differences among finalized performance products, this study longitudinally investigates how a single performer's interpretation is formed and stabilized through repeated performances. Using the third movement of Beethoven's Hammerklavier Sonata Op. 106 as a case study, six sessions by a professional pianist were recorded over three months. We apply a multi-faceted analysis, including tempo-dynamics curves, performance worms, and clustering-based performance alphabet analysis, to track how expressive strategies shift across sessions. The results suggest that interpretive stabilization is a non-linear process, with distinct patterns of exploration and convergence depending on the formal structure of the piece. Rather than treating performance as a fixed object, this study shows that repeated-performance data can make the formation and stabilization of interpretation analytically observable.

1. INTRODUCTION

The academic endeavor to understand musical performance has a long history. In the early 20th century, performance analysis was largely situated within the qualitative realms of score interpretation and performance practice. However, with the rise of Music Information Retrieval (MIR), the methodology shifted toward directly measuring and quantifying the acoustic reality of the performance itself [1]. Computational approaches to expressive performance have since matured into a well-established research area [2, 3]. Recent empirical work has surveyed this trajectory with a focus on timing, dynamics, and gesture [4]. Yet, most existing MIR-based performance research has focused heavily on reference performances—analyzing finalized recordings to compare stylistic differences among multiple artists [5–7]. This paradigm treats musical interpretation as a static product rather than an evolving process. Consequently, a critical gap remains in understanding how a performer actually constructs and stabilizes their interpretive strategy through repeated engagement with a complex work. Longitudinal MIR analysis can operationalize this process by comparing corresponding score positions across repetitions, revealing which aspects of timing, dynamics, and their coordination stabilize early and which remain exploratory.

© 2026 Korean Society for Music Informatics. **Citation:** Sein Lee, You Jin Kim, Moo Kyoung Song, and Juhan Nam, "Longitudinal Analysis of Expressive Stabilization in Repeated Piano Performance", *The First Korean Society for Music Informatics Conference*, Seoul, Republic of Korea, 2026.

Figure 1. Recording setup: Yamaha Disklavier with stereo microphones at fixed positions.

Beethoven's Hammerklavier Sonata Op. 106, specifically its third movement (*Adagio sostenuto* in F♯ minor), provides a profoundly compelling subject for investigating this process. As one of the most massive slow movements in the piano repertoire, its performance duration typically spans from 18 to 25 minutes. The performer must navigate a vast interpretive search space, balancing extreme internal stillness with high tonal tension. Op. 106/III is thus an ideal stress test for this framework: its extreme duration and sparse rhythmic pulse maximize the interpretive search space, making the process of stabilization both necessary and analytically observable.

In this paper, we propose a systematic framework to analyze six repeated performances of Op. 106/III by a professional pianist, following the multi-resolution methodology of Widmer et al. [8]. We apply tempo-dynamics curves, performance worms [9, 10], and SOM-based clustering [11] to track interpretive evolution across sessions.

2. DATA COLLECTION

2.1 Recording Setup

Performance data were collected using a Yamaha Disklavier, capturing key and pedal movements in real time [12]. The instrument records onset time, offset time, and velocity for each note in MIDI format, documenting

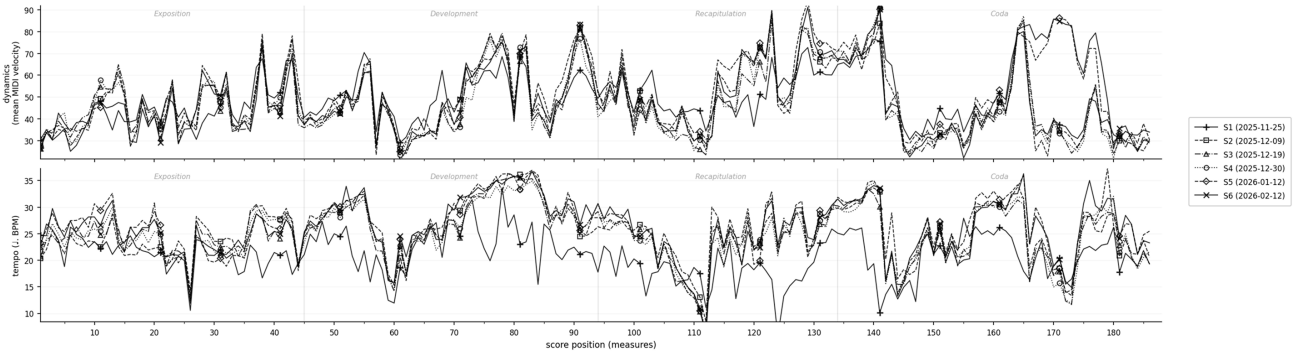

Figure 2. Dynamics curves (top) and tempo curves (bottom) across six sessions, superimposed on one coordinate system (x-axis: measure number). Vertical lines mark formal section boundaries.

expressive information without post-processing artifacts. Two stereo microphones were mounted at fixed positions, maintaining identical acoustic conditions across all sessions. All recordings took place in a dedicated recording studio at Yonsei University.

2.2 Sessions and Participants

Data collection comprised six sessions over approximately three months. In each session, the pianist performed the entire third movement. The performer agreed not to practice the movement separately between sessions, ensuring that each performance accumulated only the experience of the immediately preceding one. The participating pianist is a renowned Beethoven specialist, currently serving as a faculty member at a Korean music conservatory. It should be noted that this constraint may limit the ecological validity of the findings, as interpretive development in naturalistic contexts is rarely separable from the practice process itself.

2.3 Score–Performance Alignment and Data Structure

This study used a MusicXML score based on the G. Henle Verlag *Urtext* edition as the reference, and aligned it with the audio/MIDI data at the note level. Alignment was performed with Parangonar [13], an open-source library based on Dynamic Time Warping (DTW). File I/O was handled via Partitura [14]. The resulting match file contains the score position and the performed onset, offset, and velocity for each note, and serves as the primary input for all subsequent analyses.

3. ANALYSIS AND RESULTS

Three analytical methods operating at different resolutions are employed to observe interpretive stabilization, all based on the framework of Widmer et al. [8].

3.1 Tempo–Dynamics Curves: Structural Pacing and Dynamic Stability

Dynamics curves are constructed by averaging the MIDI velocities of all notes in each measure. Tempo curves are aggregated per measure by measuring the elapsed time between adjacent measures and converting it to BPM based on the dotted-quarter-note unit for the 6/8 meter.

Dynamics. Session 1 consistently exhibits lower velocity values across the exposition and the first half of the development compared to all other sessions. This suggests that during the initial reading, the performer consciously or unconsciously suppressed loudness to cautiously explore the movement's sonic weight and atmospheric stillness. Interestingly, toward the end of the coda (after m. 160), the session-to-session variance in dynamics reaches its maximum. This indicates that the expressive strategy for resolving the immense tension accumulated over twenty minutes remained fluid and unsettled even after six sessions. The overall dispersion in dynamics is smaller than in tempo, reflecting the anchoring role of explicit dynamic markings (*pp*, *fp*, *ff*) as objective notational guides that restrict the performer's subjective range.

The relationship between structural boundaries and dynamic stability suggests that the performer relies on the score's formal markers to regulate energy levels. In the transition to the development, where the harmonic texture becomes more complex, the dynamic curves across sessions 4–6 show a remarkable convergence. This implies that while initial sessions involve an exploration of the instrument's resonance, the final sessions are characterized by a refined control of the *piano-vocal* texture inherent in Beethoven's late style.

Tempo. Session 1's total duration of approximately 18.6 minutes is markedly longer than sessions 2–6 (ca. 15–16 min). This reflects a deliberate safety margin adopted during the first performance. Despite this absolute offset, all sessions share a common macro-level tempo contour: a pronounced peak around measures 75–85, coinciding with the highest tonal tension in the development. However, at the critical transition to the recapitulation (ca. m. 110), session 1 stands sharply apart; its tempo drops to an extreme low, reflecting a psychological yielding to the monumental structural weight.

Furthermore, stabilization was found to be slowest in the coda, where notational guidance is sparsest. This tempo yielding at structural junctions highlights the cognitive load associated with navigating such a massive form.

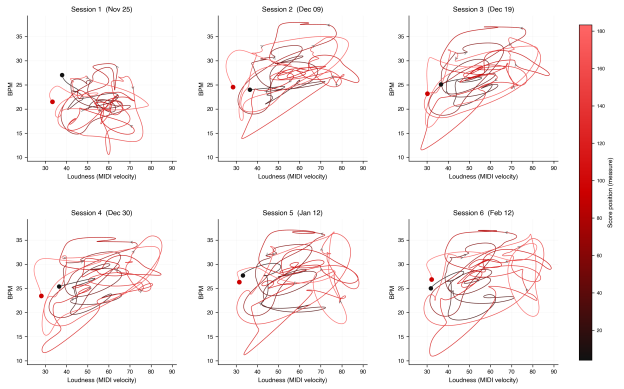

Figure 3. Performance worm trajectories across six sessions. x-axis: loudness (mean MIDI velocity); y-axis: tempo (BPM, dotted-quarter-note unit). Color indicates temporal progression from black to red; Gaussian smoothing $\sigma = 10$ applied.

As the performer becomes more familiar with the transitions, these drastic tempo drops are replaced by more subtle *ritardandi*, suggesting that interpretive stabilization involves a shift from reactive pacing to proactive structural planning. The high variance in the coda further suggests that tempo is the primary parameter through which the performer expresses open-ended interpretation, as historical interpretations vary from Kempff's 17-min to Brendel's 22-min recordings. Formal statistical comparison across sessions remains a direction for future work, given the exploratory nature of this single-performer study.

3.2 Performance Worm: Coordination of Expressive Parameters

The performance worm integrates tempo and loudness into a single 2D coordinate system as a performance trajectory [9]. Gaussian smoothing ($\sigma = 10$, ca. two beats) is applied to suppress note-level fluctuations and foreground phrase-level contours.

The most striking feature is that session 1's trajectory is qualitatively distinct: a prominent horizontal loop where loudness stays near velocity 50 while tempo remains below 27 BPM. This signature indicates that integrated tempo–loudness control had not yet been established, resulting in unstable local tempo surges despite the globally slow median tempo. From session 2, this horizontal loop disappears and trajectories utilize both axes more fully, with the loudness range expanding from 30 to 90. By sessions 4–6, the trajectories resemble one another and terminate in the lower region of the plot, indicating a settled strategy of closing the movement with a consistent diminuendo, aligning with the structural aim of gradual dissolution.

The emergence of an integrated trajectory in later sessions points to a more sophisticated coordination of expressive parameters. By session 5, the worm trajectories show a clear diagonal orientation, indicating a synchronized *crescendo-accelerando* relationship typical of professional expressive practice [10], suggesting that interpretive stabilization involves developing a unified expressive language where tempo and dynamics are no longer treated as independent variables.

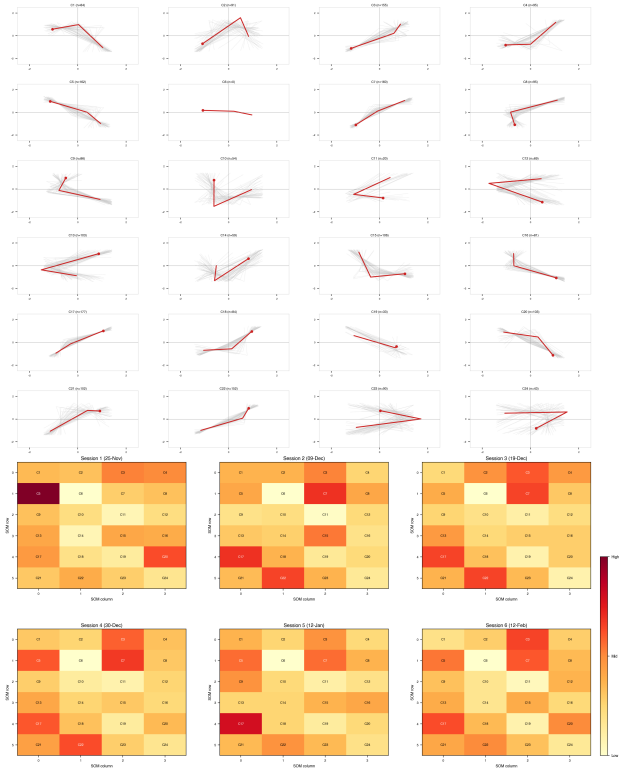

Figure 4. SOM-based performance alphabet and session-wise usage heatmaps. Top: 24 prototype clusters on a 6×4 grid (x = normalized loudness, y = normalized tempo; dots = segment end points). Bottom: cluster usage density by session.

3.3 Performance Alphabet and Heatmaps: Clustering-Based Pattern Analysis

The performance alphabet reduces continuous trajectories into a finite set of archetypal patterns. Following the segmentation-clustering framework of Widmer et al. [8], we trained a 6×4 Self-Organizing Map (SOM) [11] on the normalized trajectory segments, yielding 24 prototype clusters (C1–C24) arranged on a topologically ordered grid. Session-wise heatmaps visualize the probability density of pattern usage across this alphabet per session [5].

The heatmaps reveal a non-linear trajectory of interpretive change. In session 1, activation density is heavily concentrated in C5—an opposing-motion pattern between loudness and tempo—suggesting that the initial performance relied on a relatively limited local coordination pattern before a broader expressive vocabulary emerged. In session 2, density disperses markedly across C7 (ascending *accelerando-crescendo*), C15, C17, and C22, reflecting an active and wide-ranging exploration of opposing expressive strategies within a single performance. This bifurcation captures a critical moment of interpretive uncertainty where the performer actively tests contrasting ways of articulating the same phrase. Sessions 3 and 4 show a partial consolidation, with C7 and C22 remaining

prominent, indicating a sustained but gradually narrowing search around ascending expressive gestures. Session 5 exhibits the most striking shift: density converges sharply around C17 and C22—both characterized by pronounced loudness increases and ascending trajectories—suggesting a decisive convergence on a specific affective strategy. By session 6, the distribution broadens again slightly across C3, C7, and C20, indicating a settled but internally differentiated expressive language rather than a single rigid pattern.

This non-linear trajectory—from the dominance of a single opposing-motion pattern (S1), through wide exploration (S2), partial consolidation (S3–S4), decisive convergence (S5), to stable differentiation (S6)—suggests that the performer's interpretive path is an active search shaped by the expressive demands of the work. The eventual stabilization into clusters characterized by dynamic contrast and directional clarity is consistent with the *innigkeit* (inwardness) of the late Beethoven style and the structural goal of gradual dissolution. This suggests a possible relationship between the performer's later expressive choices and the aesthetic demands of the score, as reflected in the measured expressive parameters. Accordingly, the cluster interpretations reported here should be read as visually grounded hypotheses rather than statistically validated categories.

3.4 Prototype Web-Based Visualization Tool

As a practical extension, we developed a prototype web-based visualization tool designed to support musicological performance analysis. The tool synchronizes the aligned score, piano-roll representation, audio playback, and tempo–dynamics curves, allowing users to inspect how local expressive changes correspond to specific measures and formal regions. The resulting interface is shown in Figure 5.

The system uses an OMR-derived mapping of the PDF score image rather than a purely symbolic electronic score display. This allows the analysis to be anchored to the kind of printed score that performers and musicologists ordinarily use in rehearsal, teaching, and analytical work. Beyond synchronized inspection, the interface can also function as a quantitatively informed interpretive score: analytical observations about phrasing, pacing, dynamics, and structural tension can be marked directly on the score while being supported by the corresponding tempo, loudness, and cluster-pattern data. In this sense, the tool bridges MIR-based measurement and the traditional practice of score-based interpretive annotation.

4. CONCLUSION

This study applied a multi-faceted analytical framework—comprising tempo–dynamics curves, performance worms, and SOM-based performance alphabets—to investigate the longitudinal evolution of musical interpretation. By analyzing six repeated performances of Beethoven's *Hammerklavier* Sonata Op. 106/III recorded on a Yamaha Disklavier, we provided an empirical narrative of how a

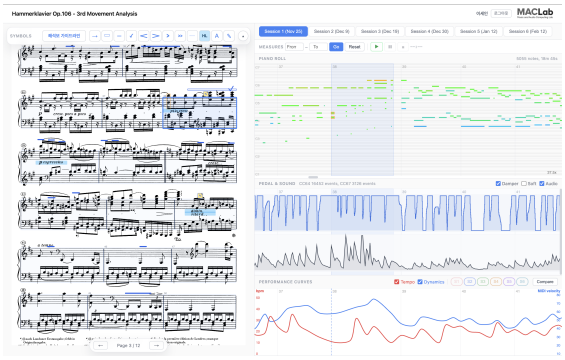

Figure 5. Prototype web-based visualization tool synchronizing score position, piano-roll data, audio playback, and performance curves.

professional pianist constructs and stabilizes an expressive strategy within a monumental structural framework. Unlike traditional performance studies that treat interpretation as a static, finalized product, our findings suggest that the formation of an interpretation is an inherently non-linear and exploratory process.

The convergence of all three analytical methods indicates that interpretive stabilization proceeds differentially according to the work's structural hierarchy. We observed that macro-structural pacing stabilizes early (by Session 2), whereas micro-level expressive nuances, particularly in sections with sparse notational guidance like the coda, remain fluid much longer. The transition from the unstable tempo surges of the initial session to the synchronized coordination in later sessions underscores that interpretive stabilization involves the integration of disparate expressive parameters into a unified interpretive intent. Furthermore, the shift toward deceleration and lower loudness in the final sessions suggests that the performer's search process gradually converges on the aesthetic demands of Beethoven's late style, specifically the affect of *innigkeit*.

While this study offers useful methodological tools for tracking interpretive change, several limitations remain. Our analysis was based on a single performer's data, and while the Disklavier provides high-precision MIDI information, the full spectral richness of acoustic timbre and the subtle nuances of pedaling and voice-leading micro-timing fall outside the scope of the present study. Furthermore, some interpretations rely on visual readings of analytical plots, and the constraint prohibiting inter-session practice may limit the ecological validity of the findings. Future research should expand this framework to include a broader diversity of performers and musical styles, and should incorporate more formal quantitative measures and audio-based spectral analysis.

The prototype visualization tool further illustrates how aligned performance data can be connected with the printed scores used by performers and musicologists in performance, teaching, and analysis. Ultimately, by shifting the focus from the *product* of interpretation to the *process* of its formation, this research argues that important insights into musical meaning are latent not in the singu-

lar performance, but in the persistent and evolving act of repetition.

5. References

[1] M. Müller, *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Cham: Springer, 2015.

[2] 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*, vol. 5, 2018.

[3] W. Goebl and G. Widmer, "On the use of computational methods for expressive music performance," in *Modern Methods for Musicology*, T. Crawford and L. Gibson, Eds. Aldershot: Ashgate, 2009, pp. 93–113.

[4] S. Jeong, Y. Kim, and D. Shin, "Research trends in empirical studies of performers' interpretations: Focusing on timing, dynamics, and gesture," *The Western Music*, vol. 26, no. 1, pp. 137–165, 2023.

[5] C. S. Sapp, "Comparative analysis of multiple musical performances," in *Proc. 8th ISMIR Conference*, Vienna, 2007, pp. 497–500.

[6] F. Foscarin, A. McLeod, P. Rigaux, F. Jacquemard, and M. Sakai, "ASAP: A dataset of aligned scores and performances for piano transcription," in *Proc. 21st Int. Society for Music Information Retrieval Conf. (ISMIR)*, Montreal, 2020, pp. 534–541.

[7] H. Zhang, Q. Kong, S. Dixon, and E. Benetos, "ATEPP: A dataset of automatically transcribed expressive piano performances," in *Proc. 23rd Int. Society for Music Information Retrieval Conf. (ISMIR)*, Bengaluru, 2022, pp. 517–524.

[8] G. Widmer, S. Dixon, W. Goebl, E. Pampalk, and A. Tobudic, "In search of the horowitz factor," *AI Magazine*, vol. 24, no. 3, pp. 111–130, 2003.

[9] S. Dixon, W. Goebl, and G. Widmer, "The performance worm: Real time visualisation of expression based on langner's tempo-loudness animation," in *Proc. International Computer Music Conference*, Göteborg, 2002, pp. 361–364.

[10] J. Langner and W. Goebl, "Visualizing expressive performance in tempo-loudness space," *Computer Music Journal*, vol. 27, no. 4, pp. 69–83, 2003.

[11] T. Kohonen, *Self-Organizing Maps*, 3rd ed. Berlin: Springer, 2001.

[12] Yamaha Corporation, "Disklavier^{tm}," https://usa.yamaha.com/products/musical_instruments/pianos/disklavier/, accessed March 2026.

[13] S. D. Peter, C. E. Cancino-Chacón, F. Foscarin, A. McLeod, F. Henkel, E. Karystinaios, and G. Widmer, "Automatic note-level score-to-performance alignments in the asap dataset," *Transactions of the ISMIR*, vol. 6, no. 1, pp. 27–42, 2023.

[14] C. E. Cancino-Chacón, S. Peter, E. Karystinaios, F. Foscarin, M. Grachten, and G. Widmer, "Partitura: A python package for symbolic music processing," in *Proc. Music Encoding Conference*, Halifax, 2022.