# FINDING A ONE-IN-A-MILLION MUSIC PERFORMANCE: HOW YESTERDAY'S MUSICOLOGIST IS BECOMING TODAY'S COMPUTING ALGORITHM

**Austin Heinz**
University of Washington, Seattle
aheinz92@uw.edu

## 1. INTRODUCTION

At the intersection of technological advancements and artistic traditions lies the potential to leverage data to provide insights, understanding, and pedagogy in ways that traditionally consume immense (and potentially unfeasible) amounts of time and effort. On the algorithmic frontier, interdisciplinary applications such as these are simultaneously overlooked and exciting, and despite lacking the greater forces of trending technology research, there is a notable presence of dedicated scholarly work striving to bring these worlds together. These efforts have been accelerating over recent decades, beginning with manual attempts at data creation to analyze musical recordings at new depth, and developing into approaches using mathematical and algorithmic processing.

One particular area of application is the study of musical recordings, in a practice known as comparative listening: comparing and contrasting different captured interpretations, for the purposes of furthering personal enjoyment, understanding of the music itself, and the quality of one's own interpretive depth as a musician or music teacher. While many classical musicians follow the practice of comparative listening to these ends, it has a natural limitation on both the necessary time and effort as well as a restriction on attention, memory, and cognitive biases; it is easy to lose focus when hearing the same piece repeatedly, even when attempting to maintain attention for subtle differences. Many mental roadblocks exist, including this challenge of extended consistent focus, influence of nostalgia from certain performances, and difficulty remembering how different recordings may have sounded at all.

For all of the important and oft-utilized artistic benefits of this practice, it has fairly immediate complications that frequently prevent a thorough exploration of recorded repertoire and thus demands the attempt to somehow process at larger volumes. Even just a handful of the most famous composers may have hundreds of individual pieces with hundreds of valuable recordings, even excluding all but the most critically-acclaimed - a lofty challenge to parse through for piano miniatures of just a few minutes in length, but nigh impossible for a 30-to-60 minute piano sonata or symphony. Especially at such a scale, the ability to mentally expose contrasts in detail among worthy interpretations is implausible. The most relevant research developments toward solving this problem have been gathered here for categorization and examination.

## 2. APPLICABLE RESEARCH

### 2.1 Chosen musical subject matter

Initial research efforts in the 1990's found ways to break musical performances and recordings into their core components in order to produce analytical results, often with manual data creation and frequently on the subject of piano music. This choice lends itself better to the development of data methodologies - piano music contains many short but distinct pieces with myriad notable recordings, and the instrument is rare in that a single note sounds the same when pressed by a beginner or a master. Most of what is truly occurring in piano playing is the result of the timing and volume of each note press, with the singular exception of the application of the pedals (primarily the sustain pedal, to hold the sounds of notes), and as artistically significant as piano pedaling can be, accurate and deep analysis can be done without it. The overall impression of sections or whole pieces of piano music are simply the larger impact created by many of these individual moments of timing and volume, and can theoretically be accurately captured by a MIDI (Musical Instrument Digital Interface) format.

Within this domain, an especially popular focus has been the famous mazurkas (named after a traditional Polish dance) by composer Frédéric Chopin, as they have many qualities that make them a prime target for research. Their short durations, countless respected recordings, and distinctly-defined rhythmic pulse help align and validate the analytical methods as they are being developed. Music without a strong pulse is far harder from which to produce good data; analyzing broad subtle tempos or aligning to structures in order to compare equivalent passages is, as of yet, very difficult without this clear auditory scaffolding. For nearly two decades, the Chopin Mazurkas have been a common recipient of methodological development, from earlier groundbreaking research by the eminent Nicholas Cook to the first massive dataset in this field, named MazurkaBL, which focuses specifically on these particular piano works (Kosta et al., 2018).

### 2.2 Types of research

As the overall aim is to find a strong analytical process that can produce meaningful analyses on volumes of recordings too large for a human to feasibly tackle, the vast majority of relevant papers deal with methodological developments. Nearly every piece of research outlines a new idea for a data-driven approach, to varying degrees of technological involvement, and then applies that approach in a case study of a certain set of recordings to both analyze interesting findings and to demonstrate its potential application and current viability. Currently, there is an explosion in many different directions of development, as the availability of new technological approaches is being explored, but it is likely that over time these methodological ventures will begin to centralize on effective approaches. To get there, however, takes a vast amount of frontier exploration, primarily concerned with hypothesizing and testing ideas for these new methods.

There exist several notable contributions that do not fit this mould, however: The production of the MazurkaBL dataset, the culmination of years focused specifically on Chopin's mazurkas, is described in detail and contributes massively to the field of research (Kosta et al., 2018). In fact, a new methodology paper about a very complex process coined "end-to-end bayesian segmentation and similarity assessment," comparing and contrasting recordings through timings and volumes, utilizes the MazurkaBL dataset and draws upon its methods (Guichaoua et al., 2024). Another notable exception is one of the earliest applicable pieces of research, one of the only examples of actual experiments, where recorded performances are stripped down to their core unique attributes (with the help of already-existing digital piano technology), making all other attributes equivalent in the process to allow for fair comparison. In two unique experiments, panels of professional pianists and scholars evaluated their listening experiences, and fascinatingly, an "average" performance was also played that resulted from calculating the middle-ground of all the other presented recordings (Repp, 1997). While the experiment results showed fascinating findings about what experts prefer, the process itself also greatly contributed a meaningful step towards processing performance elements as data.

| Author(s) | Journal or Conference | Year | Type | Subject Matter | # of Recordings Analyzed | Process |
|---|---|---|---|---|---|---|
| Repp | Music Perception | 1997 | experiment | solo piano music (Schumann & Chopin, not mazurkas) | 41 | manual |
| Cook | Musicae Scientiae | 2007 | methodology, case study | solo piano music (two mazurkas by Chopin) | 50 | both manual & algorithmic ("tempo data") |
| Sapp | ISMIR[a] | 2008 | methodology, case study | solo piano music (five mazurkas by Chopin) | 232 | algorithmic ("hybrid numeric/rank similarity metrics") |
| Liem, Hanjalic | ISMIR[a] | 2015 | methodology, case study | orchestral music (Beethoven & R. Strauss) | 31 | algorithmic ("image-based") |
| Dorfer, Arzt, Widmer | ISMIR[a] | 2016 | methodology | solo piano music | n/a | algorithmic, by way of neural network models ("score following") |
| Kosta, Bandtlow, Chew | TENOR[b] | 2018 | dataset creation, methodology | solo piano music (mazurkas by Chopin) | 2,239 | both manual & algorithmic ("score-aligned loudness, beat, and expressive markings data") |
| Yanchenko, Hoff | Annals of Applied Statistics | 2020 | methodology, case study | orchestral music (Beethoven) | 370 | algorithmic ("hierarchical multidimensional scaling") |
| Zhou, Fabian | Musicae Scientiae | 2021 | methodology, case study | solo piano music (Chopin, not mazurkas) | 2 | manual ("three-dimensional tempo model") |
| Zhang, Liang, Dixon | ISMIR[a] | 2024 | dataset creation, methodology, case study | solo piano music (Chopin) | 137 | both manual & algorithmic, by way of trained models ("pianism-labelling dataset" & "expertise ranking") |
| Guichaoua, Lascabettes, Chew | Music & Science | 2024 | methodology, case study | solo piano music (mazurkas by Chopin) | 37 | algorithmic ("end-to-end Bayesian segmentation and similarity assessment") |

[a] International Society for Music Information Retrieval

[b] Technologies for Music Notation and Representation

**Table 1.** Selected journal and conference papers that have contributed research, ordered chronologically and categorized by the type of research, the type of music used as a target of analysis, the volume of pieces analyzed, and the applied process. Colors highlight some notable features, either due to their infrequency (non-methodological paper types) or those that are repeated (categories of musical subject matter). All works in the table are cited in the full reference list.

## 3. CONCLUSIONS

It is important to note that this is an extremely small and preliminary analysis, gathering many of the most relevant research papers but not including dozens of more modern and specialized advances that are worth being explored in a proper meta-analysis. This paper should be seen as a miniature rough-draft that demonstrates the concepts and aims of a larger and properly thorough exploration of nearly all appropriate literature. It has been formatted in a journal-esque style for the sake of novelty, but likely includes many mistakes when compared to a true academic approach (appearance, poor citation work, etc.) and should in no way be taken as a genuine survey. It is clearly missing entire sections, as permitted length demanded a hard stopping point. Therefore, this merely hopes at best to serve as a basis for a substantial future endeavor.

Regarding research that ought to be included in a proper overview: there exists in particular a large overlap in technological methodologies for both the goal of comparative analysis and, far more popular in the current zeitgeist, music generation by large-language artificial intelligence models. These applications face overlapping challenges in separating music into important components, and being able to perceive what is occurring that matters to the mind and ears of a listener - by crushing everything down into data and making sense of it. Music generation is essentially the reconstruction after the deconstructive process of musical comparative data analysis, and benefits from many of the same advances; for this reason, many researchers have published work on both areas, and additionally in the field of artificial music pedagogy (with the ability to track with music, evaluate performance, and inform playing quality and style).

As for the overall state of computational musicology research, there is much to be excited about but very far yet to go. There is suddenly a vastness of possibilities for new methodological development, arriving finally thanks to new technology after decades of manual and semi-manual data labeling and minor algorithmic attempts. Results are accelerating, but there is also a risk of a lack of focus, and of too many methods being developed without enough dedication and maturing of any particular one.

The goal itself is nuanced and potentially challenging as well - what might happen if a tool is indeed created that allows for instantaneous listening analysis of countless recordings of every piece of music? It's possible that it might have a negative impact on the overall experience of musical enjoyment, by incorrectly over-centralizing performance tradition around a supposed "average" or "most typical" way to perform a certain composition (purely as a result of algorithmic bias, which could be due to coincidence), or indeed to over-value uniqueness in interpretations due to their ability to eventually be immediately flagged by analytical tools as interesting outliers. Perhaps this creates a world where musicians simply compete to stand out in an algorithm, lest they become lost in a sea of hundreds of what appear to be too-similar recordings. Ultimately, this future may never come, but if any methods truly succeed, there is some possibility of harmful ramifications on the musical arts if measures aren't taken to prevent domination by algorithms. For now, at least, the future is bright and the directions of exploration are more varied than ever before, and one can hope that continuing research indeed unlocks the bounties of comparative listening implausible by a single human mind alone.

## 4. REFERENCES


Cook, N. (2007, July). Performance analysis and Chopin's mazurkas. *Musicae Scientiae*, *11*(2), 183-207. https://doi.org/10.1177/102986490701100203

Dorfer, M., Arzt, A., & Widmer, G. (2016, December 15). Towards score following in sheet music images. *International Society for Music Information Retrieval (ISMIR)*. https://doi.org/10.48550/arXiv.1612.05050

Guichaoua, C., Lascabettes, P., & Chew, E. (2024, March 24). End-to-end Bayesian segmentation and similarity assessment of performed music tempo and dynamics without score information. *Music & Science*, *7*. https://doi.org/10.1177/20592043241233411

Kosta, K., Bandtlow, O. F., & Chew, E. (2018, May). MazurkaBL: Score-aligned loudness, beat, and expressive markings data for 2,000 Chopin mazurka recordings. *Technologies for Music Notation and Representation (TENOR)*, 85-94.

Liem, C. C. S., & Hanjalic, A. (2015, October 26). Comparative analysis of orchestral performance recordings: An image-based approach. *International Society for Music Information Retrieval (ISMIR)*, 302-308.

Repp, B. H. (1997, July 1). The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception*, *14*(4), 419-444. https://doi.org/10.2307/40285732

Sapp, C. S. (2008, September 14). Hybrid numeric/rank similarity metrics for musical performance analysis. *International Society for Music Information Retrieval (ISMIR)*, *501*, 506.

Yanchenko, A. K., & Hoff, P. D. (2020, December). Hierarchical multidimensional scaling for the comparison of musical performance styles. *Annals of Applied Statistics*, *14*(4), 1581-1603. https://doi.org/10.1214/20-AOAS1391

Zhang, H., Liang, J., & Dixon, S. (2024, July 5). From audio encoders to piano judges: benchmarking performance understanding for solo piano. *International Society for Music Information Retrieval (ISMIR)*. https://doi.org/10.48550/arXiv.2407.04518

Zhou, D. Q., & Fabian, D. (2019, September 19). A three-dimensional model for evaluating individual differences in tempo and tempo variation in musical performance. *Musicae Scientiae*, *25*(2), 252-267. https://doi.org/10.1177/10298649198731