Supporting Music Education through Visualizations of MIDI Recordings

Frank Heyen*
VISUS
University of Stuttgart

Michael Sedlmair†
VISUS
University of Stuttgart

ABSTRACT

Musicians mostly have to rely on their ears when they want to analyze what they play, for example to detect errors. Since hearing is sequential, it is not possible to quickly grasp an overview over one or multiple recordings of a whole piece of music at once. We therefore propose various visualizations that allow analyzing errors and stylistic variance. Our current approach focuses on rhythm and uses MIDI data for simplicity.

Index Terms: Human-centered computing—Information visualization Applied computing—Interactive learning environments

1 INTRODUCTION

Our work focuses on investigating how visual data analysis can be leveraged in music education and training. While several works exist on how to support static sheet music through visualization [7,11], our focus is on visualizing actual recordings of notes that musicians play while learning their instrument. We specifically hypothesize that comparative visualization [2] can be a viable approach for leveraging such data. In a traditional tutoring setup, for instance, we could record what both a student and a teacher are playing, and then use this data for comparative and easy to understand visualizations. For auto-didactic learning, we could compare actual played notes of a student to a "ground truth" piece of sheet music. Comparative visualization could then be used to help identify errors in playing and guide the process of future practice.

The data behind such an analysis can be represented in different formats, such as MusicXML [3] or MIDI (Musical Instrument Digital Interface) for abstract notes, or WAV (Waveform Audio File Format) for raw audio recordings. To keep the concept simple, we decided to start with MIDI data instead of raw audio, since abstract notes are easier to interpret and visualize by a computer. While traditionally MIDI was closely connected to keyboards, many other instruments now natively support it, such as electronic drum kits, MIDI guitars, and even MIDI saxophones.

In this work, we contribute initial design considerations and prototypical implementations for comparative MIDI visualizations. The poster focuses on MIDI drum recordings and the task of analyzing rhythm in terms of consistency and errors.

Methodologically we are inspired by design study methodology [10] and action research [4]. One of the authors taught guitar for seven years at a music school and is currently learning drums.

2 RELATED WORK

Although there is diverse research regarding music visualization [5] and visual musicology [8], there is not much usage of instrument-generated data for education yet. Rocksmith[1] and Synthesia[2] are examples for instrument education games. Such games only give immediate or coarsely aggregated feedback and do not allow analyzing

recordings on a per-note basis. In terms of visual encoding, they are mostly limited to showing coarse scores or line charts. They do not show detailed feedback on a per-note basis, restricting the insight a user might gain. There has been little work in actively leveraging visualization and visual analytics for this endeavor, which is the focus of our work.

Abstractly, our data can be viewed as a sequence of played notes. Analyzing temporal data is a core area in visualization research [1], and various approaches for visual sequence analysis exist [6,9,12]. We aim at contributing the exploration of a new application area in this field.

3 DESIGN

This sections provides details on the kinds of data we consider, as well as users and tasks we aim to support.

Data   We focus on MIDI data, which represents music as abstract notes instead of audio. This makes our design simpler and more reliable, but still works with a range of instruments. We differentiate between two kinds of data: *recordings* and *ground truth*. Recordings capture the MIDI output of an instrument, saving the played notes for later analysis, while the ground truth is the correct version of a piece of music that the users wants to learn. MIDI files containing ground truth are abundant online and easy to create with existing tools. For example, a search for *MIDI drum dataset* reveals a collection of 800,000 drum patterns[3]. Alternatively, a MIDI recording from a music teacher might be used. Without access to a ground truth, for example when improvising or composing, some of our visualizations are still able to provide insight.

Users and Tasks   Our target users are musicians who want to learn or improve their skills at playing an instrument. We want to enable those users to see where they make errors such as hitting the wrong note or playing it slightly too early or too late. To reduce effort and allow for temporal overviews, visualizations should allow analyzing multiple recordings at once, to better reveal consistent patterns. For example, a drummer might find that he consistently lags behind on the bass drum. This kind of consistent offset might even be on purpose, depending on playing style.

Visual Encodings   In order to create intuitive visualizations that reduce the effort for learning and reading, we build upon standard music notation, such as scores and tablature. We draw measures and beats with a size proportional to their duration in seconds to allow perceiving durations through pre-attentive processing (Fig. 1).

Figure 1: A simplified, notation-based representation of the ground truth drum pattern. This serves as basis for most of our visualizations.

With a single recording, a naive approach of simply plotting all ground truth notes and recording notes works fine. We use colors to encode the kind of error as follows: red for missing notes, yellow for surplus ones, and green for correct notes (Fig. 2). Notes are

---

*e-mail: frank.heyen@visus.uni-stuttgart.de

†e-mail: michael.sedlmair@visus.uni-stuttgart.de

considered correct if they can be matched to a ground truth note, such that for each recording at most one recorded note is assigned to one ground truth note. Correct notes are usually not perfectly timed; we use different shades of green to encode the timing error. While this visualization has the advantage of showing error types, it does not scale to larger numbers of recordings, as drawing all notes leads to over-plotting.
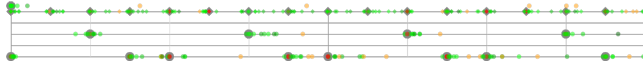

Figure 2: Recorded notes of multiple recordings (colored) drawn on top of the ground truth notes. Colors indicate missing, surplus, and correct notes.

A more scalable alternative is to draw one density estimation chart per pitch (Fig. 3). In the case of drum patterns, this means one chart for each of its components such as the bass drum and cymbals. For example, we take the start times of all hi-hat notes from all recordings and calculate the density estimation for them. Through this aggregation, an arbitrary number of recordings can be shown at once to analyze general patterns such as syncopation.
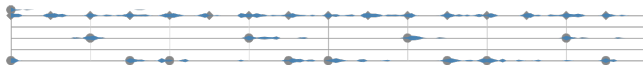

Figure 3: To reduce clutter, this visualization does not draw single notes but instead uses a density estimation area chart to show the temporal distribution of recorded notes.

The most concise encoding only draws the notes of the ground truth data (Fig. 4). Each note gets colored by the average error of the recorded notes that are matched to it. This allows users to quickly spot the most problematic parts of the piece and may serve as an overview to start a more thorough analysis from.
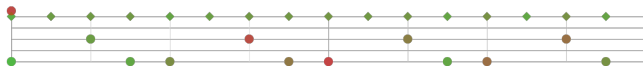

Figure 4: Drawing only the ground truth, colored by the average error over all recordings, creates a clear image that provides insight into where the problematic parts are.

Since different drum components are played differently, users might be interested in a summary for each component. We implemented such a visualization that aggregates over the whole duration of a track. With a density estimation area chart, it shows the distribution of the time differences between played and correct notes of all recordings in a compact fashion (Fig. 5). This visualization could reveal that a user makes larger errors on the bass drum, which is played via foot, than components that are played with hands.

4 CONCLUSION AND FUTURE WORK

We proposed five visual encodings for musical performance analysis using drum MIDI recordings. Currently, we are in the process of integrating these visualizations into an interactive prototype that allows users to detect errors and analyze their style of playing. In terms of interaction, we are implementing different mechanisms for selections to dynamically aggregate sub-sequences in the data. Our current version already supports synthesized playback of both the ground truth and single recordings, other visual encodings, linking and brushing, and other features. At the heart, these are different versions of the above five representations though. Qualitatively evaluating first versions of this prototype indicated that users might indeed profit from augmenting their auditory analysis with visual
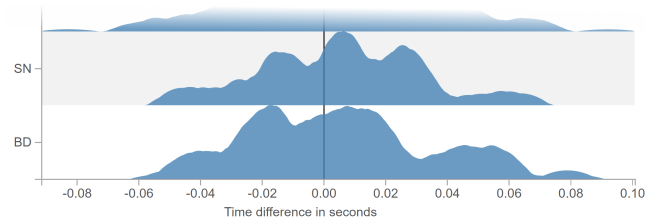

Figure 5: A time-aggregated view shows the distribution of time differences between played and matched correct notes for each drum kit component (here BD = bass drum, SN = snare).

encodings. In the future, we also plan dedicated comparative studies to better understand the design space.

As we focus on MIDI data only, our approach is limited to instruments that are able to produce the required output. We plan to extend our concept to different kinds of instruments that support MIDI, such as guitars with MIDI pickups, and will also look into incorporating instruments without native MIDI support.

REFERENCES

[1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of time-oriented data*. Springer Science & Business Media, 2011.
[2] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.
[3] M. Good et al. MusicXML: An internet-friendly format for sheet music. In *XML conference and expo*, pp. 03–04. Citeseer, 2001.
[4] G. R. Hayes. The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(3):1–20, Aug. 2011. doi: 10.1145/1993060.1993065
[5] R. Khulusi, J. Kusnick, C. Meinecke, C. Gillmann, J. Focht, and S. Jänicke. A survey on visualizations for musical data. *Computer Graphics Forum*, 2020. doi: 10.1111/cgf.13905
[6] D. Luo et al. EventRiver: Visually exploring text collections with temporal references. *IEEE Transactions on Visualization and Computer Graphics*, 18(1):93–105, 2012.
[7] M. Miller, A. Bonnici, and M. El-Assady. Augmenting music sheets with harmonic fingerprints. In *Proceedings of the ACM Symposium on Document Engineering 2019*, DocEng '19. ACM, 2019. doi: 10.1145/3342558.3345395
[8] M. Miller, H. Schäfer, M. Kraus, M. Leman, D. Keim, and M. El-Assady. Framing visual musicology through methodology transfer. *arXiv preprint arXiv:1908.10411*, 2019.
[9] M. Sedlmair, P. Isenberg, D. Baur, and A. Butz. Information visualization evaluation in large companies: Challenges, experiences and recommendations. *Information Visualization*, 10:248–266, 09 2011. doi: 10.1177/1473871611413099
[10] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18(12):2431–2440, 2012. doi: 10.1109/TVCG.2012.213
[11] M. Wattenberg. Arc diagrams: visualizing structure in strings. In *IEEE Symposium on Information Visualization, INFOVIS*, pp. 110–116. IEEE, 2002.
[12] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman. LifeFlow: Visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, p. 1747–1756. ACM, 2011. doi: 10.1145/1978942.1979196