VISUALIZING MUSIC STRUCTURE USING SPOTIFY DATA

Job Savelsberg
Eindhoven University of Technology
j.w.l.savelsberg@student.tue.nl

ABSTRACT

Understanding a song's structure is crucial to a musician's ability to learn and memorize a song. Discovering, explaining, and sharing structural patterns through visual representation is still done mostly by hand. We aim to help musicians by providing a tool capable of automatically generating a visualization that contains a clear and comprehensive overview of repetition and change of musical aspects such as harmony, timbre, tempo, and dynamics. Using abstracted audio-content data provided by Spotify, existing structure extraction strategies are combined and expanded upon. Six separate visualization modules are created to accommodate the user in understanding and navigating the music by its structure.

1. BACKGROUND

The problem of music structure extraction and visualization has been covered extensively [1]. The problem is often posed in terms of a sequence of repeating sections of music, where sequentially similar sections are given a letter as a label. This results in solutions looking like **ABACABA** (rondo form) for example. Commonly, this problem is tackled using a self-similarity matrix (SSM) as a starting point for repetition extraction [2,3] and visualization [4,5]. In the context of this research, we do not restrict the definition of music structure to song form. Despite its relative importance, visualizing other structural aspects of music is also meaningful: the change in chords [6, 7], dynamics, timbre [8], rhythm/tempo [9, 10] and tonality [7, 11] over time. The most notable attempt at including all musical dimensions is Songle [12]. One limitation to the currently available tools is the lack of data. Due to the difficulty for acquiring raw audio files from a music streaming service, we use abstracted music content data. The data used in this project is generated and made publicly available as a Spotify API using the Echo Nest Analyzer [13]. For our project we use chroma features, timbre features (time-segmented 12-dimensional spectrogram embedding), and rhythm features (bar, beat and tatum detection).

2. USERS AND GOALS

While this tool should be open for anyone to use, we mainly focus on musicians and ensembles. The visualization should aid musicians in learning and memorizing music structure, and aid ensembles by providing a common structural representation useful for referencing. For all users, the visualization should improve navigation of music; instead of timestamps, users should be able to search for specific moments using visual attributes based on the music's content. Next to this, it is important to limit the size and complexity of the visualization, make the information extraction generalize across different types of music, and present it in a visually appealing manner. Next to the visualization, this tool as a service also needs to attain three quantifiable goals (fast, vast, and last): the visualizations should be generated in a matter of seconds, it should have access to a library of music that is as large as possible, and it should be constantly brought up to date with releases of new music or improvements to the feature extraction methods. The latter two directly determine the decision for using a music streaming platform's API for data.

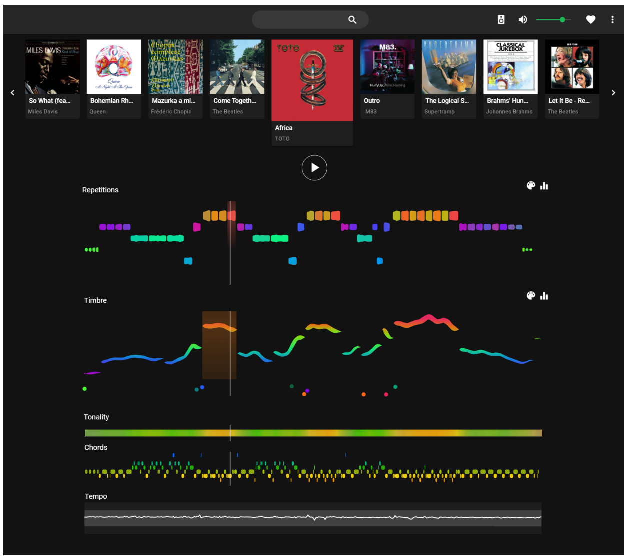

Figure 1. The complete tool, combining all visualization modules and integrating streaming platform search and playback.

3. COMPONENTS

3.1 Repetition

For repetition detection, we base our solution on the Scape plot by Müller et al. [4], which, given the chroma features of a song, visualizes sequential similarity between all possible sections. We adapt the strategy to find a discrete set of sections, where each section has a label; sections with the same label being repetitions of one another. The method tries to find an optimal decomposition of sections that minimizes overlap and maximizes fitness (a combination of coverage and sequential similarity). On top of a label, we also visualize how similar any two sections are on a continuous scale. This similarity informs the user whether it is an exact repetition, a variation, or a completely different section. Since spatial perception is essential to the perception of structural patterns the result is visualized as several rows of repeating section blocks. Using multidimensional scaling (MDS) given a sequential similarity function we map each block to a position on a color wheel, such that sequentially similar blocks have a similar color. We introduce a clustering coefficient κ that determines whether the similarity within or between groups of detected repetitions should be prioritized, see Figure 2.

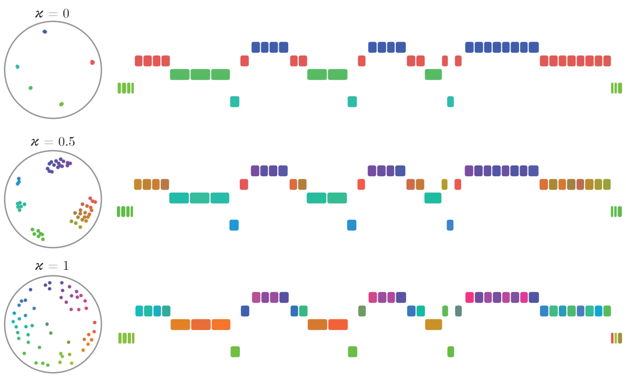

Figure 2. Visualization of the MDS mapping of repetitions with different values of κ (Song: *TOTO - Africa*).

3.2 Timbre

The timbre feature provided by Spotify is a 12-dimensional vector representing a high-level abstraction of the spectral surface. This means the data contains both concepts of individual instrument's timbre and the texture of the mix of instruments. This creates the challenge of visualizing either the continuous change in timbre and dynamics or discrete changes in the set of playing instruments. As a solution, we introduce a segmented graph visualization that shows the timbral relation between moments in time with both gradual change and segmentation. This hybrid approach combines a segmentation step and an MDS step. Segmentation is achieved with novelty detection on smoothed features to find peak moments of timbral change. MDS is used to map the gradual change within segments to a single dimension suitable for visualization. Next to a coarse visualization of timbral change, we also provide an "Event" visualization that extracts moments of timbral anomaly and visualizes it as points over time.

3.3 Tonality and Chords

Due to the ambiguity and fluidity of tonality, we visualize tonality as a continuous feature, while chords are treated as a discrete decomposition of the song. Both the tonality and chord visualization use the same mapping of the circle of fifths to colors. The tonality visualization is a combination of a common template-based key-detection method applied to a moving window and a harmonic color mapping inspired by Ciuha et al. [7]. The chord visualization also uses a template-based detection method to extract chords. Vertical positioning and coloring of the chord blocks are determined by the circle of fifths, and notably, relative major-minor chords are given the same mapping, which makes common chord detection mistakes of relative chords minimally impact structure perception.

3.4 Dynamics and Tempo

Both dynamics and tempo can be considered one-dimensional, continuous features of music. While tempo is simply visualized as a curve, dynamics is incorporated in the repetition and timbre visualizations as thickness.

4. COMBINED RESULT

All visualizations have been purposely created such that they can be combined into one comprehensive overview, see Figure 1. Users are enabled to tweak visualization options such as positioning and coloring. Generating the visualization is done in a matter of seconds, and the integration of the streaming service's API allows the search and playback of any song on the platform.

5. CONCLUSION AND FUTURE WORK

We have explored how the structure in different musical dimensions of a song can be extracted, visualized, and combined into a single comprehensive tool. We achieved the goals set for the visualization by limiting the size and complexity of the visualization, making the information extraction generalize across different types of music, and enabling the user to reference parts of the music by visual attributes. Future work includes replacing template-based methods of key- and chord-detection and timbre anomaly detection with state-of-the-art methods. Further research into optimal decomposition of repetitive structure is needed, especially given the intrinsic challenge of hierarchical structure [14]. No extensive qualitative user test has been performed yet, and collaborative error correction is to be considered [12, 15]. Despite this, we believe this tool has the potential to not only help musicians, but can also spark interest in experiencing and listening to music more intently. In other words, once completed, everyone is invited to use it [1] .

[1] Information about the tool and its progress will be available at `http://jobsavelsberg.com/musicstructure`

6. REFERENCES

[1] J. Savelsberg, “Visualizing the structure of music,” 2020.

[2] T. Cheng, J. B. L. Smith, and M. Goto, “Music structure boundary detection and labelling by a deconvolution of path-enhanced self-similarity matrix,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[3] F. Kaiser and T. Sikora, “Music structure discovery in popular music using non-negative matrix factorization.” in *ISMIR*, 2010.

[4] M. Müller and N. Jiang, “A scape plot representation for visualizing repetitive structures of music recordings.” in *ISMIR*, 2012.

[5] H. Wu and J. Bello, “Audio-based music visualization for music structure analysis,” in *Proceedings of the 7th Sound and Music Computing Conference, SMC 2010*, 2010.

[6] B. de Haas, J. P. Magalhaes, D. Ten Heggeler, G. Bekenkamp, and T. Ruizendaal, “Chordify: Chord transcription for the masses,” 2014.

[7] P. Ciuha, B. Klemenc, and F. Solina, “Visualization of concurrent tones in music with colours,” in *Proceedings of the 18th ACM International Conference on Multimedia*, 2010.

[8] T. Takahashi, S. Fukayama, and M. Goto, “Instrudive: A music visualization system based on automatically recognized instrumentation.” in *ISMIR*, 2018.

[9] F. Benadon, “A circular plot for rhythm visualization and analysis,” *Music Theory Online*, 2007.

[10] J. Foote and S. Uchihashi, “The beat spectrum: A new approach to rhythm analysis,” 2001.

[11] C. S. Sapp, “Harmonic visualizations of tonal music.” in *ICMC*, 2001.

[12] M. Goto, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano, “Songle: A web service for active music listening improved by user contributions.” in *ISMIR*, 2011.

[13] T. Jehan and D. DesRoches, “The echo nest analyzer documentation,” 2011. [Online]. Available: https://www.yumpu.com/s/yrXiztqNUMiDOBLE

[14] B. McFee, O. Nieto, M. M. Farbood, and J. P. Bello, “Evaluating hierarchical structure in music annotations,” *Frontiers in psychology*, 2017.

[15] H. V. Koops, W. B. de Haas, and A. Volk, “Integration of crowd-sourced chord sequences using data fusion,” in *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR*, 2015.