

# VISUALIZING MUSIC IN ITS ENTIRETY USING ACOUSTIC FEATURES: MUSIC FLOWGRAM

Dasaem Jeong and Juhan Nam

Graduate School of Culture Technology, KAIST

{jdasam, juhannam}@kaist.ac.kr

## ABSTRACT

In this paper, we present an automatic method for visualizing a music audio file from its beginning to end, especially for classical music. Our goal is developing an easy-to-use visualization method that is helpful for listeners and can be used for various kinds of classical music, even for complex orchestral music. To represent musical characteristics, the method uses audio features such as volume, onset density, and auditory roughness, which describe loudness, tempo, and dissonance, respectively. These features are visually mapped into static two-dimensional graph, so that users can see how the music changes by time at a look. We implemented the method with Web Audio API so that users can access to the visualization system on their web browser and make visualizations from their own music audio files. Two types of user tests were conducted to verify the effects and usefulness of the visualization for classical music listeners. The result shows that it helps listeners to memorize and understand a structure of music, and to easily find a specific part of the music.

## 1. INTRODUCTION

Music visualization is widely used in various music activities for many purposes. Because music is an auditory art, its visual representations can contain information that cannot be transferred or preserved accurately with sound. Music notation is a typical example of the visualized music representations. It is designed for communication between composers and performers. The notation systems thus have been evolved to represent and deliver a composer's intention as precise as possible.

For listeners, however, music notation has some limitations. It contains too much information for listeners to interpret and so only a small part can be understood while following the music. Especially, in the case of orchestral music, the score following task is quite difficult unless the listeners are musically trained. Another problem is that the notation does not show the entire structure of a piece of music. The time scope of a music score that can be read in a sight is limited to a few measures. The notation is focused on delivering information about what is happening in a specific time. To understand the global structure,

one needs to read through the score for a while, having a certain level of musical knowledge.

As a way of making up the shortcomings of music notation, audio-synchronized music scores have been developed [1]. Synchronized scores automatically follow music on the score so that listeners can easily track where the currently playing measure is or select the measure on the score to play the music from the position. However, such systems require a synchronization process between audio and score. Manual synchronization is too laborious to process a large set of pieces, whereas automatic one, an active research topic in the area of music information retrieval (MIR), is not accurate enough particularly for large orchestral music. Above all, these solutions still cannot show the entire structure of a piece.

In the case of classical music, particularly for long instrumental pieces, visualizing information about the entire structure can be helpful to music listeners in that they do not contain lyrics or clear storytelling to follow. So additional information about the music is required to help listeners to understand the music. A traditional way of providing the information is giving a lecture or writing a program note. But these requires professionals who can explain the music. Many researchers instead have suggested a content-based approach to visualize the entire structure of music from audio. Most of automatic music structure visualization methods are based on self-similarity between each part of a piece [2, 3, 4]. These methods show a repetitive structure of the music based on the self-similarity. Further information about the research is introduced in Section 2. In general, it is not easy to interpret the meaning of the visualizations. Finding the repetitive structure can help music structure analysis, but its usefulness on listeners has been not verified yet.

To address this problem, we present an automatic music visualization method named *music flowgram*, which aims to visualize an entire piece as an easy-to-understand image. It extracts audio features from audio files and visualizes them on a static two-dimensional graph. In our previous research, we found that a simple static graph showing the change of volume of a music piece can help listeners to concentrate more on classical music, compared to spectrum-based real-time visualization [5, 6]. We have improved this concept by adding additional features that can represent other important characteristics of music, and conducted user test to verify its effect on listening to classical music.

The later part of this paper is organized as follows. First,

related work on visualizing music structure is briefly reviewed. Then, we present our visualization method in two sections: concept of the visualization and audio features in Section 3, and its implementation in Section 4. The detailed information about user tests are described in Section 5, and results with discussion in Section 6. The last section concludes the paper with a summary and our plan for future work.

## 2. RELATED WORK

There has been some research on visualizing the structure of a music piece, both in data visualization and MIR areas. The majority of them exploited the repetitive structure of music using self-similarity within a piece. Wattenberg visualized it using an arc diagram that connects each repetitive part with an edge being drawn as a semi-circle [2]. Foote visualized the self-similarity as a two-dimensional matrix where each element is calculated from similarity between two audio frames [3]. Müller and Jiang extended it to a scape plot representation that visualizes the repetitive structure with varying segment size. Other researchers combined this self-similarity information with volume transitions over time [7]. There has been also research that applies this structural information to music listening interfaces [8, 9].

Other than those based on repetitions in music, some work visualized the structure using tonality such as key change over multiscale segments [10]. Malt and Jourdan presented a visualization method using statistical characteristics of spectral information, including spectral centroid and standard deviation of the audio spectrum [11]. They illustrated the change of those information over time on a two-dimensional graph, adding amplitude information as a color of the graph. However, the most of the mentioned research have not released an end-user application so that general users can render their own visualization. Furthermore, this research lacks user test or human side experiments that verify its effect and usefulness for listeners.

Besides the automatic visualization methods using audio files or MIDI files, visualization of semantic structure of music is also proposed [12]. This method contains a lot more information than repetitive structure, for example, traditional structure analysis of sonata form, motif development, and how the role of each instrument changes through the piece. But all of the information used in visualization is manually extracted from written explanation of the music, and cannot be automatically computed from audio files.

There is also music psychological research about visualizing whole music [13]. This research tested how people describe short music with graphical representations. Participants are asked to “make any marks” to describe five short orchestral works after listening to the music. The result showed that musically trained participants more tended to describe music with abstract representations such as symbols and lines. Most frequently used mapping was X-axis as time and Y-axis as pitch. The other type was pictorial representations, which were mostly drawn by untrained participants. Among 30 musically trained participants, 24

used an abstract representation and 21 of them were in continuous mode. This result indicates that a two-dimensional graph is natural in human sense for representing whole music piece.

## 3. MUSIC FLOWGRAM

The idea of music flowgram for music visualization is based on dramatic structure of storytelling. Freytag explained the structure of each story with two-dimensional graph visualization of tension progress [14]. Our idea is applying a similar concept to music: drawing continuous two-dimensional graph that shows the change of music by time. If listeners can see a dramatic structure of music, they could feel more comfortable to concentrate on the music because they can clearly see when the tension will increase or decrease. This is similar to watching an opera, for which people are encouraged to know dramatic structure before watching. The visualization will also help the listeners to recall the sequence of the music, as people remember the order of opera story based on the order of important events.

A similar type of visualization is waveform visualization or volume graph. It shows the volume progress of the music so that users can see which part is loud or quiet. This type of visualization is used in SoundCloud<sup>1</sup>. Though volume is a highly important factor in deciding characteristic of music, there are other quantitative parameters to explain the music. Spectrogram is another way to show the variance of music as a two-dimensional image. However, it contains too much details to deliver meaningful musical information. Thus, more compact representations, which effectively extract musical elements, is needed.

Considering that emotion is the most influential high-level concept on listeners, we focus on musical elements that are associated with the emotional aspects of music. Among many suggested elements in this regard [15], we choose loudness, tempo and harmony. For visualization, we represent them with volume, onset density and auditory roughness, respectively, as below.

### 3.1 Volume

Unlike other genres of music, classical music consists with many different sub-parts, each of which has a different loudness characteristic. Therefore, temporal differences of loudness can explain the structural information of music effectively. We represent the loudness with volume which is simply calculated as frame-level energy. Though more complex measures of loudness could be adopted, we assume that the volume is sufficiently effective in complex musical sound.

### 3.2 Onset Density

Emotion of music is highly dependent on the tempo characteristic of music, i.e, whether the music is fast or slow. Beats per minute (BPM) is a typical way of representing it. However, the single speed measure is not sufficient to describe the tempo characteristic of music because note

---

<sup>1</sup> [www.soundcloud.com](http://www.soundcloud.com)

passages can vary dramatically in the same tempo. For example, a long note and multiple short notes can be located in a single beat but they produce a different nuance. For this reason, we represent the tempo characteristic with the number of notes per second. Since we need to have overall trend of local note population rather than the exact number of notes for visualization, we use a simple onset detection algorithm which counts note onsets in a selected frame based on amplitude information.

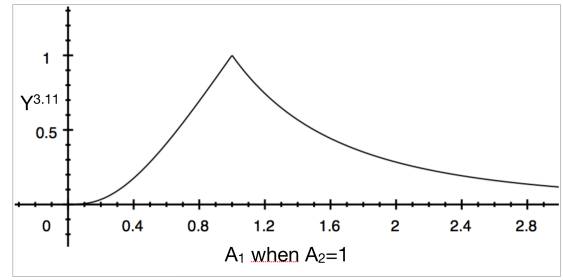
### 3.3 Auditory Roughness

Quantifying the harmonic feature of music from audio is typically carried out by chord recognition. However, recent work pointed out the limitation of automatic chord recognition [16]. Identifying chords can be arguable even for musicologists, especially for complex classical music. The research shows that the maximum agreement ratio between two chord annotations among four annotations on “I Saw Her Standing There” by *The Beatles* was only 65%. Also, classical music includes atonal music or late-romantic music like Wagner’s “Tristan und Isolde”. This makes hard to employ automatic chord recognition for classical music.

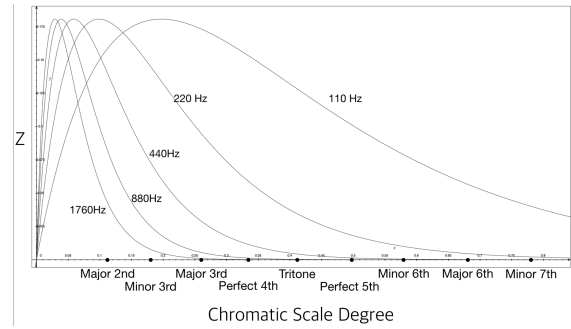
Instead, we use auditory roughness which can represent the tonality with a single value. It is a term used in the acoustics and psychoacoustics literature to describe buzzing sound quality that is produced by two sounds with different pitch that is distinguishable but close to each other like minor seconds interval. This feature is strongly associated with harmonic dissonance. For example, major seconds or minor thirds in a low register, which are usually avoided as dissonant intervals in the western musical tradition, makes high roughness. There are various models to calculate the auditory roughness quantitatively. Among others, we employ a model presented by Vassilakis [17] that uses two sinusoidal components with frequency  $f_1$  and  $f_2$  and amplitude  $A_1$  and  $A_2$ :

$$\begin{aligned}
 R &= X^{0.1} \times 0.5Y^{3.11} \times Z \\
 X &= A_{min} \times A_{max} \\
 Y &= \frac{2A_{min}}{A_{min} + A_{max}} \\
 Z &= e^{-b_1s(f_{max}-f_{min})} - e^{-b_2s(f_{max}-f_{min})} \\
 s &= \frac{0.24}{(s_1f_{min} + s_2)}
 \end{aligned} \tag{1}$$

where  $A_{min} = \min(A_1, A_2)$ ,  $A_{max} = \max(A_1, A_2)$ ,  $f_{min} = \min(f_1, f_2)$ ,  $f_{max} = \max(f_1, f_2)$ ,  $b_1 = 3.5$ ,  $b_2 = 5.7$ ,  $s_1 = 0.0207$ ,  $s_2 = 18.96$ . The term  $X$  represents the dependence of roughness on intensity. For better understanding of this equation, we illustrate how the term  $Y$  and  $Z$  change over the  $A_1$  and the frequency difference  $f_{max} - f_{min}$ , and  $f_{min}$ , respectively, in Figure 1 and Figure 2. They show that roughness is higher when the amplitude of two sin wave is similar, and the minimum frequency is lower. The roughness of complex sound can be calculated by summing the roughness of each combination of two sinusoidal components in the sound.



**Figure 1.** Change of roughness term  $Y$  by amplitude difference between two sinusoidal wave, where one amplitude is fixed to 1 and the other varies



**Figure 2.** Change of roughness term  $Z$  by frequency difference between two sinusoidal wave. The frequency difference is represented as  $f_{max}/f_{min}$

### 3.4 Visualization Scheme

As we mentioned above, representing music using a 2-D graph with x-axis in time is widely used mapping. Among three features, volume is the most accurate feature to calculate from audio data. Also it is the most dynamic feature. Therefore, we use the volume as a  $Y$  value in our visualization. For the other two features, a graph color and a background color are used as mapping targets. Since the auditory roughness is often correlated with the volume, mapping it to the graph color can be somewhat redundant. We thus map the onset density to a graph color and the auditory roughness to a background color.

## 4. IMPLEMENTATION

Our purpose is building an automatic music visualization system that can be easily utilized by general users. Previous visualization systems use symbolic data such as a MIDI file or pre-analyzed text data, which are not readily obtainable by listeners. Also, many of them ask users to install a stand-alone application, which might have some compatibility issue on user side or require some extra efforts. Considering these problems, we design our system such that, if users can access to audio content on a web browser, the visualization is immediately rendered from the audio file. We used web audio API which was developed for various audio applications under HTML5 specification. Therefore, our visualization system can be run on many web browsers such as Google Chrome and Mozilla

Firefox<sup>2</sup>.

#### 4.1 Feature Extraction

When a user loads an audio file on the system, the file is decoded to linear PCM data and saved as a buffer on the browser memory. Then, samples are segmented by a Hann window. The volume is calculated using root mean square of audio samples for each window. The algorithm for counting onsets uses local maxima of the calculated volume sequence. Specifically, it compares each volume value in the array with the next value. If it increases, the increased amount is saved. This is accumulated if the volume keeps increasing. If the volume decreases, the algorithm compares the accumulated amount and a threshold. If the accumulated amount is larger than the threshold, it is counted as an onset and the accumulated value is reset. Otherwise, only the accumulated amount is reset. Though it is not very sensitive for detecting note onsets in legato passages, it is sufficient for detecting overall onsets in the music.

To calculate the auditory roughness, we use a DSP library for fast offline FFT processing<sup>3</sup> and detect 50 peaks from the local maxima of the magnitude response. We then calculate the auditory roughness from every pair of the peaks and add them all.

The result of these three features are all normalized and scaled to the size of HTML canvas. Since the auditory roughness tends to be somewhat correlated with the volume, we make it up by dividing the auditory roughness by the volume with a constant value. This compensation can emphasize the dissonance in quiet passages.

#### 4.2 Visual Mapping

We visualize the features using a 2-D graph with a single continuous curve on an HTML canvas. The x-axis represents time and its width is fixed regardless of the length of input files. The y-axis represents the volume curve along with two color mappings. We downsample the features such that a set of values are mapped to a pixel by averaging. Onset density is mapped to the color of vertical lines below the volume curve. Lines with high onset density are colored with high saturated red on HSV scale. Auditory roughness is mapped to the color of vertical lines above the volume curve. Lines with high auditory roughness are colored with bright clear blue on RGB scale and so those with low auditory roughness is with dark dim blue.

#### 4.3 Output

Once the visualization is generated and shown on the screen, users can freely navigate the music through clicking on the visualization. It takes a mouse input and changes the playing offset of the music immediately. A progress bar shows the current playing position. Users can make a music flowgram of a very complicate contemporary orchestral work, for example, Salonen's Violin Concerto, which is about 29 minutes long as shown in Figure 3.

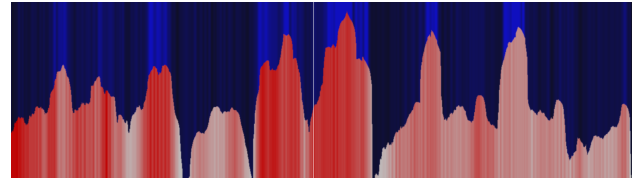


Figure 3. Music flowgram of Salonen's Violin Concerto

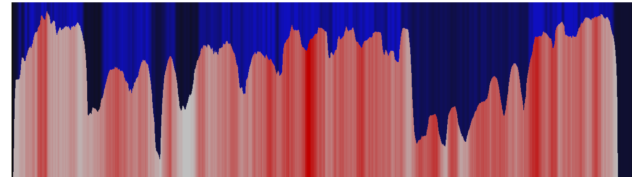


Figure 4. Music flowgram of Sibelius Finlandia

### 5. EVALUATION

To evaluate the effectiveness of music flowgram, we set up a user test for classical music listeners. The test consists of two scenarios that imitate situations where they listen to music on YouTube. The first case is listening to music without any video, which contain a static image or slide show of images such as an album cover or a picture of composer on YouTube. The second case is searching a specific part of a video when a short extracted audio clip is given.

The participants were 15 undergraduate students from Korea Advanced Institute of Science and Technology. All of them were a member of amateur orchestra, with a regular experience of listening to orchestral music. We divided them into two groups.

#### 5.1 User Test A: Listening and Recall

In our previous research [5], we found that those who listen to the music with a volume graph can identify an extracted audio clip better than those who listened to the same music without it. In this test, we extend it to a more active recall task. That is, we evaluate how much music flowgram will help a listener to concentrate on and memorize the music. After listening to a movement from a symphony, participants are asked to describe how music changed over time in an objective expression. This task requires much more accurate musical memory compared to previous research experiment, because participants need to recall the music without any audio cue.

To design the test environment to be more realistic, we used YouTube as a listening interface. We uploaded two videos for each music, one with a music flowgram and the other with an image of album cover<sup>4</sup>.

Participants listened to three music pieces and described them in three different situations: with an album cover, a music flowgram only while listening, and a music flowgram while listening and writing. Writing a note during the listening was not allowed. That is, participants had to describe music only with their memory or the music flowgram. The music for the tests were selected from rarely

<sup>2</sup> <http://jdasam.github.io/visualization/main.html>

<sup>3</sup> <https://github.com/corbanbrook/dsp.js/>

<sup>4</sup> <https://www.youtube.com/playlist?list=PLq7cRTjnYEi6fvwS3fSY1PsEDpNQdVVRXA>



Song (length)	RK1-1 (9:04)	RK1-4 (6:46)	B1-4 (7:18)
Group A	Album cover	Music flowgram (MF)	MF, maintained after the listening
Group B	Music flowgram	MF, maintained after the listening	Album cover

**Table 1.** The setting of the user test A

performed repertoire so that none of the participants had possibly listened to the piece before. In addition, we selected music with similar style to reduce the effect of difference between the music materials. Selected materials are Rimsky-Korsakovs Symphony No. 1, first movement (RK1-1) and fourth movement (RK1-4), and Borodins Symphony No. 1, fourth movement (B1-4). The detailed setting is shown in the Table 1.

To properly guide participants response, we provided an example paragraph like below:

*The piece starts with fanfare of trumpets. Then, cello plays quiet and smooth theme. Some variations of the theme are followed, and the dynamics get stronger. Then orchestra tutti play the main theme in faster tempo. Flute plays fast and virtuosic solo passages. The same orchestra tutti is followed. The key is transposed from major to minor, and woodwinds play march-like melody. This melody developed further by brass and violin in fortissimo. The fanfare from the beginning reappears in a more splendid way. Main theme is played again by cello. The orchestra tutti in the middle appears again with additional coda, which finishes the music. (Dvořák Symphony No. 8, fourth movement, translated into English by the first author)*

After the listening and recall test, participants are asked to score each visually represented feature, based on how well it represents the musical characteristics. Also, participants were asked about how the listening experience with the music flowgram was different from that without it.

## 5.2 User Test B: Searching an Excerpt

The second test was searching an excerpt of the music in the YouTube video for the purpose of validating that a music flowgram can help a listener to find a specific part more quickly.

We rendered a music flowgram from downloaded YouTube videos, and attached it below the YouTube player, as shown in Figure 5. The music flowgram image is linked to YouTube video so that users can select the playing position by clicking a specific position on the graph. We compared this setting to a YouTube player only page that contains the same video. Since the difficulty of searching task is largely influenced by a characteristic of the excerpt, we chose excerpts from movie scenes, rather than choosing arbitrarily. The selected movie is *Lorenzos Oil* and *Birdman*, which used a short clip from the third movement of Mahlers symphony No. 5 (M5-3), and the second movement of Rachmaninovs symphony No. 2 (R2-2), respectively.

Each participant watched the movie clips and searched the excerpted music parts on the YouTube video. There were two movies and a corresponding web page including YouTube videos of the music used in the movie clips. One



**Figure 5.** An example of the proposed system

Song (length)	M5-3 (18:05)	R2-2 (10:11)
Group A	YouTube with MF	YouTube only
Group B	YouTube only	YouTube with MF

**Table 2.** The setting of user test B

of the web page included a YouTube player with a music flowgram of the audio of the selected video, and the other page only included a YouTube player. We arranged a different setting for each group, as described in the Table 2.

## 6. RESULT AND DISCUSSION

### 6.1 A. Listening and Recall

We checked the descriptions written by participants and matched the description to the corresponding part in the music. Any phrase or sub-phrase that can specify a rehearsal letter from music was counted as a correct answer. Considering the participants are all non-professional musician and the description was written after only one listening attempt, we allowed a certain level of wrong descriptions, for example, incorrect instrument and melody identification. Many participants had confusions whether the melody was reoccurred or newly introduced, and whether the solo instrument was flute or oboe. But the confusion between string and wind instrument was not allowed. The paragraph below is an example of participants' answer, which is translated into English by the first author. Because Korean does not usually use a definite or indefinite article, melody is translated without an article. A letter in a parenthesis is annotation made by authors, which means a rehearsal letter of corresponding part in the score.

*Music starts with brass and bass(A). Main theme starts with bass and cellos, violas, violins takes over the theme and the pitch register goes higher(B). All the instruments play main theme in fortissimo(C). Clarinet plays melody and string plays melody(F). Then, the brass section is added and play majestic chord(G). The same pattern is repeated (Repetition). The flute plays melody and similar pattern is played(I). At the last part, flute and oboe appear(T). After majestic brass, timpani finish music (U).*

Song	RK1-1	RK1-4	B1-4
Group A (n=8)	6.75 (SD: 1.49)	6.88 (SD: 2.23)	8.38 (SD:2.45)
Group B (n=7)	9.14 (SD: 1.07)	6.86 (SD: 1.68)	6.71 (SD: 1.98)

**Table 3.** Average score and standard deviation of correct description in user test A

By this criteria, we scored how many parts of the music is described in each description. The rehearsal letters are referenced from an edition of Muzgiz, Soviet State music publishing house, for both symphonies. The data analysis was in blinded name to avoid bias.

The result on the Table 3 shows that the group with the music flowgram recalled more parts of the music than group with the album cover image, regardless of whether the music flowgram was provided until the end of the writing step. This result shows that the music flowgram can help a listener to memorize music more precisely as we expected.

There was almost no difference between two groups in the case of fourth movement of Rimsky-Korsakovs symphony, for which the only difference was the presence of the music flowgram during the writing step. From this result, we infer that the music flowgram was easy to remember and recall, so that there was almost no disadvantage of not watching it again while writing the description.

During the analysis, we found that the ratio of participants who mentioned the repetition of the piece was higher in the group with the music flowgram. Both the first movement of Rimsky-Korsakovs Symphony No. 1 and the last movement of Borodins Symphony No. 1 are in sonata-allegro form that includes a repetition of the exposition. Four out of seven participants of group B mentioned the repetition of the first movement of Rimsky-Korsakovs symphony, while only one out of eight participants mentioned in the other group. In the case of Borodins symphony, six out eight participants mentioned it in the group A, while two out of seven in the group B mentioned it.

Recognizing the repetition is important for understanding a structure of music. Most of musical forms in classical music include repetition of main part. This is one of the reasons why the former research about music structure focused on repetitive structure. The repetition of exposition is an important characteristic of a sonata-allegro form. This result also shows that our music flowgram is helpful for listening and understanding classical music.

Comments from participants, which is shown in the Table 4, also support the previous results. Many participants answered that the music flowgram was helpful for understanding the entire structure of music. There was only one participant who wrote negative comments only. Some of the comments contradicted each other; some participants answered that the listening was more interesting because he can anticipate, but the other answered that it was less interesting because it was easy to anticipate. Three out of fifteen participants answered that they preferred audio only listening environment. But they also commented that just taking a look at the visualization was helpful for under-

Positive comments	<b>It was helpful for...</b> understanding an entire structure of music. anticipating when will music get excited or relaxed. knowing length of each section. recognizing which part is repeating. memorizing change of dynamic. concentrating on music by comparing my own anticipation and the actual music
Negative comments	It was disturbing to concentrate on listening to the music. It made listening less interesting because it was easy to anticipate music. It made me keep thinking about how much the graph is accurate.

**Table 4.** Examples of comments on listening with music flowgram

Parameter	Average score (1 to 5)
Volume	4.6 (SD: 0.51)
Onset Density	3.4 (SD: 0.74)
Auditory roughness	1.9 (SD: 0.74)

**Table 5.** Evaluation for individual features by participants

standing the entire structure of the music.

Participants evaluation on each visual parameter is shown in the Table 5. Most of the participants gave the highest rate to loudness. The evaluation of the auditory roughness was not positive. This is partially because of its visual mapping to background color. Some of the participants pointed out that the change of background color was not easily recognizable.

## 6.2 B. Searching an Excerpt

The result of user test B is on Table 6. Because the task is largely dependent on participants pre-knowledge about the music material, we put answers from participants who know the material well separately in Table 7.

For both pieces, it took less time for the group with music flowgram to find the target part as we expected. In the case of Mahlers symphony, the excerpt was in the very last part of the music, which contains a loud brass section. So it was a tough task to find it by navigating from the beginning of the music. But when the music flowgram was provided, the participants could only concentrate on the loud part. This advantage greatly reduced the searching time, especially for the participants who do not know the piece very well.

On the other hand, the excerpt form Rachmaninovs symphony is in an early part of the music. So many participants can easily find the part by navigating without music flowgram. This is the reason why the difference between the two group is slight. But it is worth mentioning that the shortest record was only 6 seconds, which included the

Song	M5-3	R2-2
Group A (n=6)	170 seconds (SD: 138)	59 seconds (SD: 32.9)
Group B (n=6)	377 seconds (SD: 153)	40 seconds (SD: 29.6)

**Table 6.** Results of user test B (average of consumed time)

Song	M5-3	R2-2
Group A (n=2)	14 seconds 43 seconds	15 seconds 21 seconds
Group B (n=1)	85 seconds	6 seconds

**Table 7.** Results of user test B with participants who know well the material

loading time for the YouTube player. One of the participants found the excerpted part with a single click on the music flowgram. Since the excerpt contains legato passage of strings, the participant could easily find it by searching a part with low onset density.

## 7. CONCLUSION

In this paper, we have presented an automatic visualization method for representing music in its entirety. The goal of our visualization is showing how the music changes from beginning to end. The method visualizes music with three audio features like volume, onset density, and auditory roughness, which are highly associated with loudness, tempo, and dissonance, respectively, in musical characteristics. These features are visualized as a two-dimensional graph. We implemented the method on a web page using Web Audio API and conducted user test for verifying the usefulness of our method in the listening and searching task. The results showed that listening to music with a music flowgram helps listeners to memorize the music more precisely. A music flowgram was also helpful for searching a specific excerpt from music.

Despite of the overall positive results, there is still a large margin for improvement. Auditory roughness, which is intended for representing the harmonic characteristic of music, was not satisfactory for many participants. For the future work, we are planning to improve our algorithm for detecting onset and calculating audio roughness. We are also considering other audio features that can replace auditory roughness such as tonal complexity [18]. Another important challenge will be finding more intuitive and visually pleasing mappings for each parameter.

## 8. REFERENCES

- [1] M. Müller, F. Kurth, and T. Röder, “Towards an efficient algorithm for automatic score-to-audio synchronization,” in *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)*, 2004.
- [2] M. Wattenberg, “Arc diagrams: Visualizing structure in strings,” in *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*. IEEE, 2002, pp. 110–116.
- [3] J. Foote, “Visualizing music and audio using self-similarity,” in *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*. ACM, 1999, pp. 77–80.
- [4] M. Müller and N. Jiang, “A scape plot representation for visualizing repetitive structures of music recordings,” in *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*. Citeseer, 2012, pp. 97–102.
- [5] D. Jeong and Y. Noh, “Music visualization using volume graph and its effect on classical music listening,” 2014.
- [6] D. Jeong, “Music visualization using flow graph and its effect on listening to classical music,” Master’s thesis, Korea Advanced Institute of Science and Technology, 2015.
- [7] N. Kosugi, “Misual: music visualization based on acoustic data,” in *Proceedings of the 12th International Conference on Information Integration and Web-based Applications & Services*. ACM, 2010, pp. 609–616.
- [8] M. Goto, “Smartmusiciosk: Music listening station with chorus-search function,” in *Proceedings of the 16th annual ACM symposium on User interface software and technology*. ACM, 2003, pp. 31–40.
- [9] M. Goto, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano, “Songle: A web service for active music listening improved by user contributions,” in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*. Citeseer, 2011, pp. 311–316.
- [10] C. S. Sapp, “Harmonic visualizations of tonal music,” in *Proceedings of the International Computer Music Conference*, vol. 1001. Citeseer, 2001, pp. 423–430.
- [11] M. Malt and E. Jourdan, “Le «bstd»—une représentation graphique de la brillance et de lecart type spectral, comme possible représentation de levolution du timbre sonore.” in *Proceedings of L’analyse musicale aujourd’hui. Crise ou (r)’evolution ?*, 2009.
- [12] W.-Y. Chan, H. Qu, and W.-H. Mak, “Visualizing the semantic structure in classical music works,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 16, no. 1, pp. 161–173, 2010.
- [13] S.-L. Tan and M. E. Kelly, “Graphic representations of short musical compositions,” *Psychology of Music*, vol. 32, no. 2, pp. 191–212, 2004.
- [14] G. Freytag, *Die Technik des Dramas*. Hirzel, 1872.
- [15] A. Gabrielsson and E. Lindström, “The role of structure in the musical expression of emotions,” *Handbook of music and emotion: Theory, research, applications*, pp. 367–400, 2010.
- [16] E. J. Humphrey and J. P. Bello, “Four timely insights on automatic chord estimation,” in *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, 2015.

- [17] P. N. Vassilakis and K. Fitz, "Sra: A web-based research tool for spectral and roughness analysis of sound signals," in *Proceedings of the 4th Sound and Music Computing (SMC) Conference*, 2007, pp. 319–325.
- [18] C. Weiss and M. Muller, "Tonal complexity features for style classification of classical music," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on.* IEEE, 2015, pp. 688–692.