Music learning through visualization

Conference Paper · February 2002

DOI: 10.1109/WDM.2002.1176199 · Source: IEEE Xplore

Rumi Hiraga
National University Corporation of Tsukuba University of Technology

Issei Fujishiro
Keio University, Yokohama, Japan

Music Learning through Visualization

Rumi HIRAGA
Bunkyo University,
Department of Information Sciences
rhiraga@shonan.bunkyo.ac.jp

Fumiko WATANABE
Ochanomizu University,
Graduate School of Humanities and Sciences
fumiko@imv.is.ocha.ac.jp

Issei FUJISHIRO
Ochanomizu University,
Graduate School of Humanities and Sciences
fuji@is.ocha.ac.jp

Abstract

We have developed a prototype music visualization system that shows the whole piece in a computer window, however long the piece and no matter how many parts it includes, as well as selected part. The selected part is presented using either performed data or staff notation. Users can better understand the performance expression by visualizing the performance with expressive cues of the qualitative music terms. Based on this understanding, users can exchange with other users their comments, recorded on visualized figures, through the Internet. Though still a prototype, we expect to make our system for one that will enable users to better understand a musical piece and its performance, especially for a cooperative performance.

Keywords: Viewing tool, Visualization, Cooperative work, Learning assistance

1 Introduction

New technologies related to computer music make it possible to not only create new types of music by ourselves, but also to learn conventional music more easily. We propose a music system for assisting music learning that uses a new type of interface and a function. It helps the user to visualize music—both musical scores and their performance. Target users are those who want to improve their understanding of a musical piece and their performance of it, especially for a cooperative performance. When working on a musical piece, such as a piano piece for four-hand, chamber music, jazz quartet, and orchestra, without other people, users can use our system in several ways: For example, they can individually deepen their understanding of the piece through visualized figures of the musical score and their performance. They can share opinions and suggestions by exchanging their messages attached to the visualized score and performance through the Internet.

Consider a case in which four people are going to perform as a quartet at a festival. They are not musical novices but they are too busy to spend much time practicing together. They are given the data of their performance at some point and listen to it at home. Using our system, which provides an environment for visualizing music both the conventional staff notation and the performance, they can visualize music as well as listen to their performance of it. They can then analyze their performance, put their comments about it on a figure on the system, and send their comments to the other members.

Our system consists of the two parts—one for browsing and administering of sound sources and one for showing performance expression. Both parts are based on our use of desktop music (DTM) software. For browsing of score on a PC, our system focuses on the conceptual design of the spatial substrate, the attributes of the environment, and the related level-of-detail operations. In contrast to DTM systems that show MIDI parameter values on each note, our system shows the performance expression appeared in the move of music, i.e., not by the direct mapping of MIDI parameter values to a picture but by using a more abstract level of music data: the dynamics change, articulation, and local tempo, namely the expressive cues. We have developed a prototype of this system.

In Section 2, we describe related work in the areas of music visualization and information visualization. We also describe some of the problems with using current DTM software in terms of visualization. In Section 3, after briefly describing our system, we describe score visualization and performance visualization in detail. Then in Section 4, we describe the user interface, which is a new way of visualizing digital scores and performance expression. User comments on the prototype's performance are also given. Finally in Section 5, we summarize the key points and describe our future work.

2 Music Visualization

In a broader sense, music has been combined with visual scenes for a long time especially in artistic areas. Operas, ballets, musical dramas, and movies (such as Fantasia) can be categorized as music with visualization effects. Several contemporary composers have invented their own forms of notation to describe music [10]. Conversely, there have been several trials to explain painting masterpieces by using sonification. Music visualization of artistic forms enables people to interpret a figure in several ways.

The relationship between a musician's physical movements and the music generated has been researched on especially in terms of music perception. In this case, computer systems do not play a role in help visualizing this relationship.

With the aim of facilitating music learning, we focused on ways to visualize a score and its performance, the accessibility of information, and the usability of the system.

2.1 Related work – music visualization

When "music visualization" is simply visualizing music using a computer graphics and not for artistic purposes, visualization of sound is already indispensable in research and applications such

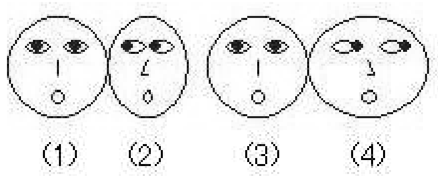

Figure 1: Sample faces

as sound analysis and synthesis. Besides showing that sound as a wave shape, Pickover showed an interesting and impressive symmetrical figure of sound data [17]. Sobieczky visualized the consonance of a chord on a diagram based on a roughness curve [21].

There have been some research on music visualization, but not very much. There are two types of music visualization.

- augmented score visualization

  Conventional staff notation is limited because it does not necessarily represent all of the composer's expressive intentions. Oppenheim proposed a tool for representing the composers' expressive intentions [16]. Kunze proposed defining several figures in three dimensions for composers to use in composing their works [12].

- performance visualization

  Hiraga proposed using simple figures to help users analyze musical performances [9]. Smith proposed a mapping from MIDI parameters to 3D graphics [20]. Foote's checkerboard-type figure [4] shows the resemblance among performed notes based on the data of a musical performance. Watanabe proposed a system with a unified 3D interactive interface both for browsing and editing sound data [22]. Miyazaki's *comp-i* tool with a 3D performance-visualization interface enables users to generate music using a rich set of functions [15].

Our system provides an interface for an augmented score (though not for composers) and performance visualization for understanding a performance.

2.2 Related work – information visualization

We benefited from the research on information visualization in our efforts to help users visualize scores and their performances. As will be described in the next section, one of the features of our score browsing is to enable users to use more of the display resource to correspond to interest of the user's attention— called $focus + context$ [2].

There are several ways of viewing a particular data from a large quantity of data; Fisheye views [5] [19], Perspective Wall [13], LensBar [14] and Table Lens [18] are able to generate a small display of a large structure and and enable users to use a $focus + context$ function to display a particular part while maintaining the relationship of the part to the whole.

Our way of displaying the score is like Fisheye views.

2.3 Problems with music visualization from the user interface point of view

The user interface of commercial DTM software provides users several types of visualized music. Sequence software usually includes staff notation and a piano-roll figure, as well as a table of MIDI events. An introductory DTM system shows icons of musical instruments and their players. Although DTM systems are now widely used, both for generating performance and a score, they have some problems from user interface point of view.

- Users are unable to see the whole piece at one time. It is natural for users to start with the whole and then move down to the details of a piece, as they do in the real world: they open a score book, turn to a page, find a phrase, and take a closer look at a melody.

  Even with a the digitized score generated by Finale, searching for specific measures is accomplished either by specifying the measure number or scrolling down a window. For a full score for an orchestra, which includes many parts, users either have to look at musical symbols in small print if they want to see everything at one time, or give up displaying some parts if they want to see other parts with readable symbols.

- The visualized figures of a performance are shown by a direct mapping from the attribute values of the MIDI data (velocity, onset time, and pitch, for example). A piano-roll figure is an example of such a mapping. We understand music not at the musical abstract level of MIDI attributes, but rather by the relative relationship among notes (horizontal relationship), parts (vertical relationship), and the whole and parts (two-dimensional relationship).

  The MIDI parameters in DTM systems are handled individually; the user interface does not take the musical structure into consideration, which means the musical expression is not shown on figures.

- The information about a performance is visualized each channel by channel, not several channels at one time. This make it difficult to compare the performance of several channels at once when generating a performance by several instruments.

  Sequence software has a track view window in which several channels are shown with their melodic outline, but not their other musical attributes such as the change in dynamics, articulation, or tempo. Furthermore, it is impossible to see these musical attributes for two or more channels, which would enable comparing the performance in the other windows.

3 Proposed System

The system we propose can be used by the players of a piece to deepen their understanding of their performance and to exchange comments for improving it. The system partially solves the problems listed in the previous section.

3.1 Description

As described in Section 1, our system aimed at helping users learn a piece and improve their performance. The target users thus do not include music novices. Although the system can be used by

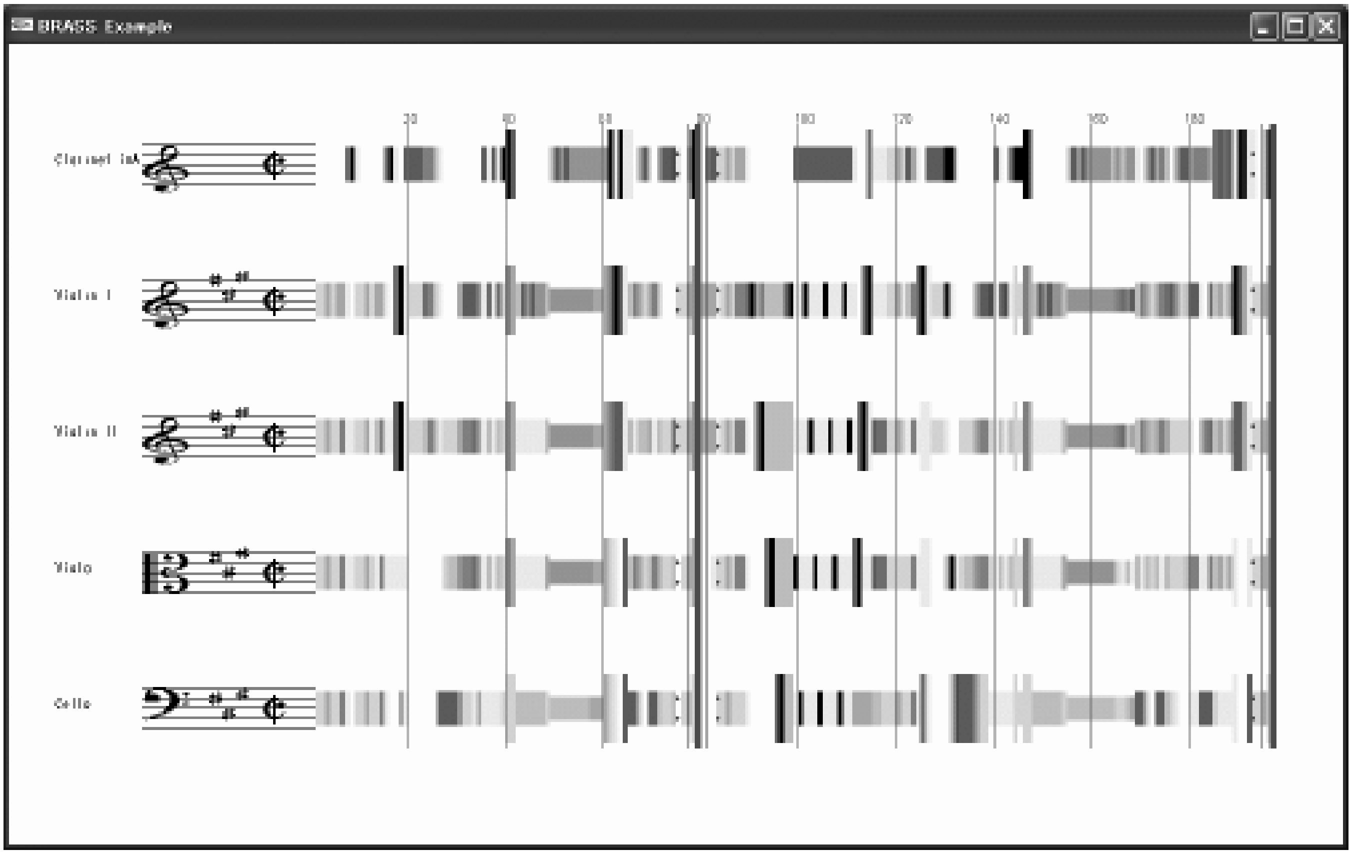

Figure 2: Overview of a score ("Clarinet Quintet A-major, K. V. 581" by W. A. Mozart)

a single user, such as a pianist, it will be more beneficial to users who are going to perform a musical piece together.

The system combines the results of research in two areas: BRASS (BRowsing and Administration of Sound Sources) by Watanabe and Fujishiro [23] and visualizing performance using Chernoff faces by Hiraga [8]. The former is used for visualizing the score and the latter for visualizing the performance respectively.

The relationships among parts are presented in the visualization in which the whole score is compressed; the details are examined using $focus + context$. Relationships among notes are expressed in terms of expressive cues. They are described using qualitative music terms (tempo, sound level, articulation, tone onsets and decays, timbre, IOI (Inter Onset Interval) deviations, vibrato, and final ritardando) which are quantifiable with performance data. Gabrielsson reported that listeners find basic emotions (anger, sadness, happiness, and fear) in human performances by intentionally controlling expressive cues [6]. Bresin showed a makeup experiment with his performance rendering system [1].

Based on their understanding of the performance with the score, users can place comments on a visualized figure and exchange them with other users in order to share and deepen their interpretations of a musical piece. An archive of such comments among players or between a tutor and a student would reveal the progress in improving the performance.

Because of this development progress, each of the two visualization programs and the message-exchange function use independent data sets. The score is shown using the ETF format of Finale. The performance is shown using MIDI data. The messages are stored in a file. Though there are not many studios for MIDI recording of a performance using various instruments and there is not a convenient and simple tool for extracting MIDI data from audio data, the system is even now applicable to piano duos.

3.2 Visualizing a score

Our system visualizes a score by using a spatial substrate and visual properties. The substrate is the framework for spatial information representing a musical structure with staves, notes, vertical bars, and brief marks (e.g., the *D. C.*). The visual properties represent the performance indicators on a score such as for the dynamics and the signs of *glissando*, *tenuto*, or *fermata*. They are categorized according to the range of their effects–on a note or on a phrase. Symbols on a note are, for example, *staccato* and *marcato*. *Fermata* is distinguished because it is applied to all parts at the same time. *Legato* and *leggiero* are effective on a phrase. The tempo indicator is usually applied to all parts.

The compressed figures are shown as follows:

1. spatial substrate
    - A staff is represented as a line with width.
    - Clefs are shown at the beginning of the score.
    - Symbols for rhythm and tonality are shown at the beginning and changing points.

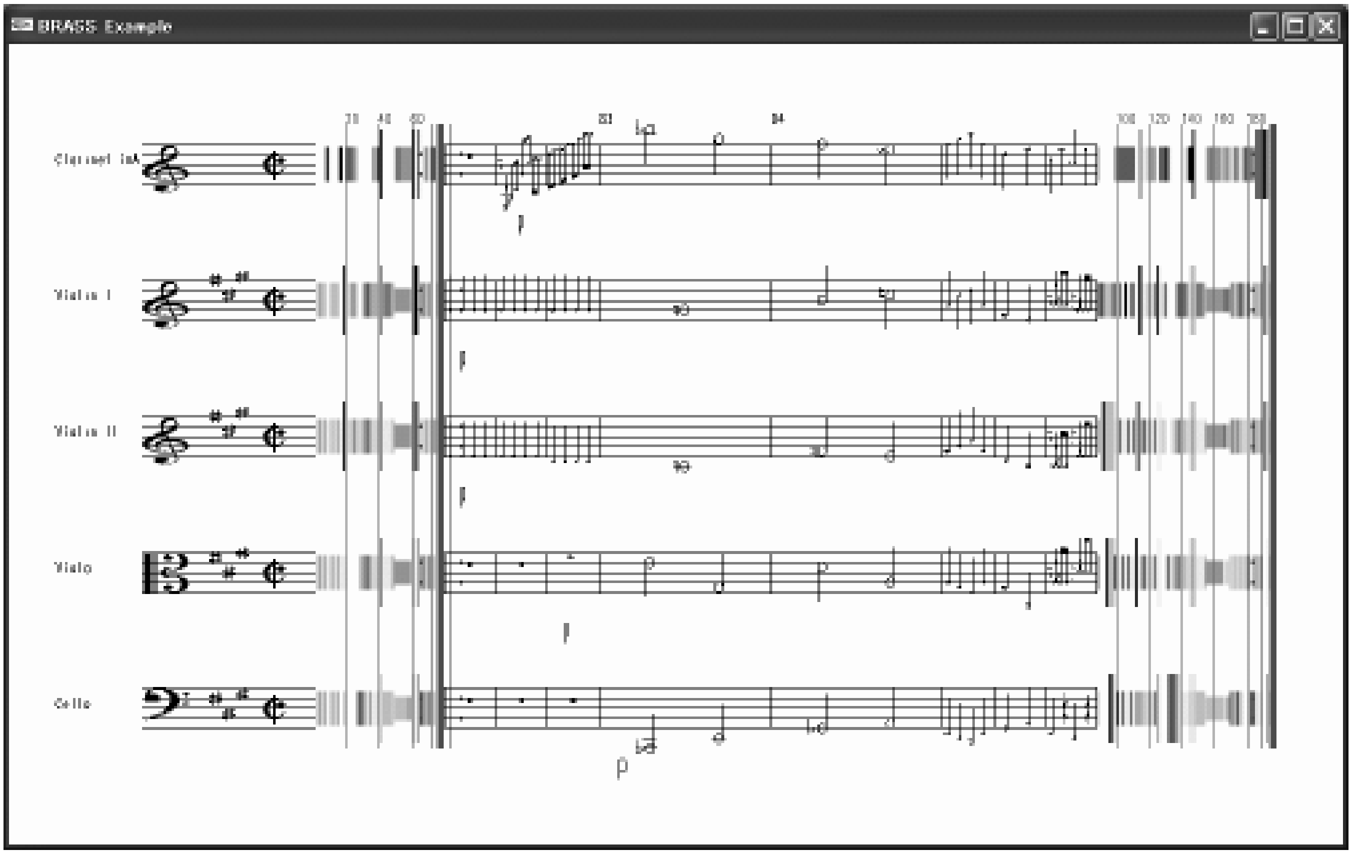

Figure 3: Focus + Context: a score

- Each note is not shown. The number of notes in a measure defines the brightness of the line.

2. visual properties

- Dynamics are shown on the width of each line.
- Tempo is not shown on the figure. We will use either the texture or the brightness.
- *Fermata*, *legato*, and the other specifications for each instrument are shown by glyphs [11].
- Specification of the instrument is shown at the beginning of the score.

The score is compressed horizontally; it is also compressed vertically if there are too many parts.

3.3 Visualizing a performance

The performance corresponding to part of the score is visualized by placing expressive cues on Chernoff faces [3]. Chernoff, a statistician, proposed using computer-drawn faces to visualize data for many dimensions. Each face corresponds to a note whose tempo, articulation, and sound level are visualized.

One of the biggest advantage of using Chernoff face is that the dimensions can be spontaneously extended. Three of the proposed expressive cues can now be shown on a Chernoff face, while more and more will be able to be shown in the future. In contrast, intuitive extension of 3D computer graphics cannot be used to show many dimensions.

The sample faces in Figure 1 represent expressive cues. We modified the shape of parts of the original Chernoff faces and omitted other parts (eyebrows, for example). Each face corresponds to a note whose tempo, articulation, and sound level are visualized. The performed note sequence is from left to right.

Local tempo, calculated as the IOI normalized by a note value[1], is mapped at the position of the eyeballs. The eyes look back (Figure 1 (2)) if the tempo is slower and look forward (Figure 1 (4)) if it is faster. The more extreme the eye movements, the more the tempo deviates from the one specified on the score.

The contour of the face and the shape of the mouth both represent articulation. Articulation is calculated as the overlap/detachment of two consecutive notes divided by a note value. If a note is played somewhat in *staccato*, the mouth is narrow (Figure 1 (2)), while in *legato* it is round (Figure 1 (4)).

The value of the sound level can be calculated in several ways. Though the sound level is described as the velocity value in MIDI parameters, the MIDI value does not correspond to a specific decibel level. Currently, we express the sound level as the difference between two consecutive notes. If a note is played more strongly than the previous one, the nose is drawn as ’<’ (Figure 1 (2)), at the same velocity ’|’ (Figure 1 (3)), and softer ’>’ (Figure 1 (4)). Since the first note has no sound level to compare to, the first nose

[1] The length of a note specified on a score.

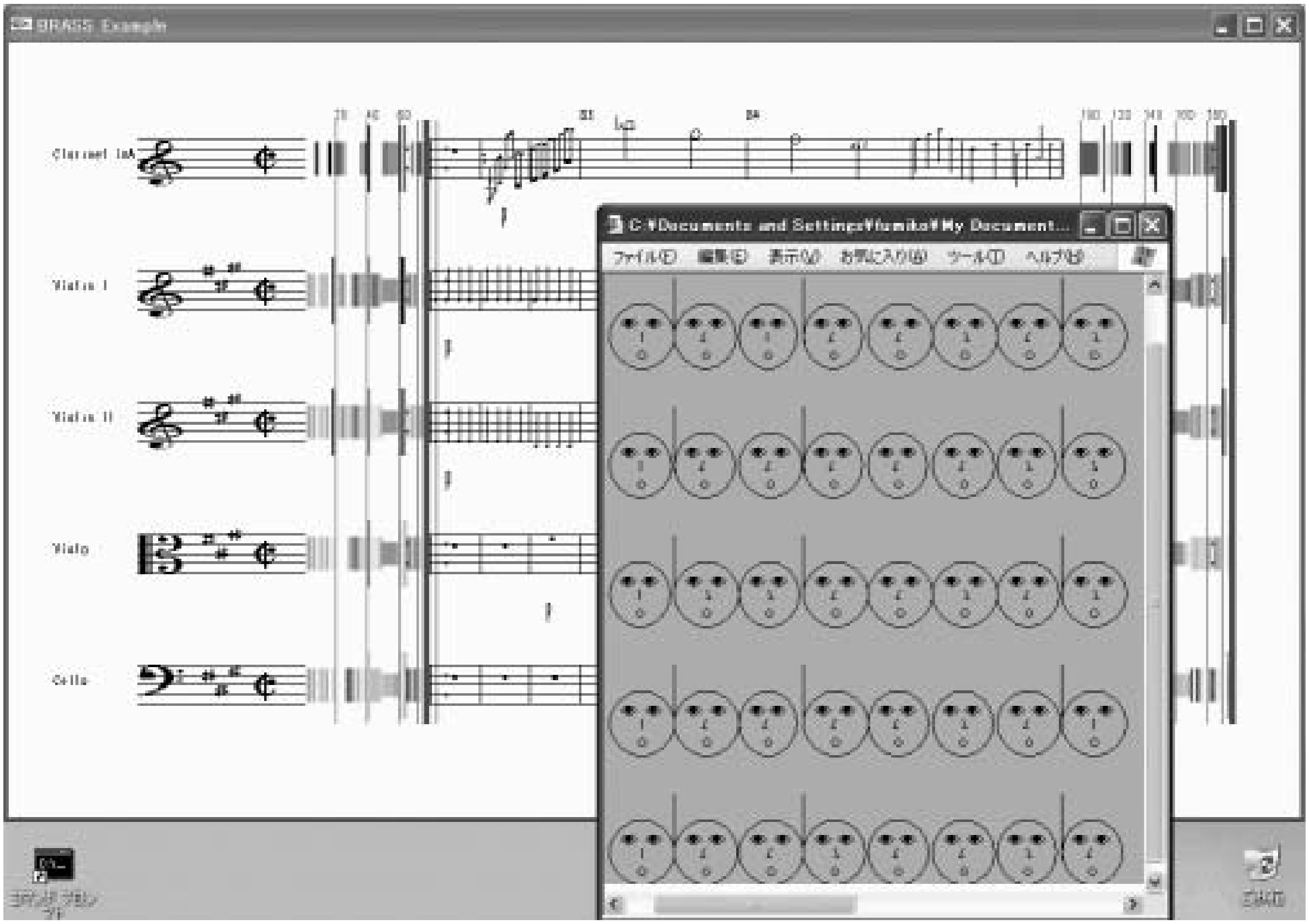

Figure 4: Focus + Context: performance

is always '|'. Faces (1) and (3) represent notes that are played with fidelity along a score with local tempo and articulation.

The visualized performance in Figure 1 is artificially made. By merely looking at the figure we can see that the performance starts softer and lighter, and ends lengthy in this phrase. In this way, local tempo and articulation show the deviation of a performance from the score.

A program BRASS is written in C++ with OpenGL. The score is displayed by BRASS using the input file in ETF format generated by Finale. The Chernoff face is written in Java 2 SDK, and it is invoked by BRASS.

4 Example

Using the first movement, consisting of 197 measures, of "Clarinet Quintet A-major, K. V. 581" by W. A. Mozart as a sample, we will explain the figures and functions in our system.

4.1 Visualizing "Clarinet Quintet" by W. A. Mozart

The user interface provides the full score in a window. Figure 2 shows the piece. Some of the vertical bars, such as repeat and the double bar, are shown in different colors on the figure since they show musical structures. Though it is not clear in the figure, there is a double bar and a repeat at the end of the exposition (the eightieth measure); they are in red.

Users can set the focus to be either vertical or horizontal. Reducing the number of parts shown realizes a vertical focus. If a place is selected for closer examination, it is magnified in the same window in which the uncompressed score for the 83rd and the 84th measures appears (Figure 3). This is the development of the first movement in which the clarinetist plays the first motif in the exposition. Since the neighbors of the selected place include the next most important information, they are visualized with a lower compression rate than the selected place. This method of visualization in which a selected section appears as part of the whole space lightens the user's psychological load imposed by visual perception and system operation, enabling the user to devote his or her attention to the music.

Performance for the selected section is shown on the score (Figure 4). The vertical bars show the measure bars. From the upper to lower in the window, the performance for each part is shown. The difference in performance expression between two parts can be visualized using the same shape. In case of showing a performance by multiple instruments or persons, we have to decide which notes to visualize.

From the 83rd to the 86th measures, almost all the parts have

Figure 5: Sample score: Which onset timing to show as performance visualization to compare several parts?

the same rhythmic structure, i.e., the onset timing of all parts has the simultaneity. In this case, the performance for all notes on the score is as shown in Figure 4. The exception is the 83rd measure played by the first and second violinists, during which the other three parts are playing using two half-notes while they play one whole-note. In Figure 4, we show the expression of the first note in all parts for this measure. Though not shown, a way to visualize which onset timing to show needs consideration such as in the 87th and the 88th measures (Figure 5). Onset timing of the five parts never happens simultaneously. How to treat the *trill* is another issue. Its treatment affects the way to visualize the difference in expression between two parts.

We used the performance of the piece in the “RWC Music Database: Classical Music” [7]. This database gives researchers free research use of the performance data. It provides audio and SMF data. Since the SMF for this piece was recorded uniformly as for tempo, eyeballs of Chernoff faces in Figure 4 do not move at all.

Figure 6 shows the score with inserted comments. When a position is selected, any comments at that position become readable. The player of the violoncello’s comment is shown (Figure 7). For the focused place, he or she suggests to play better in articulation, since the performance figure in Figure 4 shows the dispersion of articulation in five parts, in spite of the similarity in the move of their melodies and rhythms.

4.2 User comments

We have made an initial test of the system using two users. Comments from 1 to 3 below were by a professional musician with more than ten years of sequencer programming experience (creating musical performance using sequence software). The fourth was by an amateur musician who has studied music theory.

1. Currently only one place can be selected. Selecting two or more places at once would be useful in editing music.
2. When displaying a score with many parts, the track view of the conventional sequence software may be a hint.
3. A replay function is desirable.
4. Visualizing the musical structure helps in understanding the performance.

We will take these comments into consideration as we work to enhance the system.

5 Conclusion

We have described our prototype visualization system for learning music and for working cooperatively to perform a piece of music. The system provides users a new way of accessing musical information related to both score and performance expression. No matter how long a musical piece, the whole piece is shown in a window, giving users better access to places of interest. After selecting a place to focus on, they can exchange messages with each other to deepen their understanding of the music, which they can then share with the other performers.

We plan to make the following improvements.

- Score visualization
  - It is natural to compare two or more places since a musical piece often involves repeated phrases. To make this possible with our system, we have to devise a variation of fisheye views to show multiple foci.
  - The comment given by the user is valuable when we prepare a score for a full orchestra. We should also consider using a software system, such as SharpEye Music Reader, to convert from a scanned score image to a Finale file format.
  - Since we use MIDI in our system, it is possible to replay the performance, we are considering using XML for music representation and its related tools including replay a performance.
- Performance visualization
  - When visualizing the performances for multiple parts, deciding on the expression for which timing to show is a problem. We thus have to consider which information to use for comparing performances.
  - Though we visualize all notes uniformly in the current Chernoff face figures, each note has a different role in a music. By introducing some music theory, information of the role of each note will be used in visualizing performance.
  - We will focus on grouping information based on performance data, such as *accelerando* or *diminuendo*, clearly on the figures to enhance understanding of the performance.
  - Another challenge is to introduce continuous data, such as that of a half-pedal on a discrete note icons. Since grouped information also describes a continuous change over a range, this is an inevitable problem.
  - Using audio data for performance visualization is another issue. To use performance data that are closer to what we hear, we may have to handle audio data.

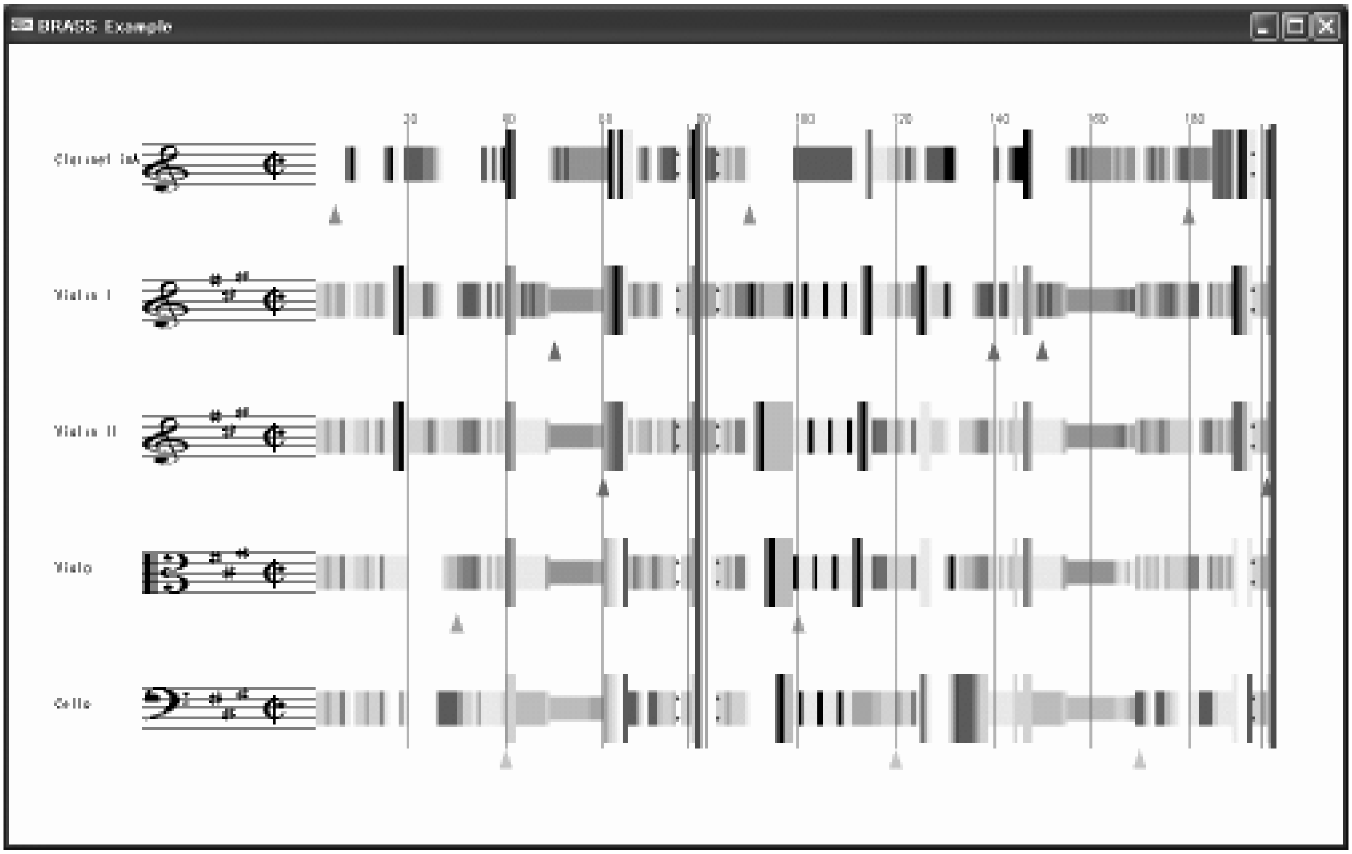

Figure 6: Inserted comments (Overview)

- Data set integration
  - When integrating the data set for our system, we will consider using XML for music representation. Using WEDELMUSIC and we will extend our system in a systematic way besides data integration; to include replay, to retrieve the similar places in a piece by analyzing the melody, and to manage a large quantity of data. These improvements will help users to understand music.

Acknowledgements

This research is supported by The Ministry of Education, Culture, Sports, Science and Technology through a Grant-in-Aid for Exploratory Research.

References

[1] R. Bresin and A. Friberg. Emotional coloring of computer-controlled music performances. *Computer Music Journal*, 24(4):44–63, 2000.

[2] S. Card, J. Mackinlay, and B. Shneiderman (edt.). *Readings in Information Visualization, using vision to think*. Morgan Kaufmann, San Francisco, California, 1999.

[3] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *J. Amer. Statist. Assoc.*, 68:361–367, 1968.

[4] J. Foote. Visualizing music and audio using self-similarity. In *ACM Multimedia99*, pages 77–80. ACM, 1999.

[5] G. W. Furnas. Genaralized fisheye views. In *Proc. of ACM Conf. on Human Factors in Comp. Sys.*, pages 16–23. ACM, 86.

[6] A. Gabrielsson and P. Juslin. Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, 24:68–91, 1996.

[7] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. Rwc music database: Classical music database and jazz music database. In *2002-MUS-44*, pages 25–32. IPSJ, 2002.

[8] R. Hiraga. Case study: A look of performance. In *Proceedings of IEEE Visualization*. IEEE, 2002.

[9] R. Hiraga, S. Igarashi, and Y. Matsuura. Visualized music expression in an object-oriented environment. In *Proc. of ICMC*, pages 483–486. ICMA, 1996.

[10] E. Karkoschka. *Das Schriftbild der Neuen Musik*. Moeck Verlag, Germany, 1966.

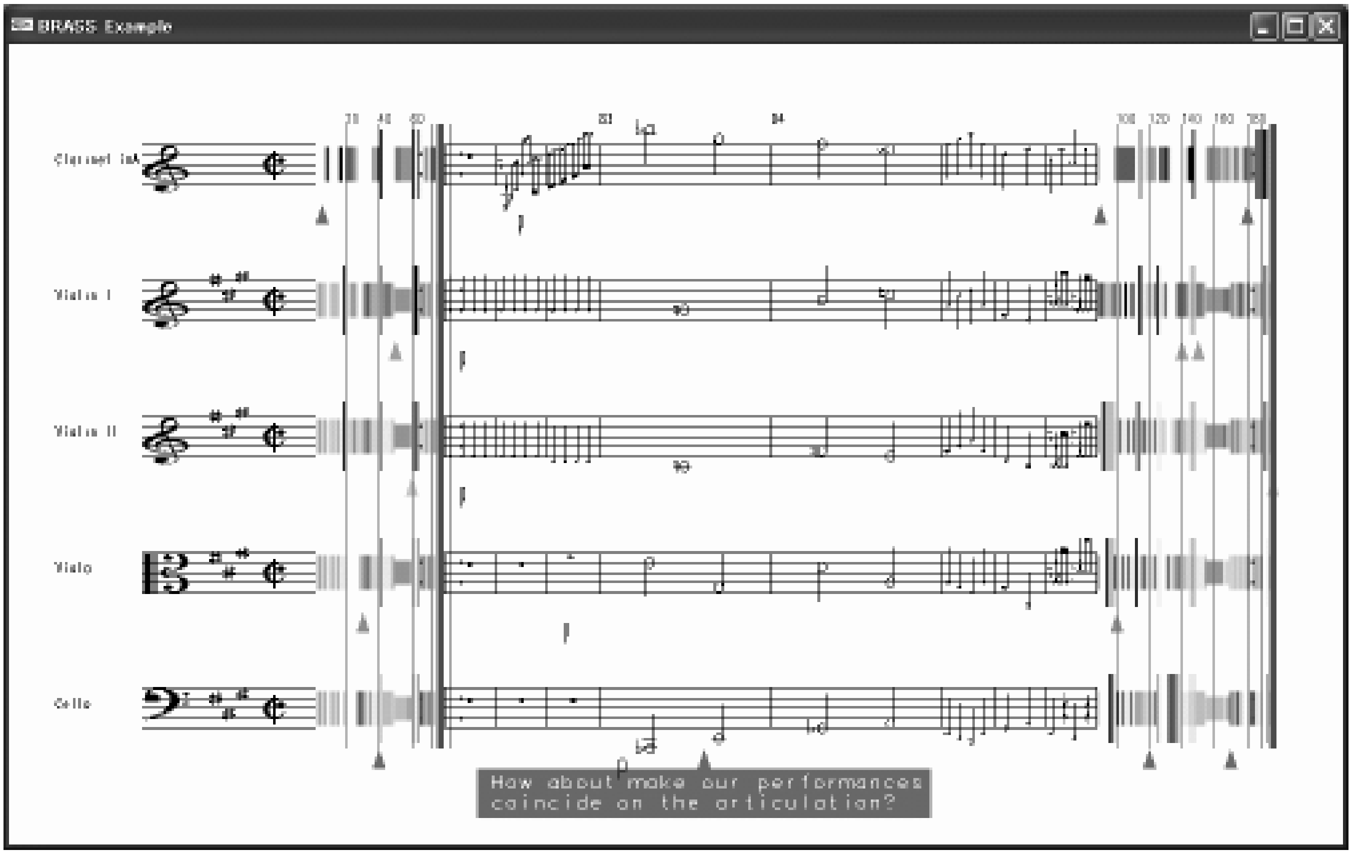

Figure 7: Inserted comment (Focus)

[11] P. R. Keller and M. M. Keller. *Visual Cues: Practical Data Visualization*.

[12] T. Kunze and H. Taube. See: A structured event editor – visualizing compositional data in common music. In *Proc. of ICMC*, pages 63–66. ICMA, 1996.

[13] J. D. Mackinlay, G. G. Robertson, and S. K. Card. The perspective wall: Detail and context smoothly integrated. In *Proc. of ACM Conf. on Human Factors in Comp. Sys.*, pages 173–179. ACM, 1991.

[14] T. Masui. Lensbar - visualization for browsing and filtering large lists of data. In *Proceedings of InfoVis'98*, pages 113–120, 1998.

[15] R. Miyazaki and I. Fujishiro. comp-i: 3d visualization of midi datasets. In *2002-MUS-46*, pages 55–60. IPSJ, 2002.

[16] D. Oppenheim. Compositional tools for adding expression to music. In *Proc. of ICMC*, pages 223–226. ICMA, 1992.

[17] C. Pickover. On the use of computer generated symmetrized dot-patterns for the visual characterization of speech waveforms and other sample data. *J. Acoust. Soc. Am.*, 80(3):955–960, 1980.

[18] R. Rao and S. K. Card. Exploring large tables with the table lens. In *Proc. of ACM Conf. on Human Factors in Comp. Sys.*, pages 403–404. ACM, 1995.

[19] M. Sarkar and M. H. Brown. Graphical fisheye views of graph. In *Proc. of ACM Conf. on Human Factors in Comp. Sys.*, pages 83–91. ACM, 1992.

[20] S. M. Smith and G. N. Williams. A visualization of music. In *Proceedings of IEEE Visualization*. IEEE, 1997.

[21] F. Sobieczky. Visualization of roughness in musical consonance. In *Proceedings of IEEE Visualization*. IEEE, 1996.

[22] A. Watanabe and I. Fujishiro. tutti: A 3d interactive interface for browsing and editing sound data. In *The 9th Workshop on Interactive System and Software*. Japan Society for Software Science and Technology, 2001.

[23] F. Watanabe, I. Fujishiro, and R. Hiraga. Brass –an interactive interface for supporting score reading–. In *2002-MUS-46*, pages 49–54. IPSJ, 2002.