comp-i:
A System for Visual Exploration and Editing of MIDI Datasets

R. Miyazaki, I. Fujishiro, R. Hiraga †
Graduate School of Humanities and Sciences, Ochanomizu University, Japan
reiko@imv.is.ocha.ac.jp, fuji@is.ocha.ac.jp
† Faculty of Information & Communications, Bunkyo University, Japan
rhiraga@shonan.bunkyo.ac.jp

Abstract

Conventional sequence software systems, which are commonly used to edit MIDI-encoded music, possess two kinds of problems due to interactions with MIDI data through multiple independent windows. We address the problems by developing a system, called comp-i (Comprehensible MIDI Player - Interactive), which provides music composers and arrangers with a novel type of 3D interactive virtual space, where the users are allowed to explore global music structures and to edit local features, both are embedded in a time-series of multichannel asynchronous events of MIDI datasets while keeping their cognitive maps.

1 Introduction

A MIDI (Musical Instrument Digital Interface) dataset contains a multi-channel sequence of note events to represent the control information for musical composition. When editing and playing a MIDI dataset, users usually rely on so called sequence software systems, that offer a variety of editwindows such as staff, piano-roll and track view for different tasks. However, some users, including musical composers and arrangers, might want to capture the entire musical structure for their intensive work. To that end, the sequence software systems have the following two problems.

1. They allow the users to interact with MIDI datasets only through specific task-oriented editwindows. Since a limited number of related parameters appear in each of the editwindows, the users are forced to open multiple windows of small sizes side by side for accomplishing their complicated tasks.
2. Each of the editwindows displays only a short section of the given musical piece, so that this makes users difficult to grasp the global musical structures.

To visualize a particular piece of music intuitively and/or to increase the number of parameters presented in a single window, several 3D music visualization techniques have been proposed so far. Smith and Williams (1997) visualized a MIDI dataset including pitch, volume and timbre in a 3D virtual space. Kaper and Tipei (1998) proposed another 3D virtual space that is more immersive than the one in (Smith and Williams 1997). A virtual space presented by Kunze and Taube (1996) strives to visualize a musical structures along the time axis. However, all these systems show only a short section like editwindows offered by the sequence software systems do, and thus there still remains a difficulty to grasp the global musical structures.

We have developed a system, called *comp-i (Comprehensible MIDI Player - Interactive)*, which provides a sophisticated 3D virtual space, where the users are allowed to perform visual exploration and editing of a given MIDI dataset in an immersive and intuitive manner. Interim reports can be found in our early articles (Miyazaki and Fujishiro 2002; Hiraga, Miyazaki, and Fujishiro 2002; Miyazaki, Fujishiro, and Hiraga 2003). We have carefully designed two types of spatial substrates, called *Timeline Space* and *Structure Space*, for organizing the comp-i virtual space, along with a rich set of operations according to Shneiderman's Visual Information Seeking Mantra (Shneiderman 1998). One of the key features of the comp-i system is to offer a 3D animation-based focus + context mechanism (Card, Mackinlay, and Shneiderman 1999), which can make the users keep their cognitive map during their work.

The remainder of this paper is organized as follows. In the next section, we clarify our aim and give an overview of the comp-i system, with special focus on its five major functions. In Section 3, we show the organization of the virtual space and visualization techniques deployed in the system. In Section 4, we describe how to use the operations to perform their information seeking and editing tasks. In Section 5, the comp-i system is evaluated in comparison with a representative sequence software system and the existing 3D music visualization systems. Finally, in Section 6, we conclude this paper with a few remarks on future research directions.

2 System Overview

The aim of the comp-i system is to construct a visual exploration environment which enables the users to grasp the global musical structures of a given MIDI dataset and to edit events in the dataset while keeping their cognitive maps (Spence 2001) with animated focus + context views.

Figure 1 shows the relationship between the aim and major tasks executed with the comp-i system. The two extreme levels of functionalities treated in this system are to allow the users to "grasp the global structures" and to "edit MIDI events of interest". The two functionalities should interact with each other in a single virtual space, and the system connects them seamlessly through five tasks, that is, defining musical structures, information seeking, playback, generating virtual sound space and editing.

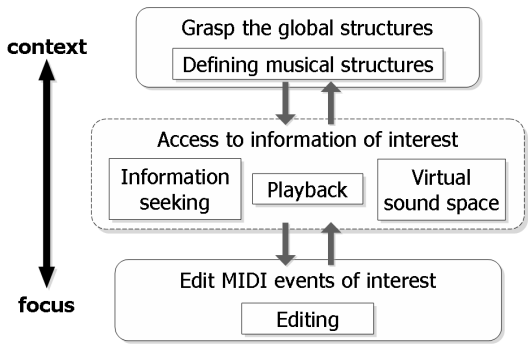


Figure 1: The relationship between the aim and the tasks with the comp-i system.

Defining musical structure. "Musical structures" embedded in a piece can be analyzed in several ways. The results of a score analysis, such as GTTM (Generative Theory of Tonal Music) (Lerdahl and Jackendoff 1983), are different from each other according to applications. The results of performance analysis are not always the same as those of score analysis. The notion of musical structure, that a composer or an editor has wanted to suggest, is not necessarily perceived identically by players and listeners.

Therefore comp-i takes an approach which enables the users to define their musical structures as they like. The users who compose or edit a piece are allowed to split the piece into several parts, and the users who playback the piece are allowed to take a look at the visualized structures.

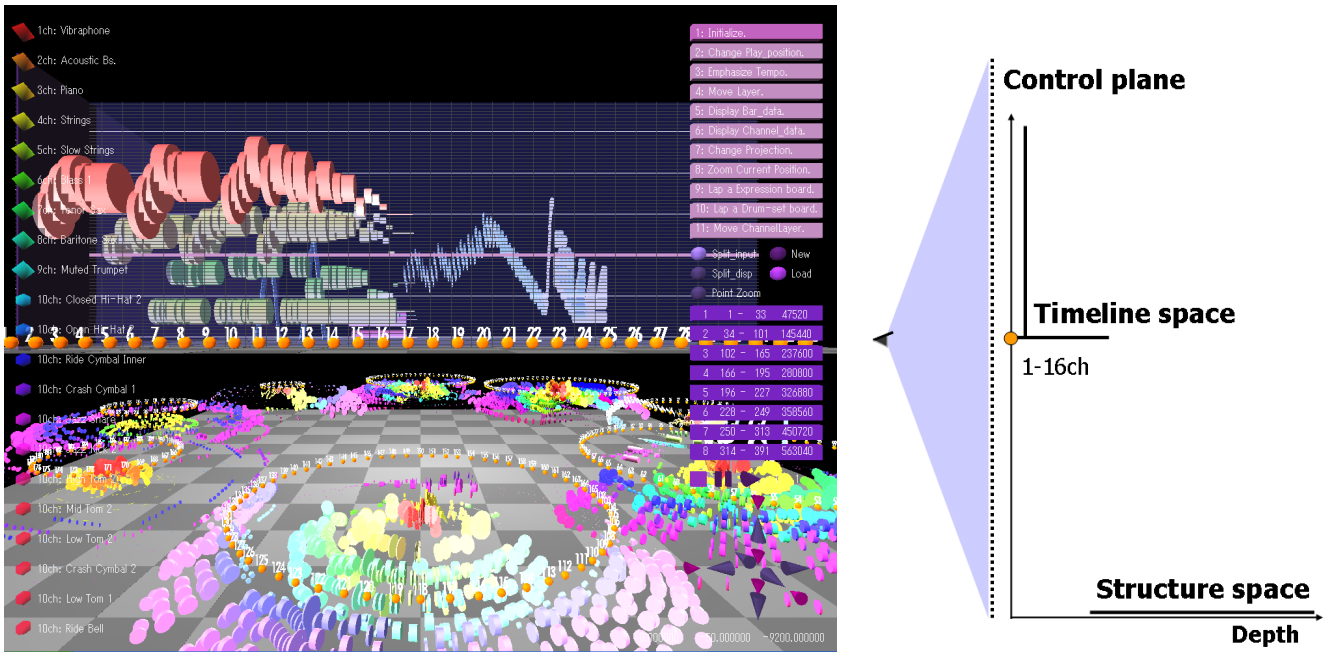
Information seeking. Information seeking is the key task with the comp-i system. To enable the users to grasp the global musical structures and local features, we decided to offer several operations each of which falls into one of Shneiderman's Information Seeking task categories (Shneiderman 1998). The Visual Information Seeking Mantra, "Overview first, zoom and filter, then details on demand." summarizes many visual design guidelines and provides an excellent framework describing the information finding process with the supported methods. In addition to these operations, comp-i makes it possible to switch global view and local view of the piece smoothly by using animated focus + context views of the target piece.

Playback. Conventional media players as well as sequence software systems provide the users with the music decoder interface which contains many buttons, such as play, pause, stop, and moreover, fast-forward, rewind, jump to previous/next track, plus the playback slider. However, the users can only change the playback position through listening to the pieces repeatedly, because the fast-forward and rewind buttons change the playback position without any reference to structural information about the piece.

The comp-i system enables the users to perform the playback-operation on a visually-encoded sequence of musical structures, instead of using the decoder metaphor, and they can change the playback position by pointing the position directly on the sequence in the virtual space.

Generating virtual sound space. To perceive the playback position and emphasize a selected channel of MIDI datasets auditorily as well, comp-i provides the virtual sound space where the users can enjoy listening to the virtual 3D sound of the target piece.

Editing. An intuitive way of visualization used in comp-i is reasonable for editing task. Sequence software systems, for example, can display only a limited number of related parameters in a single editwindow. So the users cannot verify the results of their editing work completely without juxtaposing multiple editwindows. However, comp-i enables the parameters of MIDI datasets to be edited in an easier manner because the MIDI datasets are visualized intuitively and displayed within a single window. The users can point an object corresponding to a single note sound to change its MIDI parameters directly and they do not need to open any other windows.



(A) Overview of comp-i virtual space. **(B) Right side view of comp-i virtual space.**

Figure 2: The comp-i virtual space.

3 Organization of Virtual Space and Visualization Techniques

In this section, we describe how to construct the comp-i virtual space. Figure 2 (A) shows a complete overview of the comp-i virtual space, and Figure 2 (B) illustrates the construction of the virtual space projected from right side. The lower space named *Structure Space* provides a context view which shows the whole span of the piece and visualizes its global structures. The upper space named *Timeline Space* presents a focus view which expands a musical component selected from the Structure Space. Also a head-up-display named *Control Plane*, which is placed in the front of the window, provides the operation buttons and displays several data, such as playback position with the subsection number and the musical component number, and the position of the viewpoint in the virtual space. The Control Plane can be commonly used for these two subspaces.

Figure 3 illustrates the comp-i workflow. First, the comp-i system accepts a standard MIDI file (SMF) and selects from the file, primary MIDI elements to construct the designated 3D virtual space. And second, comp-i allows the users to load a set of pre-defined musical structures, or to define new structures of the input piece. When pre-defined musical structures are loaded, comp-i represents it in the Structure Space and the first musical component is expanded in the Timeline Space as the default. On the other hand, when defining a new structure, comp-i represents the local features in the Timeline Space, and gives the split operation which enables the users to define the global musical structures of the piece. After that, comp-i utilizes several types of visualization according to the information seeking and editing operations. Lastly, the information of musical structures which the users have defined in comp-i is saved as *meta data* of SMF, and the edited data of each event is saved as *event data* of SMF.

The following subsections describe how to visualize selected MIDI events for constructing the comp-i virtual space.

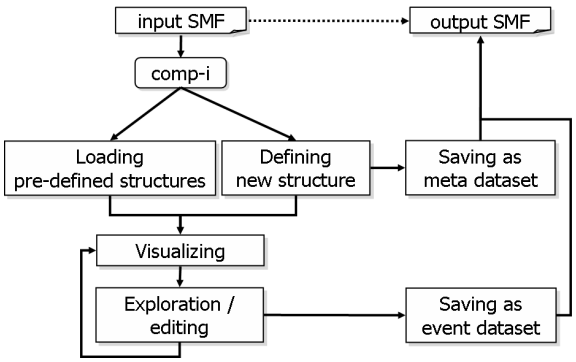


Figure 3: The comp-i workflow.

3.1 Events and parameters to be visualized

We decided to visualize the following five primary MIDI events, each of which includes timestamp and other parameters to set or change the attributes concerned with note sounds, channels, and the whole span of music piece.

- **Note-on and note-off:** Note-on and note-off are basic MIDI messages that start and stop a single note sound, and include channel, pitch, and velocity parameters.
- **Channel volume:** Channel volume sets a max volume for each channel and includes channel and sound volume parameters.
- **Expression:** Expression sets a ratio of channel volume and helps representing a performance stress. It includes channel and sound volume parameters.
- **Set-tempo:** Set-tempo is a MIDI event that sets the tempo parameter.

3.2 Construction of virtual space

Timeline Space. Figure 4 shows visualization techniques employed in the Timeline Space. A MIDI dataset has 16 channels and each of which is mapped to a channel layer. A single note sound is depicted with a cylinder. The three parameters, pitch, volume, and tempo are encoded respectively as the height, diameter, and color saturation of the corresponding cylinder. A scan-plane is drawn with a vertical plane and it is orthogonal to the channel layers. As a given MIDI dataset is played through a 3D virtual sound space, this scan-plane is moving from left to right to indicate the current playback position in the timeline. Channel volume is indicated as the height of a rectangle board that is on a channel layer, while expression is represented as the contour of another board on a channel layer (Figure 4).

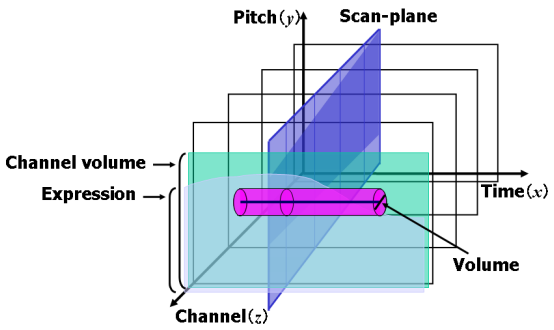


Figure 4: Parameters mapping in Timeline Space.

Every musical instrument plays notes within its limited pitch range. The maximum and minimum values of this range are typically instrument-dependent. According to the MIDI standard, pitches have values restricted to the range [0, 127]. The range of pitches for each instrument is mapped to some locations along the *y*-axis. Because music seldom makes use of the entire range of pitches, the upper and lower portions of pitch range may not be utilized. The comp-i system provides two ways to define the coordinate of the *y*-axis. One way is to use the pitch value embedded in SMF as it is. Another way is to determine the maximum and minimum pitch values and the scales of the vertical axis accordingly. Consequently, the displayed maximum and minimum values represent the maximum and minimum pitch values for a particular piece of music.

The volume of each note sound is defined by three MIDI events: velocity parameter of note-on, channel volume, and expression. The velocity of a single note specifies the loudness or softness of the note. In the visualized sequence, the velocity of a note is represented by the diameter of each cylinder. The velocity value for a note is restricted to the range [0,127]; this corresponds to each cylinder's diameter. The channel volume indicates a default volume of each channel instrument and the expression specifies the ratio of channel volume. Because it is necessary to represent these two events independently of the velocity of each single note, a board for channel volume and a wave shape board for expression are lapped over each of the channel layers. In the default display mode, comp-i does not display these boards to avoid visual cluttering[1].

The tempo parameter is represented with the color saturation of each cylinder. Therefore, the more slowly the segment played, the whiter the corresponding sequence of cylinders is colored. According to the MIDI standard, the tempo parameters are set independently of the timestamp of each note sound, so it is more appropriate to visualize the tempo independently of the length of timeline.

Drum-set parameters need to be visualized in a different way from usual instruments, because the note number, in this case, can be used to specify the kind of instrument such as Hi-Hat, Cymbal, Snare and Tom. In default display mode, each of the drum-set instruments is visualized in different layers along the *z*-axis.

Structure Space. We adopted a circular form as a spatial substrate to visualize the global structures of a long MIDI sequence effectively. Music is composed hierarchically with sentence, phrase, and motif. We attempt to employ a Cone-Trees (Robertson et al. 1991)-based multi-resolution tech-

[1] To be hard to see the necessary object because of other objects.

nique to find some similarities among these musical components.

Figure 5 is a context view which the users can see from the top of the Structure Space. Each circle is placed annularly, and the size of which suggests how large the scale of the corresponding musical structure is. The time axis starts at an angle of 0 degrees and is defined right-handedly in both each circle and the entire circular form composed of those circles. And each of the circles includes cylinders for note sounds. These cylinders displayed on different concentric layers along the time axis are assigned different colors: cylinders in red correspond to the winds, and those in blue to the strings. The tempo parameter is encoded as color saturation. Thanks to the orthographic projection, the users can recognize which instrument is used for each as well. In Figure 5, we can recognize intuitively that the music piece entitled "Valse des Fleurs" has four kinds of melodies: A-B-C-D, and takes a multiple ternary form: (ABAB)-(CDC)-(ABAB). The first circle corresponds to introduction, second, third, seventh and eighth circles to the sets of (AB), and each of the other circles to each of the component melodies.

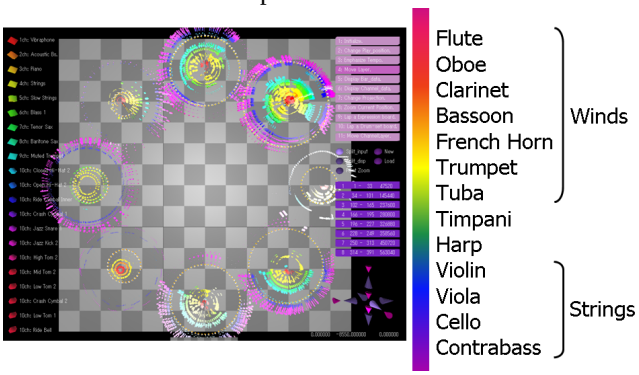
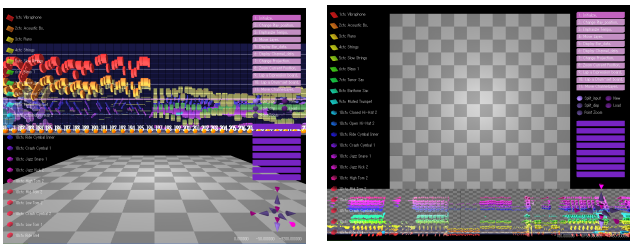


Figure 5: A view of musical structure.

Control Plane. In the front of display, comp-i provides several types of operation buttons. On the left side, the octahedra named *channel nodes* are presented to manipulate each channel in several view and playback modes. On the right side, menu bars, split buttons, structure bars, playback buttons and 3D cursors are provided. Menu bars switch one playback and nine view modes. These operation menus are described in the next section. Split buttons are used to define the structures of input SMF by splitting the timeline. Structure bars are displayed by turns as the timeline is split, and comp-i displays the structure number and start and end positions with subsection numbers. We describe more details of this operation in the next subsection. Playback-operation buttons are to STOP, PAUSE, and PLAY, from left to right. Because it is sufficient for the users to change playback position by clicking a point node along the time axis in the comp-i virtual space, the comp-i system offers just three playback operation buttons, whereas the screen interface, as used in the conventional media players, displays PREV TRACK, RAW, FF, and NEXT TRACK. By using 3D cursors, the users can move their viewpoint at will in the comp-i virtual space, take any directions that they want, jump to a particular position (Top view and front view of the comp-i virtual space, top view of the Structure Space and Timeline Space, and side view of the Timeline Space) to grasp the visualization effectively.

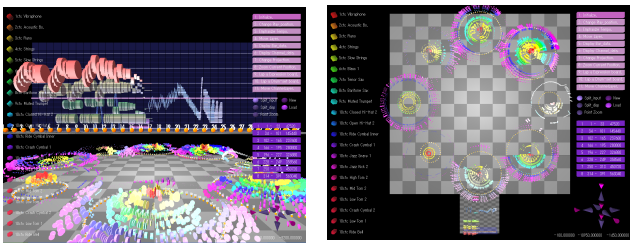
3.3 Definition of musical structure

By splitting the timeline in the Timeline Space, comp-i reorganizes the Structure Space and the users can grasp the musical structures of the input SMF. The users can find the split positions by referring to the sequences of cylinders represented in the Timeline Space. Figure 6 shows an orthographical top view of the Timeline Space. First, the users click the NEW button on the right side of the Control Plane, and the full view of the input piece appears in the Timeline Space in the form of piano-roll. Second, they find an appropriate split position by perceiving the saturation of sequences representing the tempo change, the variations in the diameter of individual cylinder visualizing crescendo or decrescendo, and the existence of cylinders in the each channel's sequence. After that, they click a point node in the Timeline Space to split the piece along the timeline, and comp-i constructs the Structure Space and visualizes each structure in the single window, regardless of the length of the input piece. Figure 7 shows the views before/after defining musical structures.



Front view. **Top view.**

(A) Before



Front view. **Top view.**

(B) After

Figure 7: Effects of musical structure definition.

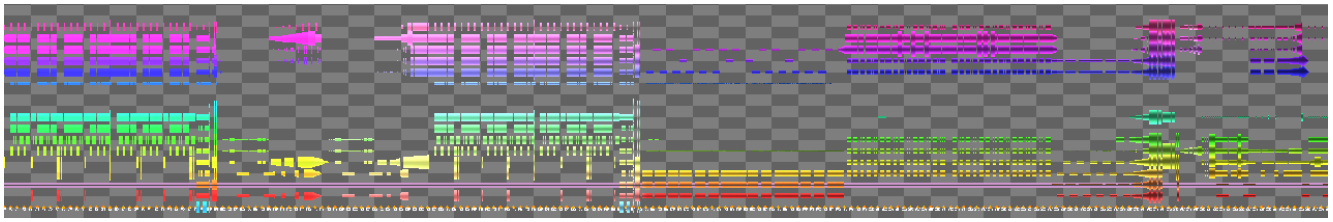


Figure 6: An orthographical view from the top of Timeline Space.

4 Operations

In this section, we will turn our focus onto the functionalities of information seeking and editing tasks with the comp-i system. All the operations offered by comp-i can be processed at an interactive rate on a standard PC (Pentium4 2.8GHz CPU, 1GB RAM). The code was written using a 3D software library (World Tool Kit[1]) and requires two 3D virtual sound devices (RSS-10[2]) to generate 3D virtual sound.

4.1 Information seeking tasks

The users can take advantage of perspective transformation to grasp the entire dataset within the virtual space. 3D illumination gives object shadows on the floor, and conveys the right information of MIDI object geometry, whereas the users can make their viewpoints as close to selected objects as they like. The users are also allowed to permute the channel layers of interest, control the visibility of retinal/auditory properties of the objects, and choose arbitrary positions to start/stop playing. Furthermore, the scales and quantitative properties are possible to be displayed along with the channel layers and the scan-plane.

Table 1 lists up all the eleven information seeking operations that the current version of the comp-i offers. **Change playback position** operation belongs to the playback mode, and the rest of the items to the view mode. These operations can also be categorized according to the Shneiderman Visual Information Seeking Mantra (VISM), which was described in Section 2. According to the VISM, **Initialize** operation is to see the overview and **Change playback position, Emphasize tempo, Zoom current position, Move channel layer, Lap an expression board, and Lap a drum-set board** operations are to adjust zooming. **Change projection** operation is for information filtering, and **Move scaled layer, Display bar data, and Display channel data** operations are used to obtain details-on-demand.

[1]World Tool Kit is the registered trademark by Roland Corporation.
[2]RSS-10 is the registered trademark by Sense8.

Table 1: List of comp-i information seeking operations.

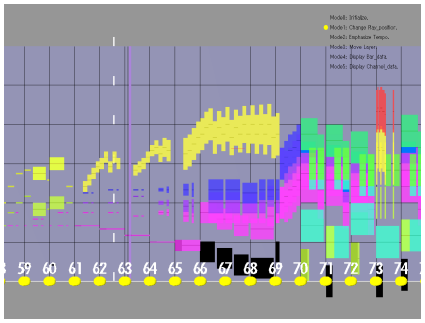
Operation	Mode	VISM category
Initialize		overview
Change playback position	playback	zoom
Emphasize tempo	view	
Zoom current position		
Move channel layer		
Lap an expression board		
Lap a drum-set board		
Change projection		filter
Move scaled layer		details
Display bar data		-on-demand
Display channel data		

Change playback position operation makes the users change the playback position in an interactive and global manner by clicking a point node in both subspaces offered by the comp-i system. When the point node located in the Structure Space is clicked to change the playback position, the structure including the position is expanded into the Timeline Space. **Move channel layer** operation is to reorganize the channel layer along the depth axis as the users like. **Move scaled layer** operation is to see the details of object's position by lapping the scale on the channel layer.

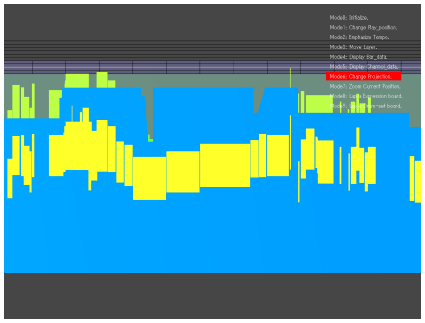
4.2 Editing tasks

After the users comprehend the outline of a given MIDI dataset, they are allowed to alter the projection to orthographical for accurate editing work (Figure 9). The comp-i system offers two ways to edit the MIDI dataset: direct manipulation of objects for novice users and textbox-based specification for experts.

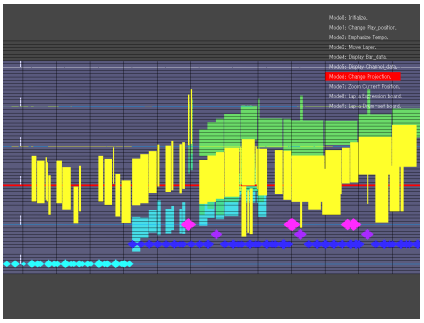
The comp-i system provides editing functions through direct manipulation of objects in the Timeline Space. After the users grasp the structures and find the positions they want to edit, they can expand the structure which includes the editing position to the Timeline Space. The cylinder objects rep-



(a) Default.

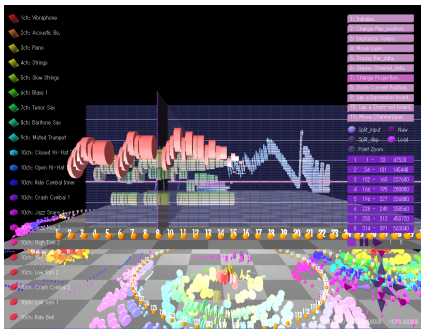


(b) Expression.

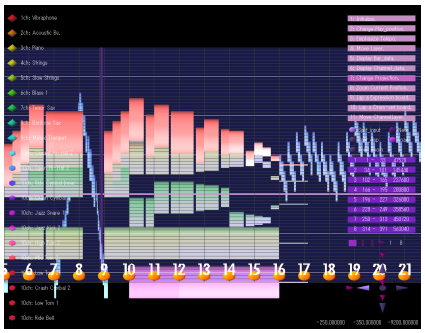


(c) Drumset.

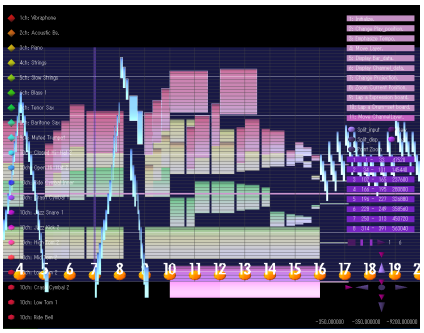
Figure 8: Lapping operations on channel layer.



(a) Perspective view.



(b) Orthographical view.



(c) After an editing operation.

Figure 9: The effect of editing task.

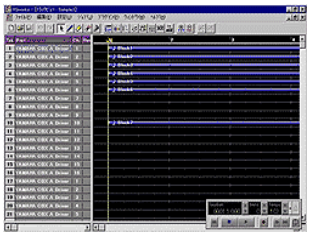
resented in the Timeline Space can be affinely transformed by using a mouse. When clicking and holding down the left mouse button on the object, the users can point and drag the object. If the object is moved along the *x*-axis, the note-on and note-off timestamps of the corresponding note are modified. If the object is moved along the *y*-axis, the pitch parameter of the corresponding note is modified. When clicking and holding down the right mouse button on the object, the users can point and scale the object. If the object is moved along the *x*-axis, the note-on and note-off timestamps of the corresponding note are modified and the length of the note is changed. If the object is moved along the *y*-axis, the velocity parameter of the corresponding note is modified.

Modified parameters are reflected to the input SMF immediately. The edited objects are accentuated with white wire frame, thus the users can recognize which objects have been modified.

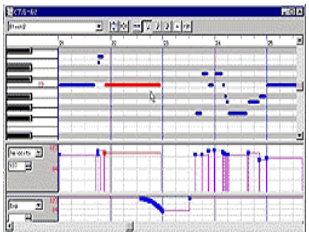
5 Comparison with Sequence Software Systems

Choice of particular projections of the Timeline Space can offer better GUIs than the traditional sequence software systems do. For example, the top orthographical view can give a clear view of multi-channel information involving both volume and tempo. The pitch parameters can be recognized by changing projection to perspective (Figure 10 (a)-(b)). The side view of the Timeline Space makes it possible for the users to look over both pitch and volume of multiple channels simultaneously (Figure 10 (a')-(b')). In addition to the fact that each of the projections that comp-i offers has a corresponding editwindows of the sequence software systems, the comp-i system provides more global musical views in the same window by constructing the Structure Space while the sequence software systems can only show a particular section of the piece in each editwindow.

It should be emphasized here that transforming viewpoint continuously to particular projections in a single virtual space can minimize the deterioration of the user's cognitive maps, and thus leading to a sort of usability. This is the salient feature which differs the comp-i system from the conventional sequence software systems.

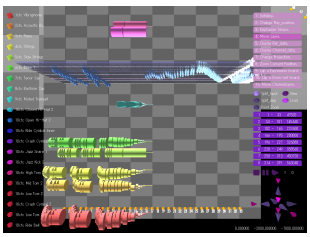


(a) Track view window.

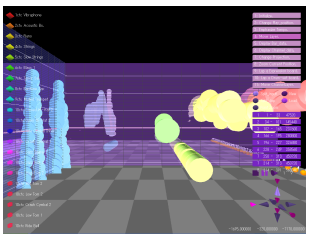


(a') Piano-roll window.

A sequence software system.



(b) Top view.



(b') Side view.

The comp-i system.

Figure 10: Choice of particular projection of the comp-i virtual space supersedes typical GUIs of a sequence software ((a)(a'): sequence software, (b)(b'): comp-i).

6 Conclusion and Future Work

In this paper, we have described the comp-i system for exploration and editing of MIDI datasets by visualizing MIDI encoded music through a 3D virtual space and providing an interactive and intuitive interface.

In the future, we plan to estimate the musical structures automatically by paying more attention to repetition structure of music. Foote (1999) shows that a similarity matrix applied to well-chosen features (MFCC in (Foote 1999)) leads to a visual representation of the structural information of music piece. The studies examining repetition section of a piece of music can be found in the literature (Wattenberg 2002; Goto 2003; Peeters and Rodet 2003). It can be said that detecting repetition sequence is effective in music visualization including chorus search function and music summary. We propose to detect the repetition section of music for automatic construction of the Structure Space. We also attempt to offer visual MIDI data mining with customizable data mappings and the detecting techniques, especially interactive similarity search in the present virtual space.

References

Card, S., J. Mackinlay, and B. Shneiderman (1999). *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Pub.

Foote, J. (1999). Visualizing music and audio using self-similarity. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, pp. 77–80. ACM Press.

Goto, M. (2003). Smartmusickiosk: Music listening station with chorus-search function. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST 2003)*, pp. 31–40.

Hiraga, R., R. Miyazaki, and I. Fujishiro (2002). Performance visualization — a new challenge to music through visualization. In *Proceedings of the ACM Multimedia 2002*, pp. 239–242.

Kaper, H. G. and S. Tipei (1998). Manifold compositions, music visualization, and scientific sonification in an immersive virtual-reality environment. In *Proceedings of the International Computer Music Conference*, pp. 339–405. International Computer Music Association.

Kunze, T. and H. Taube (1996). See — a structured event editor: Visualizing compositional data in common music. In *Proceedings of the International Computer Music Conference*, pp. 63–66. International Computer Music Association.

Lerdahl, F. and R. Jackendoff (1983). *A Generative Theory of Tonal Music*. The MIT Press.

Miyazaki, R. and I. Fujishiro (2002). 3D visualization of MIDI dataset. In *IEEE Visualization 2002 Posters Compendium*, pp. 96–97.

Miyazaki, R., I. Fujishiro, and R. Hiraga (2003). Exploring MIDI datasets. In *ACM SIGGRAPH2003 Full conference DVD-ROM*.

Peeters, G. and X. Rodet (2003). Signal-based music structure discovery for music audio summary generation. In *Proceedings of the International Computer Music Conference*, pp. 15–22.

Robertson, G. G., J. D. Mackinlay, and S. K. Card (1991). Cone-trees: Animated 3D visualizations hierarchical information. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'91)*, pp. 189–194. ACM Press.

Shneiderman, B. (1998). *Designing the User Interface Strategies for Effective Human-Computer Interaction, Version 3*. Addison-Wesley.

Smith, S. M. and G. N. Williams (1997). A visualization of music. In *Proceedings of the IEEE Visualization 1997*, pp. 499–503.

Spence, R. (2001). *Information Visualization*. Addison-Wesley.

Wattenberg, M. (2002). Arc diagrams: Visualizing structure in strings. In *Proceedings of the IEEE Information Visualization 2002*, pp. 110–116.