VISUALIZING THE SEMANTIC STRUCTURE
IN CLASSICAL MUSIC WORKS

by

WING YI CHAN

A Thesis Submitted to
The Hong Kong University of Science and Technology
in Partial Fulfillment of the Requirements for
the Degree of Master of Philosophy
in Computer Science and Engineering

January 2009, Hong Kong

Copyright © by Wing Yi Chan 2009
Authorization

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science and Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science and Technology to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

______________________________
WING YI CHAN

13 January 2009
VISUALIZING THE SEMANTIC STRUCTURE IN CLASSICAL MUSIC WORKS

by

WING YI CHAN

This is to certify that I have examined the above M.Phil. thesis and have found that it is complete and satisfactory in all respects, and that any and all revisions required by the thesis examination committee have been made.

DR. HUAMIN QU, THESIS SUPERVISOR

PROF. MOUNIR HAMDI, HEAD OF DEPARTMENT

Department of Computer Science and Engineering

13 January 2009
ACKNOWLEDGMENTS

First and foremost I offer my sincerest gratitude to my supervisor, Dr Huamin Qu, for his continuous guidance and patience, and granting me freedom to work in my own way. Without his help, this thesis would not have been possible. I joined the visualization research team in Summer 2005 when I had just survived my freshman year; throughout the years we have bitter and sweet moments, but eventually when I look back I treasure all the bits and pieces that constitutes the five years of fruitful studies at HKUST. I will always be proud of what I have achieved in my early twenties and have no regrets at all.

I am grateful to have met numerous great professors here. In particular, I would like to thank my undergraduate advisor Professor Lionel Ni for his advice and encouragements. I am sincerely thankful to Dr. Nevin Zhang, who was the Undergraduate Coordinator at the time I was confused about my future and career. It was his informative phone call that brought the indecisive 18-year old girl to the Department of Computer Science and Engineering at HKUST, and made her determined to pursue for a master degree after introducing her to the Honors Study Track. Thanks are also due to my thesis committee members Professor Andrew Horner and Dr. David Rossiter.

I wish to express my thanks to everyone who has helped me with my research work. I specifically acknowledge some of my past and present group-mates: Ming-Yuen Chan, Charles Chung, Peter Chung, Weiwei Cui, Jacky Tang, Haomian Wang, Yingcai Wu, and Hong Zhou. I also appreciate the help of Ying Tang and Anbang Xu.

During my college life I have matured and better discovered myself; thanks for the unconditional tolerance and support from my dearest friends to assist me through this painful yet rewarding process. My best friend Janet is always hatefully adorable, however adorably hateful she is sometimes. I also miss the gossip lunch dates with my baby Mona at LG7 Deli where we poured our updates over the tonkatsu pork fillets served with tomato sauce and iced lemon tea. Cindy, Sze Ning, Veronica, and many other old friends have made my study much more colorful.

Finally, I would like to thank my parents for their gratuitous love, and my beloved fuzzy rabbit Gilbert for accompanying and torturing me every night. Just as important is my ultimate soulmate Wallace, the most precious and dear takeaway from HKUST.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title Page</td>
<td>i</td>
</tr>
<tr>
<td>Authorization Page</td>
<td>ii</td>
</tr>
<tr>
<td>Signature Page</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Abstract</td>
<td>ix</td>
</tr>
</tbody>
</table>

## Chapter 1 Introduction

1.1 Background 1
1.2 Motivations 1
1.3 Formulation 2
1.4 Contributions 3
1.5 Organization 3

## Chapter 2 Related work

2.1 Musical Structure Visualization 4
   2.1.1 Arc Diagrams 4
   2.1.2 Isochords 5
   2.1.3 ImproViz 6

2.2 Abstracted Scores 7
   2.2.1 Graphic Scores 7
   2.2.2 Simplified Scores 8

2.3 Performance Expression Visualization 9
   2.3.1 Vertical Bar Display 10
   2.3.2 Chernoff Faces 10
   2.3.3 Hierarchical Approach 11
   2.3.4 Music and Emotion 12

2.4 3D Music Visualization 12
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.3 Theme and Variations</td>
<td>39</td>
</tr>
<tr>
<td>8.4 Macro-Micro Confluence</td>
<td>40</td>
</tr>
<tr>
<td><strong>Chapter 9  Evaluation</strong></td>
<td></td>
</tr>
<tr>
<td>9.1 User Study</td>
<td></td>
</tr>
<tr>
<td>9.1.1 Tasks</td>
<td>42</td>
</tr>
<tr>
<td>9.1.2 Results</td>
<td>42</td>
</tr>
<tr>
<td>9.1.3 Discussion</td>
<td>43</td>
</tr>
<tr>
<td>9.2 User Survey</td>
<td></td>
</tr>
<tr>
<td>9.2.1 Results</td>
<td>44</td>
</tr>
<tr>
<td>9.2.2 Discussion</td>
<td>44</td>
</tr>
<tr>
<td><strong>Chapter 10  Conclusion</strong></td>
<td></td>
</tr>
<tr>
<td><strong>References</strong></td>
<td>49</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1.1 Traditional ways of learning music with musical score and technical essay. 2
2.1 Arc Diagrams example. 4
2.2 Isochords example. 5
2.3 ImproViz example. 7
2.4 Graphic Scores example. 8
2.5 Simplified score BRASS example. 8
2.6 Performance visualization with bars. 9
2.7 Performance visualization with Chernoff faces. 11
2.8 Performance visualization with hierarchical approach. 11
2.9 Performance visualization of the mood of music. 12
2.10 Screenshots of the comp-i system. 13
2.11 Examples of immersive approaches in virtual reality. 13
2.12 TimeSketch example. 14
2.13 Music Animation Machine graphical modules. 14
2.14 Screenshots of Hyperscore system 15
2.15 Visualization in computer music. 16
2.16 Tonal content visualization. 17
5.1 Color Scheme for layers. 23
5.2 Theme glyph example. 24
5.3 Basic type anatomy in typography. 26
6.1 Layer braid example. 28
6.2 Layer braid prototype. 28
6.3 Layer braid with different levels of detail. 29
6.4 Layer braid with focus+context technique. 29
6.5 Theme fabric prototype. 30
6.6 Theme fabric styles in textile art. 31
6.7 Theme fabric in plain style example. 31
6.8 Generating bundled style from plain style theme fabric. 32
6.9 Theme fabric in bundled style example. 33
6.10 Collapsing a theme fabric. 34
6.11 Theme fabric in collapsed style example. 34
9.1 User study result with the general audience. 43
9.2 Survey results with music experts. 44
VISUALIZING THE SEMANTIC STRUCTURE IN CLASSICAL MUSIC WORKS

by

WING YI CHAN

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

ABSTRACT

One of the major obstacles to appreciating classical music is that extensive training is required to understand musical structure and compositional techniques towards comprehending the thoughts behind the musical work. In this thesis, we propose an innovative visualization solution to reveal the semantic structure in classical orchestral work, such that users can gain insights into musical structure and appreciate the beauty of music. We formulate the semantic structure into macro-level layer interactions, micro-level theme variations, and macro-micro relationship between themes and layers to abstract the complicated construction of a music composition. The visualization has been applied with success in understanding some classical music work as supported by highly promising user study results with the general audience and very positive feedback from music students and experts, demonstrating its effectiveness in conveying the sophistication and beauty of classical music to novice users with informative and intuitive displays.
CHAPTER 1

INTRODUCTION

1.1 Background

While pop music is intended to be friendly to the general public, difficulties arise when we listen to classical music. The actual structure and form of the music are usually only understood by music lovers, who have received extensive training in music theory and history. This high learning curve thus makes classical music seemingly intricate and unapproachable. In this thesis, we demonstrate how information visualization techniques can be novelly applied to visualize the semantic structure of music such that users are immediately impressed by the sophistication and beauty of classical music with the vivid visual displays, and are then able to visually gain insights into musical structure for brand new listening experience and facilitating music appreciation, learning, and teaching.

1.2 Motivations

Realizing musical structure is an important step in appreciating music, but it is often difficult for non-experts due to its abstract and temporal nature. While one can readily identify the concrete visual features in a painting, it is unlikely that unskilled ears could easily recognize complicated musical elements from multi-layered music. Some believe that the lack of visual equivalence is what makes music different from other art forms, but this is also why music appears to be more abstract. Together with its dynamic and time-varying features, appreciation and comparisons become more laborious.

The most common way to learn music is to study the musical score and technical essays (Figure 1.1). Reading the score is, however, extremely demanding. Beginners have to spend considerable time in learning the basics of music theory before being capable of understanding the notations; and there is still a long journey toward mastering a score for in-depth understanding. Even for amateurs with sufficient background, it could be tedious to go through the original score which is flooded with technical details. Another alternative is to read pages of technical essays, but it is certainly not a very fascinating experience in particular for the general audience, as the essays are often very abstract and full of jargon.
The first theme begins presented in a minor key by the violins. Then part of the first theme is repeated by the violins. The transition begins in the violins; it contains a short motive and many scalewise passages. The second theme is presented by the violins and woodwinds. The second theme is repeated by the violins. The music moves on to some transitional material. A motive from the first theme is played by the clarinet and bassoon. The codetta begins with rapidly moving notes played by the violins. The codetta concludes with a solid chord. The development section begins with two short chords and notes played by the woodwinds. A portion of the first theme is played by the violins three times in sequence, each time one note lower than before. A portion of the first theme is played by the lower strings as the upper strings play a contrasting line of rapidly moving notes. The upper and lower strings trade what they play. The upper and lower strings again trade what they play. The violins continue playing portions of the first theme. The motive from the first theme alternates quietly several times between the violins and the flute. Fragments of the first theme are exchanged more energetically between the upper and lower strings. A short link leads to the recapitulation. The first theme returns played by the violins in the tonic key. The first theme continues. Part of the first theme is repeated by the violins. The transition begins…

Figure 1.1: Traditional ways of learning music with a musical score (left) and a technical essay [27] (right).

To aid the general public in appreciating the sophisticated structure in classical music works, visualization is one of the most promising approaches, thanks to our strong visual cognition ability. Although numerous efforts have been devoted to the visualization of sonic characteristics, few attempts have been made to visualize the semantic structure of music instead of the low-level physical properties for understanding the music perceptually. Moreover, the important layer interactions and theme variations are barely addressed.

1.3 Formulation

The beauty of classical music comes from its elegantly complicated yet beautifully arranged structure. The lowest semantic level is constructed by themes, which are recurring main melodies and principal musical ideas. Composers achieve balance in time by arranging and setting them in order, the basic framework of which is then clothed with interesting details by applying a variety of musical materials, resulting in different variations [3]. Apart from this horizontal foundation across time, a composition can be viewed from the layer perspective. Different layers in the music collaborate, accompany, blend, and contrast with one another, giving the rich vertical texture and depth of music. In brief, the micro-relationship between themes contributes to the breadth of the music, while the macro-relationship among layers gives the depth of music. Furthermore, these two equally important relationships are not independent. A layer contains information about themes and themes are played by various layers; this interaction results in an extra macro-micro relationship between themes and layers, ultimately yielding a piece of sophisticated music.
1.4 Contributions

We pioneer several intuitive design principles and effective visualization prototypes to convey the semantic structure in classical music which novice users without musical backgrounds can effortlessly master within seconds. A native color scheme is proposed to depict orchestral instruments for macro-level layer information. The micro-level themes are represented by configurable glyphs, with theme variations being encoded by a glyph grammar based on typographical concepts for amplified cognition. With these building blocks, the layer braid and theme fabric prototypes are implemented to visualize the layer interactions and theme repetitions across layers respectively. By introducing an analogy to the familiar plaiting and weaving in textile art, the general audience can rapidly gain insights into complex musical structure with these impressive visualization. They can rely upon vivid visual displays, instead of dull technical essays written by experts, towards understanding the semantic structure and thus appreciating the sophistication and beauty of music for a more enjoyable listening experience. Music teachers, music students, and musicians also benefit as the visualization reveals musical structure in an intuitive way for teaching and learning different compositional and musical styles, as well as providing interesting visual clues for musical features that may be otherwise overlooked easily without the visual aid.

1.5 Organization

This thesis is organized as follows: After introducing the related work in Chapter 2, we explain the musical structure data in Chapter 3. Chapter 5 presents the design principles for encoding the important layer and theme components, from which the braid and fabric prototypes in Chapter 6 are constructed to reveal the sophisticated relationships in music. Chapter 7 outlines the user interactions provided by our system for visual exploration. Next we discuss some case studies in Chapter 8 to illustrate the merit of our techniques. Chapter 9 describes the user study conducted with the general audience and the user survey feedback received from music students and experts. Finally it is concluded in Chapter 10.
CHAPTER 2

RELATED WORK

The visualization of musical structure has not been explicitly addressed in the literature. Most of the music visualization approaches are highly mathematical and physical, while we emphasize on the human interpretation of musical structure. The visualization techniques summarized below emphasize on novel visualizations of musical elements, forms and structure. Sound visualization, 3D immersive visualization of music in virtual reality, music structure analysis or extractor, and music information retrieval are not included in this scope of study. Since the semantic structure is mainly qualitative, general quantitative techniques \cite{8,57,12} cannot be deployed directly. Instead, new approaches are demanded to address the special qualitative features.

![Arc Diagrams: (a) Three repeated substrings 1234567 connected by two arcs. The width of the arc is proportional to the length of the repeated subsequence while the radius is proportional to the distance between the matching pair. (b) An example illustrating Für Elise.](image)

2.1 Musical Structure Visualization

2.1.1 Arc Diagrams

Arc Diagrams \cite{58,59} visualize complex patterns of repetition in string data by connecting a translucent arc between a pair of matching pair (Figure 2.1). The main application is to visualize the repeated elements in a composition. The author derived an automatic analysis method to extract the repeated patterns from a MIDI file. While the outcomes are aesthetically pleasing, it is uncertain whether the repeated substrings identified by the algorithm essentially correspond to the musical repeated units such as theme, subject, motif or the more specialized *leitmotif* and *idée fixe*. 

---

4
2.1.2 Isochords

Isochords [4] [5] apply Tonnetz, a two-dimensional triangular isometric coordinate grid invented by Euler (Figure 2.2(a)), to visualize the chord structure, progression and voicing of a composition in a MIDI file. Tonnetz, the German word for tone-network, is popular in modern musical analysis. It is a conceptual lattice diagram to place each chord type in a 2D space according to the relationships between musical pitches. Hence the major contribution of this paper is to propose an animated display of the usually static Tonnetz, and to reveal various chord properties in the same representation.

Each note is represented by a dot, and the dots of the chord notes form an Isochords geometry. Under the Isochords configuration, upward and downward pointing triangles represent major and minor chords respectively; diminished, augmented and other chord types are constructed using the same principle, yielding Isochords in different shapes (Figure 2.2(b)). An edge is drawn between two notes only if they are consonant\(^1\) to highlight these more important intervals. Chord progression\(^2\) is encoded with adjacent structure in the display, and color may additionally show the modulation\(^3\). The size of the triangle vertices denotes further the octave of the note with a larger dot referring to a lower octave. The inversion\(^4\) of chords is not explicitly handled though.

![Figure 2.2: Isochords: (a) Traditional Tonnetz grid. (b) Seventh chords rendered as Isochords. (c) An example of Isochords display.](image)

\(^1\) **Consonant**: harmonious combination of tones that provides a sense of stability in music [44]

\(^2\) **Chord Progression**: the movement from one chord to the next [54]

\(^3\) **Modulation**: the process of changing from one key to another [44]

\(^4\) **Inversion**: the position of a chord when the root position is not the lowest note [54]
While chords are important in music analysis, the more general descriptions such as phasing, orchestration and variation of themes would be more useful for audience to understand the high-level structure of the composition. As a result, they would not get lost when listening to the performance and can therefore possibly grasp the basic musical notations and techniques.

It is stated that Isochords display is a visualization of structure in music. However, the structure defined here, which emphasizes on low-level units like notes and chords, is slightly different from what we would like to examine. Structure can be interpreted from a technical, quantitative point of view on the detailed constructive and structural materials used, or in a more general, qualitative manner about the abstract, semantic structure of a piece of work. Most of the research on visualizing the structure of music tackles the former for music theory training especially in the computer music aspect (Section 2.6), yet the latter, with a huge number of potential users for the purpose of music appreciation, is rarely studied in a formal context.

### 2.1.3 ImproViz

ImproViz [53] is a visualization technique showing jazz improvisations to discover melodic and harmonic patterns. A melodic landscape shows the contour of a melodic line and a harmony palette visualizes the use of chords (Figure 2.3). The only example examined is All Blues by Miles Davis. The design of ImproViz was solely based on this composition, hence imposing numerous constraints if one wants to visualize other jazz work. Unlike the previously discussed techniques, the ImproViz results were created in Adobe Illustrator and were based on a book of transcriptions (sheet music) instead of MIDI files.

Improvisations are about impromptu. The use of transcriptions in fact violates the spirit of improvising during performance time. Therefore, although the visualization was designed for jazz improvisations, the technique is fairly general that shows the underlying melodic and harmonic structure of an input sheet music not only limited to jazz. Nevertheless, melodic contour is effective to reveal the shape of the music visually. Also, it is easier for the listeners to follow the graphical representation with the melodic hints.

---

5. *Orchestration*: technique of setting instruments in various combinations [44]

6. *Improvisation*: creation of a musical composition while it is being performed [44]
2.2 Abstracted Scores

2.2.1 Graphic Scores

In [15], the author crafted graphic scores for two electro-acoustic compositions. The graphical elements applied are primitive, and the graphs do not aim at generalization for other pieces of music work (Figure 2.4). In spite of this, the idea behind this work is to provide a graphic study scores for listeners. The author suggested that traditional musical score is written for performance but not analysis. Although a score is regarded as a graphic description of the sonic events comprising a composition, many constructional details are often missed. The followings are the considerations provided by the author in making an effective study score for the listeners:

- Sonic events should be simple, but visually identifiable
- Temporal logic should match the spatial logic
- The full study score should be visible at a glance
- Score reading is not the most important
- Study scores are for listening, in service to the ears, not for quantitative analysis
Other work by the author include sonic map, time slice and heterophonic map [16] [17], which focus on how to map color (still images) to sound (music) and vice versa. These techniques could also be incorporated to visually reinforce the sound perceived by the listeners.

![Figure 2.4: The graphic study score of an electro-acoustic composition *Incantation* (1953) by Otto Luening and Vladimir Ussachevsky.](image)

2.2.2 Simplified Scores

A simplified score called BRASS is exploited to enables users, potentially with limited background in music, to view the whole piece at once [43]. Each measure is represented as a vertical line. The number of notes in a measure defines the brightness of the line and the dynamics determines its width; texture as well as glyph are further utilized to

![Figure 2.5: The simplified and compressed score: (a) Initial overview of the whole piece. (b) With fisheye for displaying the original score.](image)
visualize tempo and other specifications (Figure 2.5). The conventional focus+context technique fisheye is then introduced to facilitate details navigation. The mapping and methods presented are straightforward; neither the effectiveness nor the practical usage of such scores were demonstrated.

### 2.3 Performance Expression Visualization

Performance expression visualization focuses on the expressions brought by different performances of the same composition. These expressive attributes do not appear in the original musical scores concisely, or are often added by the individual performer. Augmented scores [34] [46] that include these additional expressive parameters are used as input for such kind of visualization. The objective is to visualize the depth of performance so that users can compare and learn from distinct performances. A number of related work was conducted by a research group in Japan [23] [24] [25] [26] [42] [43] for understanding the music performance in cognitive terms described by melody, rhythm and phrasing.

The philosophy behind performance visualization is to display qualitative musical characteristics so as to deepen our understanding of the composition. It is suggested that MIDI parameters do not have any music sense that can be connected to human perception. Moreover, the result does not necessarily render the original scores faithfully; it should provide an insightful display for comprehending the underlying music and structure in the abstract level. The major difficulty is to detect the musical role of each note, as music perception is usually generated by a group of notes.

![Figure 2.6: Performance visualization with bars: (a) The mapping. (b) A performance that follows a score precisely for any composition. (c) Visualization of a performance of Beethoven’s Piano Sonata Op.13 in C minor, “Pathetique”.
](image)
2.3.1 Vertical Bar Display

In [25], quantifiable expression elements include tempo change, articulation\(^7\), and dynamics change were chosen as they can be appreciated qualitatively and have an affinity with music cognition. They are mapped to a 2D graph (Figure 2.6(a)), with the x-axis denoting the time and the y-axis showing the relative dynamics. In the basic grid, a vertical line indicates the start of a note in the performance. The local tempo variation, that is the difference between the performed tempo and the original tempo on the score, determines the interval between the two lines. If there is no tempo disagreement, all the intervals will be equal (Figure 2.6(b)). Each played note is then assigned to a rectangular bar placed between two vertical grid lines. The height and the width encode the dynamics and articulation respectively. A legato\(^8\) passage results in a series of rectangles connected one by one, whereas a staccato\(^9\) passage creates discrete bars. The gray scale of the bar further visualizes the expressiveness of that note, with the darkest denotes the most impressive notes having biggest impact on the listener. However, the authors did not explicitly mention how these expressive elements were defined and maintained. It is suspected that the input was an augmented score consisting of these expressions. The repeating patterns observed (Figure 2.6(c)) additionally show us the player’s phrasing.

Finally, the authors ran a pilot experiment on matching performances to the graphical displays. In the first task, the same section of music is played in two interpretations; in the second task different sections of the same composition are used for comparisons. As user-interaction is limited, the result of task one was unsatisfactory by showing just the static images without any indication for locating the current position being performed and for relating the visualization to the actual performance.

In summary, the visualization proposed here is more informative in terms of listening comprehension and music appreciation. But still, the authors suggested that other musical features should also be incorporated to increase the effectiveness. Animation should also be explored to amplify listeners’ cognition.

2.3.2 Chernoff Faces

Chernoff faces [9] were used to visualize the performance expression in [23] [43]. The tempo, articulation and dynamics change of each note are mapped to the eyeball pos-

\(^7\)Articulation: characteristics of the attack, duration, and decay of a given note [54]

\(^8\)Legato: smooth and connected notes [44]

\(^9\)Staccato: short and detached notes [44]
tion, the contour of the face and the shape of nose of the corresponding face (Figure 2.7).
The multidimensional characteristic of Chernoff faces makes it highly flexible to visualize more expressive cues. The true musical abstract features are shown instead of the MIDI mechanical data values. However, this is a proposed prototype and the results are artificially made. Also, all notes are visualized uniformly in the current schema; priority and grouping should be included to produce more meaningful outcomes.

![Chernoff faces](image.jpg)

**Figure 2.7:** Visualizing musical expression using Chernoff faces.

### 2.3.3 Hierarchical Approach

A hierarchical approach was proposed [26] [41] [42] towards visualizing musical structure. A circular form is first applied to display the global structure of a long sequence. Then Cone Trees [48] layout is used to visualize the musical structure of recursive hierarchies including form, section, phrase and motif (Figure 2.8). The implementation details were not given in the papers and it is not certain whether the visualization results are generated automatically or crafted manually. Yet the idea of employing typical hierarchical methods in information visualization to show the structure of music should be thoroughly considered.

![Hierarchical approach](image.jpg)

**Figure 2.8:** Hierarchical approach: (a) Circular form. Each level is rendered as a circle for long sequence overview. (b) Cone Trees layout. (c) Viewing from the top. We can observe that the upper layer is inherited to the lower in terms of musical structure.
2.3.4 Music and Emotion

[24] derives a graphical expression of the mood of music (Figure 2.9). Musical mood is subjective and is usually left to the listeners' interpretation. However, people may find it difficult in translating the invoked mood into words and phases, unless they have been trained. This work is part of the performance visualization which they visualize musical mood as a snapshot of a performance with a rectangular texture. Each note is represented by a small colored square based on its expression attribute values; the size does not have any specific meaning though. The squares are then arranged along a zigzag line based on the importance of the note calculated by a music analysis method time-span tree [40]. Despite the linear nature of music, the mood generated does not necessarily preserve the temporal sequence. Therefore, the visualization does not depend on the duration, and the time order of the data is not important. Also, all notes, not just the melody, are included because mood is generated as a whole from the composition.

Apart from that, a series of music-emotion structural rules is proposed in [37] for deciding the emotion of music from the underlying musical elements. [14] presented a different emotional data extracting algorithm and mapped the emotion to facial expressions.

![Figure 2.9: An example of a graphical expression of the mood of music.](image)

2.4 3D Music Visualization

[51] [52] discussed the possibility of visualizing music using color and 3D space. A MIDI data file is taken as input. The x-axis refers to the pitch range, the y-axis refers to the instrument type and z is the time axis denoting the start time of the notes. A note is represented by a sphere being placed in the 3D space, with the height, radius and color of the sphere indicating the pitch, volume and timbre respectively. As a result, color can differentiate instrument groups with different timbre. As each tone ends, the sphere will be replaced by a history marker which is a relatively smaller and lighter sphere. Each marker is scaled along the time axis according to the original note duration; long notes thus generate markers in ellipsoids. Furthermore, the intensity of the history markers will eventually decrease with time.
Figure 2.10: Screenshots of the comp-i system.

Figure 2.11: Examples of immersive approaches in virtual reality.

The comp-i (Comprehensible MIDI Player - Interactive) system [26] [41] [42] allows visual exploration on MIDI data in an immersive manner (Figure 2.10). The system again accepts MIDI files as input and visualizes five primary MIDI parameters in the 3D virtual space, namely note-on, note-off, set-tempo, expression and channel-volume. Multiple channel layers are stacked along the $z$-axis, while the $x$- and $y$-axis correspond to the time and pitch accordingly like a typical musical score.

In concerning the other immersive approaches in virtual reality which is beyond the scope of study, they focus on visualizing the synthesized sound or MIDI data attributes in a realistic 3D environment using objects or landscape [30] [47] [56]. They provide an intuitive way for users to manipulate the sound produced to achieve some music performances in the virtual environment (Figure 2.11). Obviously the potential of visualizing semantic musical structure and elements in the 3D space has not been well explored yet.

2.5 Commercial Products

2.5.1 TimeSketch

TimeSketch [28] is a commercial system designed to facilitate listening and analyzing music, as well as creating guided listening lessons from audio files. Experts or educators first define the input data via an editor. They can specify the appearance of each theme and how a theme is related with other themes. Descriptive text can also be added and will be shown at specific offsets as the audio track is played. The musical themes are
encoded by bubble chart (Figure 2.12), which are half-disks of varying sizes depending on the scope and length of the theme. Hierarchical relationships are obvious by enclosing the same-level themes in a larger semi-circle. Color is used to indicate related passages, but the outcomes are sometimes confusing. TimeSketch graphics are seen in the multimedia CDs accompanying textbooks on introduction to classical music and music appreciation [27].

![Figure 2.12: TimeSketch of the first movement of Symphony No.40 in G Minor, K.550 by Mozart.](image)

### 2.5.2 Music Animation Machine

In the basic bar-graph of Music Animation Machine [39], each note is represented by a bar with its length corresponding to the exact duration as performed. The vertical position of the bar is mapped to the pitch and the horizontal position indicates the timing. The bar color categorizes instruments and structural parts in a composition, or to show the pitch class implying the harmonic structure. For the Music Animation Machine MIDI Player, different kinds of displays are also included to visualize chords, intervals, melody and harmony (Figure 2.13). However, these modules are separated and cannot be rendered synchronically. Again, the objective is to eliminate the conventional notations on a musical score solely for performers by abstracting the melodic motion, texture and structure to listeners.

![Figure 2.13: Different graphical modules available on Music Animation Machine MIDI Player; top left-hand corner is the basic bar-graph.](image)
2.5.3 Hyperscore

A closely relevant application Hyperscore [18] [36] is a computer-assisted composition system that takes freehand drawing as input, such that users with limited musical training can literally sketch their compositions (Figure 2.14). Users first construct the melodic lines of a motive in the motive windows by adding droplets representing notes, where the \( x \)-axis indicates pitch and \( y \)-axis refers to time. Users can then specify how the motive is repeated over time in the sketch windows by drawing colored line categorizing different motives. A motive is melodically varied depending on the position and curvature of the sketched line, and is rhythmically altered according to the length of the sketched line. The line texture further suggests the instrument being used. The advanced harmony information is described by a central line where different shaping of the line would affect the harmonic progressions. The exact key is determined by its vertical location. Despite being a composition application, the graphical interface shares some similarities with the visualization of musical structure as the underlying musical elements should be expressed intuitively. It also suggests that using simple colors, textures, and curves are effective to convey musical information to novice users.

Figure 2.14: Screenshots of Hyperscore system. (a) The motive windows. (b) The sketch windows with motives in different colors.
2.6 Visualization in Computer Music

The computer music community does not deliberately study the visualization of musical structure, yet various graphical tools are developed for visually analyzing music. Some representative ones are included in this section.

A self-similarity grid [19] [20] is deployed to visualize the time structure of musical waveforms which helps comparing audio recordings for their acoustic similarity (Figure 2.15(a)). A composition is often self-similar which contains several repeated passages or variations of themes. A 2D array based on dotplot [11] is utilized where the brightness of the pixel at \((i,j)\) is proportional to the audio similarity for instants \(i\) and \(j\). This method can therefore visualize different structural repetitions such as melodic and rhythmic, or can be generalized for other types of data with similar behavior. The major contribution of this work is to propose sophisticated analysis algorithms for computing the similarity, but the details are excluded in this report.

Harmonic visualization of tonal music [49] [50] maps tonality\(^{10}\) to color; the \(x\)-axis represents time in the score, and the vertical axis stands for the analysis window size used to select notes for key finding (Figure 2.15(b)). When the window size increases, more notes are included and may affect the analyzed tonality. These hierarchical key analysis diagrams are useful for comparing the impact of using different time scales, and for viewing the harmonic structure\(^{11}\) and relationships between key regions in the composition. Similarly, several analysis methods were discussed but the technical details are omitted here.

Figure 2.15: Visualization in computer music: (a) Self similarity grid of Bach’s Prelude No.1 in C major, BWV 846. (b) Harmonic visualization of the first movement of Mozart’s Viennese Sonatina No.1 in C Major, K545, with logarithmic vertical scale.

\(^{10}\)Tonality: principle of organization around a tonic, or home, pitch, based on a major or minor scale [44]

\(^{11}\)Harmony: simultaneous combination of notes and the ensuing relationships of intervals and chords, which is central to most Western music [44]
The geometry of musical structure for keys and chords is described in [33]. These 2D arrangements of tonality are long-established from which we could consider to base our mapping, such that the encoding scheme will be less arbitrary. Besides, self-organizing maps (SOM) are exploited to visualize the tonal content of a composition [55]. They clearly show how the keys are related to each other (Figure 2.16(a)). MuSA.RT Opus 2 [10] is another tonal visualization of music at multiple scales that uses a spiral array model for tonality layout with dots indicating keys and triads; the closet triads will also light up as a triangle (Figure 2.16(b)). The spiral array represents tonal elements in the three-dimensional space so that perceptually close entities are near to each other. It also shows the evolution of tonal structure over time.

In conclusion, the visualization tools created in the computer music aspect are mainly for visual analysis of detailed musical structure, for example, the tonal content and the chord progression in a composition. The self-similarity grid can show repeated patterns with a proper algorithm; users can observe the difference between two grids rapidly but in-depth identification of the recurring musical units may be missed. As a result, the techniques summarized in this section are useful for music learners to study music theory, but general listeners may not be interested in knowing how the keys are changed, or how the chords are employed throughout the piece. Knowing the overall key of a theme or a passage could be already sufficient for them to excel in comprehending the musical structure.

**Figure 2.16: Tonal content visualization: (a) Self-organizing map. (b) Spiral layout showing tonal evolution.**
CHAPTER 3
THE MUSICAL STRUCTURE DATA

The major musical structural elements discussed in this thesis are layers and themes, as defined below:

Layer. A composition essentially comprises various layers, which are musically known as parts or voices. A layer usually refers to an instrumental part, such as the flute, the first violins, or the cellos. It is determined by musical significance rather than physical section. For example, a group of cellos is equally important to a flute in terms of musical function and each is regarded as one layer, whereas the first and second violins have different functional roles and are thus split into two layers. The audience may not be aware of how the layers are actually played, though they would have a general idea that groups of layers sound distinct from one another from the macro-perspective. The term “layer” is used instead of “track” to provide a generic view of musical parts. For instance, the sound effect “barking dogs” could be a layer in some contemporary music.

Layer Roles. It refers to the functional role of a layer in the harmony of all layers at a time, which contributes to the overall sounding effect of the musical texture. A layer probably plays different roles throughout the music, and some layers may belong to the same functional group at a time. For example, violins are initially dominant and later have a contrast role playing counter-melodies, while the flute and the oboe still share the accompany role supporting other layers. A layer may also have multiple roles like accompany and contrast at the same time, when it accompanies with contrasting materials. Sometimes a layer may be playing transitional materials that do not have any specific role. Overall, layers exhibit some macro-relationships among themselves by having different functional roles over time, regardless of what melodies they play at the micro-level.

Theme. Musical themes, also known as motifs and subjects, are main melodies and musical ideas persisting in the music. Composers usually construct the entire piece with only a few themes since too many ideas would make the music appear to be aimless and lacking in form. Themes are repeated to bring unity to the music, and are contrasted to introduce variety and interest [3].

18
**Theme Variations.** Each theme occurrence is usually varied by melodic, rhythmic, instrumental, and transformative means, forming the micro-relationships. Also, a theme is often played by different layers when it reoccurs, with some solo layers and perhaps other layers as accompaniment and reinforcement, contributing to the macro-micro relationship between layers and themes.

The uniqueness of musical structure and relationships make the visualization a challenging task. Layers form the macro-contexts where they interrelate with different roles. At the micro-level, themes are repeated with different variations along the time-line. Furthermore, the macro layer roles and the micro theme variations are possibly related, resulting in a complicated macro-micro relationship. For example, it would be interesting to know the functional role of a particular theme variation. These coordinated relationships with different levels of viewpoints lead to a challenging visualization problem.
CHAPTER 4

DATA PREPROCESSING

The musical structure data are retrieved manually from descriptive essays, since presently there is no existing algorithm to parse the high-level structural information automatically from MIDI files or raw sound data. This is feasible since there are common, objective agreements for musical structure; the difference between essays is usually in their comprehensiveness but not perspective. It also allows educators to customize the input for teaching purposes, as there are strong human elements in musical structure. Nevertheless, the front-end preprocessing can be automated if such algorithms become available, but it is beyond the scope of this thesis and would not affect how the back-end visualization is generated.

All the examples in this thesis are based on the listening guidelines from an introductory textbook [27]. Without any computer assistance, it takes an amateur with high school music background two days to process the layer and theme information of the first movement of Beethoven’s Symphony No. 5, which include the beginning and ending beat of an interval, the instruments involved, the layer roles, and the theme variations. The most demanding part is, surprisingly, not making musical judgements but digitalizing the data from the score. It is therefore estimated that, with a simple graphical interface that allows users to mark on a digital score as they always do on sheet music, the preprocessing time can be dramatically reduced to at most a few hours.
CHAPTER 5

DESIGN PRINCIPLES

Since the score is already a special visualization of music with demanding notations, we aim to propose an intuitive design which is not yet another functional but overwhelming new language for penetrating musical structure. Challenges lie in how the design can strike a balance between aesthetics and effectiveness. In general, there are four design principles for the visualization of the semantic structure in music:

1. **Consistent**: The encoding scheme should be consistent in all visualization prototypes such that users would not be confused or misled. Thus, the basic visual channels should be carefully assigned.

2. **Intuitive**: Novice users should be able to comprehend the visual displays rapidly without much training. This could be accomplished by introducing some familiar metaphor in the visualization.

3. **Effective**: The underlying musical structure should be shown effectively for novice users to gain insights into the structural features.

4. **Aesthetic**: Being a visual representation of some beautiful and sophisticated compositions, the display should be of high aesthetic to reflect both the artistry and complexity of music.

The design starts with layers and themes, which are the most important and persistent structural components in the data. Apart from being intuitive and efficient, their encoding schemes should also be highly general and flexible as building blocks of diverse visualization prototypes. For instance, while an instrument icon can naturally represent a layer, it considerably limits the design of visualization tools as an icon needs to take a block of pixels to become visible, and does not fuse well with other visual channels such as color and shape. An intuitive color scheme is therefore used to represent layers in our visualization, which incorporates well with different visual designs. For themes, instead of listing the variations with filled boxes and plain texts, the visual representation should convey the semantic meanings of the variations effectively so as to give a feel of how
a theme is varied. It should also allow users to spot which variations a theme applies, and which themes apply the same variations easily. Therefore, a specific pictorial glyph grammar with typographical concepts is proposed to encode the important theme entity in our framework, so that each variation can be intuitively mapped to an iconographical feature of a configurable theme glyph.

5.1 Color Scheme for Layers

5.1.1 Music and Color

The correlation between music and color has been of mutual interest to researchers from various fields. In spite of the many works on music and color, none can be directly applied to the visualization of musical structure data as the semantic context has not been addressed.

Tone Color. In music, the color of a tone, also known as timbre, refers to the quality of a sound. Even if two instruments play the same musical note at the same amplitude for the same length of time, one can still easily distinguish them as different instruments have unique timbres. Conceptually, tone color is associated with sensation such as “warm” and “mellow” rather than optical colors. Thus, there is no explicit mapping from tone color to the optical spectrum. Also, an instrument can have different tone colors depending on performance styles, making it impractical to represent any structural data.

Psychology. In music psychology and psychoacoustics, researchers are particularly interested in how colors are related to music in terms of emotions invoked, such as joy and sadness. They study the human perception of sound which is again not relevant to musical structure.

Physics. Sound pitch can be encoded by a physical mapping between sound and color spectrum. For instance, the color of noise is loosely determined by mapping sound frequencies to equivalent light wave frequencies. Similar mappings exist for people suffering from a form of Chromesthesia called color hearing, which they see consistent colors when they hear the same musical notes. This kind of mapping is rational but lacks perceptual justification in visualization.
5.1.2 Proposed Color Scheme

For effective visualization, the use of color should be coherent and unambiguous. We thus assign colors to layers as traditional orchestral instruments can be divided into four families, string, woodwind, brass, and percussion. This categorial nature is ideally represented by colors of unique hues, with lightness distinguishing instruments in the same family [6]; the lower the pitch range, the darker the color. Although we can randomly assign the hues, it is less desirable since confusing emotions could be invoked by a misleading color assignment. For example, if strings are represented in blue, listeners might expect some sad violin sounds even for a cheerful violin solo. Instead, the physical color of the instrument families is used, as it is objective and intuitive that users can easily recognize the instrument without memorizing the hues. It follows that strings are in reddish brown, woodwind are silver, and brass are golden (Figure 5.1). Percussion instruments, however, do not share a common color like the other families. Hence, some degree of arbitrary assignment like mapping to an outlier purple color is feasible, because they are defined as those not belonging to the other three families. Similarly, non-orchestral layers can be assigned some distinct colors based on users’ preferences.

5.2 Theme Glyph Grammar

ThemeRiver [22] proposed a river metaphor to depict thematic variations in document collections, but it is less intuitive to represent musical themes since the semantic context of musical variations is lost. Contrarily, we use an iconographic glyph metaphor to represent each theme occurrence. However, existing glyph encoding schemes [23][9][31] and icon designs [35][32] are not appropriate as the qualitative meanings of musical variations cannot be conveyed effectively with these quantitative methods.

We first experimented with different glyph designs to determine the perceptually best candidate to represent a musical theme, including primitive shape, human figure, cartoon
head, concrete object, and symbol (Figure 5.2(a)). Primitive shape is simple and functional but the cognitive linkage with music is minimal, while an appealing human figure, cartoon head, or object could be better candidates for entertaining presentation and education. Symbol generally receives positive feedback, yet it does not have clear advantages over other designs. It is concluded that there is no universal design which is outstandingly the best to represent musical theme. Thus, users can import any glyph design and our system will configure the glyph to symbolize theme variations according to the proposed glyph grammar. In this paper, a musical symbol is chosen to be the default theme glyph to deepen the connection with music (Figure 5.2(b)).

The proposed theme glyph grammar for five basic types of variations is as follows. It defines a glyph based on several typographical concepts (Figure 5.3) such that any glyph design can be configured conveniently. Also, only mutually exclusive variations, like lengthen and shorten, are assigned to the same visual channel to encode all possible combinations of theme variations.
1. **Theme** (*e.g.* theme 1, theme 2)
   **Duplicate a sub-part of the glyph**
   As the number of themes is usually limited, it is feasible to duplicate a glyph part to indicate which theme it is.

2. **Form** (*e.g.* form A, form B, form C)
   **Change the core ‘x-height’ shape**
   A theme may be further classified into different forms. This categorical attribute is encoded by changing the shape of the core ‘x-height’ part, such that the overall glyph appearance is preserved with obvious changes to the major ‘x-height’ component only.

3. **Transformation** (*e.g.* augmented, diminished, inverted)
   **Transform the core ‘x-height’ part**
   A theme may be structurally transformed into other shapes. For instance, a theme is augmented by enriching it with extra musical notes and patterns, instead of simply elongating it. Again, only the ‘x-height’ part is modified as it is otherwise difficult to remain similar to other glyphs if the whole glyph is transformed. It also ensures a consistent type size for better legibility.

4. **Duration** (*e.g.* lengthened, shortened)
   **Adjust the ascender or descender part**
   The duration of a theme may be altered due to changes to rhythmic pattern or partial appearance. It is denoted by adjusting the length of the ascender or descender to reflect this one-dimensional variation.

5. **Singleton** (*e.g.* echoed, transposed, in different key)
   **Add indicative decorations**
   Some special variations are more like tags than actually varying the theme. These singleton variations are represented by indicative decorations that can cope with more information if available, such as the relationship with the new key or the degree of transpositions.
Typography is the study and process of typefaces, including how to select, arrange, size, and modify type glyphs, so that type is integrated with other design elements for a complete composition. Instead of proposing some arbitrary design rules, we establish a glyph grammar with typography so that any glyph type can be varied easily with sound design and cognitive foundation for an informative representation. In typography, glyph type generally communicates on three levels [13]:

1. **Content**: The written meaning of the word
2. **Visual impact**: Use of type as a design element
3. **Context**: The content and visual impact of the text in relation to the entire composition

In our framework, content is determined by the choice of glyphs to address different visualization purposes from presentation to analysis, whereas visual impact is achieved by configuring the glyphs according to the proposed grammar. The context level is then reached when the theme glyphs are used as building blocks to construct different visual designs as in Section 6.2. Eventually, by following the typography principles, content is clearly defined, visual impact is formally applied, and context to be delivered is maximized. In fact, the proposed glyph grammar can be generalized to any glyph designs for optimal visual impact and legibility of the visualization. Also, the similarities between type glyphs and general glyphs may indicate a possible future direction to introduce more well-established design formulations from graphical design into information visualization.

![Figure 5.3: Basic type anatomy in typography [13].](image-url)
CHAPTER 6

VISUALIZATION PROTOTYPES

With the careful encoding designs for layers and themes, numerous visualization prototypes can be built with these two basic structural units to reveal the macro-relationships among layers, the micro-relationships of theme occurrences, and the macro-micro relationships between layer roles and theme variations. Among the many possible visual modules, the final candidates in our framework should be intuitive and effective in revealing the musical structure, as well as being able to arouse aesthetic sensation to reflect the beauty and sophistication of music. We thus propose two modules, namely the layer braid and theme fabric, which satisfy the above criteria by being founded on a textile art metaphor. We observe that the vertical structure formed by layers and the horizontal foundation contributed by themes share some common features with weaving and knitting. A piece of delicate cloth is generated by interlacing and intertwining threads along both the $x$- and $y$-axis, while the weaving patterns are still clearly perceived for examining the textile structure in this sophisticated art piece. Also, this visualization metaphor enables novice users to comprehend the displays rapidly without being cognitively overloaded by complex encoding schemes and unfamiliar visual layouts.

6.1 Layer Braid

Inspired by plaiting, layer braid uses a braid analogy to visualize how different threads, representing different layers by the color scheme discussed in Section 5.1, interact and correlate with each other over the horizontal time-line (Figure 6.1). Initially, the threads are laid down one-by-one following the musical score order of woodwind, brass, percussion, and string section. The intertwining process then begins and the braid patterns are determined by the interweaving positions in 2D plane, which are in fact the control points for thread interpolation. In our analogy, the $x$-position corresponds to the time and the $y$-position refers to the current layer roles of the thread, so that layers belonging to the same functional group would be plaited together to create informative braids (Figure 6.2). The order of roles can further impose some semantics of the layer relationships. The dominant and ensemble roles can be placed aside, encapsulating the more supportive accompany,
Figure 6.1: Layer braid for the first movement of Mozart’s Symphony No. 40. Layer threads are plaited according to their functional roles over time. The threads are blended ($\alpha = 0.7$) to better reveal the braid pattern.

Figure 6.2: Layer braid prototype. Layer threads interwind to show how layers interact at the macro-level. Thread clusters suggesting functional layer groups are resulted over time. Control points for interpolating the cello thread are marked. Notice how clarinet behaves exceptionally compared with its woodwind peers (silver) in this example.

contrast, and collaborate roles in between for better visual impact. When a layer has more than one functional roles at the same time, the thread is spread among several streams which plait at the respective locations and rejoin afterwards. Moreover, segments not having a specific role can be darkened upon user request to eliminate confusion caused by threads crossing some roles that they do not belong to during interpolation.

More interactive techniques can be introduced to the basic braid layout. Visual enhancements and filtering in parallel coordinates [21][45][29] can be adopted to better convey the braid structure from the cluttered display. Users may specify the opacity value of each thread for different levels of detail (Figure 6.3). Existing focus+context approaches can also be deployed to the long braids created from music that was originally printed on hundred pages of musical score (Figure 6.4). Most importantly, the dynamic nature of music can be visualized by animating the braids. Simple highlighting of the threads can
Figure 6.3: Different levels of detail for the layer braid in Figure 6.1. By adjusting the thread opacity values, users can better explore the cluttered display such as to highlight the woodwind instruments in (a) and the strings in (b).

Figure 6.4: Deploying some focus+context technique like bifocal display [2] to the long layer braid for better navigation.
indicate the current playing intervals. Alternatively, the braids may appear in accordance to the music being played, simulating a flow of musical threads that illustrates how the layers interact in real-time.

6.2 Theme Fabric

While the layer braid visualizes the macro-interactions of layers with intervening threads, theme fabric aims to reveal the micro-relationships between theme occurrences across layers. A theme not only varies, but is also played by different layers when it reappears. As a result, these micro-relationships are found along both the $x$-axis over time and the $y$-axis for different layers.

![Figure 6.5: Theme fabric prototype. The glyphs are first placed in a 2D plane according to when and where the theme appears at the micro-level. The same glyphs can then be connected aesthetically in various weaving styles.](image)

The theme fabric is first constructed by overlaying the theme glyphs, which are encoded by the grammar explained in Section 5.2, on a conventional musical score layout according to which layers play the theme and when the theme is played (Figure 6.5). If a theme occurrence is played by more than one layer, the glyph is duplicated at every layer that plays the theme. To avoid glyphs overlapping, the space between the glyphs is scaled non-linearly such that a minimum glyph gap is guaranteed. This is applicable since the relative gap proportion is more useful than the actual gap distance in understanding thematic structure when we listen to the music. Typography also suggests that letterspacing would affect legibility of the text, because it would be easier to distinguish one letter from the next with sufficient space. Therefore, slight adjustment to the spacing is made to ensure that every glyph is equally visible while retaining a relative gap proportion.

Although themes having the same variations are represented by the same glyph, the repetitions are still not obvious in a large graph without strong visual clues to connect the glyphs together. Originating from the textile metaphor, glyphs can be treated as stitches connected by threads in various styles, weaving a sheet of theme fabric that effectively
reveals how the glyphs are organized in a 2D plane. We include three types of stitch in our framework, namely plain style, bundled style, and collapsed style to provide decreasing levels of detail of the micro-relationships of theme variations.

![Figure 6.6: Theme fabric styles in textile art. (a) Plain weaving. (b) Butterfly stitch in knitting that bundles yarns. (c) Open strips produced by drop-stitch technique.](image1)

6.2.1 Plain Style

The most straightforward way is to connect repeating glyphs in the orthogonal direction only. That is, identical glyphs of a layer are chained horizontally, and those being played by different layers within a short time interval are linked vertically. This basic layout is an analogy to plain weaving in textile art, where two sets of threads, known as warp and weft, are interlaced to create a fabric (Figure 6.6(a)). The horizontal weft threads are woven under and over, back and forth through the vertical parallel warp threads. Consequently, glyphs across layers are connected with straight yarns in our plain weaving

![Figure 6.7: Theme fabric with plain style weaving.](image2)
metaphor, and those appearing at the same layer are woven by curve yarns interwoven with the vertical lines (Figure 6.7). Moreover, each set of identical glyphs is stitched with threads of unique lightness for a clearer weaving pattern. The plain theme fabric serves as the most fundamental and expanded view of thematic structure at the micro-level as every detail on theme variations and repetitions across layers are literally visualized in a plain weaving style.

Figure 6.8: Generating bundled style from plain style theme fabric. Vertical threads (warp) are removed and horizontal threads (weft) are bundled, retaining the betweenness centrality of the theme glyphs with a concise layout.

6.2.2 Bundled Style

As a theme is usually played by various layers at the same time, the glyphs are duplicated at these layers and are connected by replicated horizontal threads. The plain style may therefore be enhanced by bundling all horizontal weft threads originating from glyphs that are initially connected by vertical warp threads, creating a less cluttered view while maintaining the betweenness centrality of the theme glyphs (Figure 6.9 and 6.8). However, threads would not be bundled for successive glyphs that are not repeated elsewhere. Alpha blending of threads are included to tackle the overlapping of thread bundles and knots. Moreover, line patterns may be used to outline special glyph groups like orphan theme occurrences by double dashed lines.

The bundled style is in fact derived from butterfly stitch in knitting (Figure 6.6(b)). It is an abstract form of the plain style since the weft threads are bundled and the warp threads are removed. Although repetitions at the same layer are less obvious, the bundled view effectively reveals the more high-level thematic structure by visually clustering all identical glyphs across layers. The actual musical structure is further highlighted by thread bundles and knots which figuratively exhibit the sophistication of theme and variations at the micro-level.
Figure 6.9: Theme fabric in bundled style for the first movement of Mozart’s Symphony No. 40. Each theme occurrence is represented by a musical symbol glyph encoding its variation. Identical glyphs are connected by bundled threads.

6.2.3 Collapsed Style

Conceptually, the collapsed style is very similar to drop-stitch knitting that can produce any number of open stripes by dropping stitches intentionally (Figure 6.6(c)). Apart from clustering similar threads as in the bundled view, the numerous redundant glyphs can be further eliminated by collapsing identical glyphs at different layers into one single glyph only, resulting in some Arc Diagrams [58] layout (Figure 6.10). As the original design cannot accommodate multiple tracks in one display, we propose to encode the layer information by rainbow arcs that are divided into color bands based on the layer color scheme. A band is colored if the corresponding layer contributes to the current theme occurrence. The arc is split into two quarters to show the layer details of the former and latter occurrences respectively. In contrast to the original Arc Diagrams where the arc height equals the duration between two repetitions, this visual channel is further exploited by assigning a specific arc height to every unique glyph, so that users can locate all repetitions of the same theme glyph efficiently.

The arc height can be randomly assigned for showing solely thematic structure, or seamlessly combined with the spatial order of layer roles as in the layer braid display to reveal the interesting macro-micro relationships between layers and themes (Figure 6.11). Theme glyphs that are correlated with certain layer roles would have their arcs frequently appear in the roles’ allocated partition. Similarly, layers that are strongly related to a
certain functional role would have their respective colors often filled in the rainbows at a particular height. If a layer has multiple roles at the same time, duplicated arcs at different heights may be drawn.

### 6.2.4 Summary

In summary, the three theme fabric styles provide different levels of detail on the micro-relationships by connecting and manipulating the theme glyphs with reference to several textile art techniques, in order to aesthetically express the sophisticated nature of theme and variations in music. The plain style contains the expanded thematic structure, while the bundled view highlights groups of identical glyphs by thread bundles. The collapsed style is the most abstract view of thematic structure with compressed layer information. By mapping layer roles to arc heights, it can further reveal the unique macro-micro relationships on how themes and layers correlate in a compact display.

Figure 6.11: Collapsed theme fabric for a segment of the first movement of Beethoven’s Symphony No. 5. Macro-micro relationships are shown by mapping arc heights to layer roles. In this example, no theme occurrence is associated with the dominant or accompany layer roles. Pale reflection of the arcs is added for visual reinforcement.
CHAPTER 7

USER INTERACTIONS

Various user interactions are implemented in our system to aid visual analysis and knowledge discovery among different visualization prototypes [7]. The system takes a MIDI file and its respective layers and themes information in a pre-defined text format as input. The basic layer braid and theme fabric displays are then automatically rendered, on which users can further interact to reveal the sophisticated structure of the input music.

7.1 Customization

Users may configure the layer colors and the theme glyph type via a graphical interface, while the proposed color scheme and the musical note glyph are used as default. In addition to the pre-loaded theme glyphs, users may import their own glyphs from which the system will vary according to the glyph grammar.

7.2 Zooming and Panning

Basic zooming and panning are provided, but users more often utilize the focus+context tool for overview and navigation, with selection and filtering for details inspection. Thus, awkward zooming and panning that cause users to lose sight of the original context are minimal.

7.3 Selection

Each visual entity, such as a layer thread and a theme glyph, can be selected directly from the display to alter its behavior. Attached properties like track volume or theme duration are also shown for the selected item. Furthermore, when users point at a theme glyph, all relevant theme repetitions will also be chosen for rapid selection from the comprehensive visual modules.
7.4 Filtering

Users may adjust the opacity value of the layer threads, theme bundles, theme glyphs, and rainbow arcs to filter insightful visual patterns from the complex layout. In particular, hierarchical slidebars and transfer function windows are implemented to facilitate the opacity specification for the layer braid. Apart from setting the opacity value, users may overshadow the parts they are not interested in, or contrarily highlight some special structural features. For instance, thread segments not having any functional roles can be darkened, while the orphan theme glyphs can be spotlighted.

7.5 Linking and Brushing

Users can use simple brushing to interactively explore different visual modules. For example, when they select a segment from the layer braid, the corresponding part in the theme fabric is brushed. Moreover, users may overlay the prototypes for some interesting effects given their similar layouts. The layer braid can be fused with theme glyphs, or embedded to the arc height stack representing layer roles in the collapsed theme fabric to generate an inclusive visualization of the macro-structure with the layer braid, micro-relationships with the theme fabric, and macro-micro confluence with rainbow arcs at semantic heights.

7.6 Abstraction

Visual abstraction of the theme fabric is achieved by rendering in different styles, or by retaining only a sketchy outline of the theme glyphs. For smooth transition, fading is added when the theme fabric style is changed. Abstraction of the layer braid can be performed at the data level, where short intervals that are not musically significant can be aggregated to eliminate the abrupt zig-zag leaps during thread interpolation.

7.7 Focus+Context

Focus+Context is critical for navigating the visualization of lengthy musical compositions, as users often wish to maintain a context of the overall musical structure while examining some interesting details. Bifocal display [2] [38] is currently employed since it is similar with browsing sheet music. Other focus+context techniques can be added to boost the interactivity. Also, the focused area will be shifted accordingly when the music is played.
7.8 Details on Demand

Apart from showing the details of a layer thread or a theme glyph when it is selected, more information are also provided if users zoom into the visual entity. For example, instrument icons, lists of variations, or the actual musical score can be shown on demand.

7.9 Synchronization with Music

The visualization is synchronized with music in real-time. Users can set the playing position and range by a progress bar. Moreover, the animated layer braid may flow with the music concurrently, likewise the theme bundles can appear according to the music. For the collapsed theme fabric, not only the rainbow arcs emerge along with the music, the current arcs may also be rippling to reinforce thematic structure.
CHAPTER 8

CASE STUDIES

In this chapter we demonstrate how the proposed visualization prototypes, the layer braid and the theme fabric, express visually the sophistication of musical structure in terms of layer interactions and theme variations, from which the general audience can quickly gain insights into the semantic structure without going through the overwhelming technical essays. The analysis tends to be more general to avoid specific knowledge on classical music, but it already provides many interesting observations on musical structure and compositional styles. Classical symphonies are examined as their clear-cut forms are relatively easier to understand. Nevertheless, our techniques are not limited to orchestral music but are capable for handling any kind of music.

8.1 Overview

The analysis is based on two famous symphonies, Mozart’s Symphony No. 40 in G minor, K.550 and Beethoven’s Symphony No. 5 in C minor, Op. 67. For illustration purpose, only the first movement that is commonly regarded as the signature of a symphony is discussed. The first movement is often built up in sonata form including exposition, development, and recapitulation. The exposition and its repetition are found at the beginning to expose the basic musical ideas; more complicated relationships are seen in the middle development section and the music eventually returns to some familiar materials in recapitulation. It definitely requires some effort for untrained ears to recognize sonata form with plain text. With the proposed visualization, novice users can also understand the semantic structure easily and obtain a general overview of the composition.

8.2 Macro-Relationship

The macro-relationship of layers for Mozart’s Symphony No. 40 is first studied (Figure 6.1). Remarkably the development section exhibits highly complicated relationships with rapid exchanges of roles and multiple roles at the same time. Overall, ensembles are only heard once at the end of exposition. More specific patterns can be obtained by
modifying the thread opacity values for different levels of detail (Figure 6.3). Lower strings including violas, cellos, and double basses mostly play the accompaniment especially when the violins are dominant. Instruments of the same family tend to behave similarly, except bassoon which becomes dominant more frequently compared to other woodwind instruments.

The layer braid also figuratively illustrates Mozart’s talents in handling an orchestra at the apex of the mature Classical era, including the frequent interchanging relationship between strings and woodwinds, the rapid alternating predominance of instrumental group combinations, and the careful sharing of musical interests across layers [1]. In effect, the compact layer braid display intuitively portrays the sophistication of musical texture with a familiar braid analogy to novice users.

8.3 Theme and Variations

As compared to the previous macro-interactions of layers that generates musical texture, the horizontal framework of repeating theme provides an alternative perspective for understanding the micro-structure of melodies as shown by the corresponding theme fabric (Figure 6.9).

The overall textile patterns in bundled style gives a lasting impression of thematic structure, with the thread bundles depicting how the themes are structured in the entire piece. The sonata form is usually framed by two subjects or themes in contrasting styles. The bluish thread bundles indicates that the principal subject is mostly played by the strings, and is occasionally heard in echo among clarinets, bassoons, and violins following the second subject stitched with orange yarns. Its normal form without variations clearly marks the beginning of exposition, repeated exposition, and recapitulation, suggesting how the music is restated in sonata form. The transposed principal subject not only appears in exposition, but also directs the music into the development section in violins and violas only. Vivid variations that never repeat are seen in development and recapitulation, such as shortened, augmented, or in different keys. These orphan variations with double-lined threads imply how the composer built up musical variety and tension towards the ending climax. Ultimately, the shortened principal subject from the development section returns in shift in the basic key after some key changes in recapitulation, ending the music in unity.
Overall speaking, strings are the primary theme players in this Mozart’s symphony. Woodwind instruments are less significant; in particular oboe contributes to only one theme occurrence. Interestingly, though the score also calls for horns as seen from the layer braid, they do not participate in the thematic structure at all. It aligns with the historical development of orchestra that the string section was the foundation of a typical Classical orchestra, and brass had only become increasingly significant in early 19th century under Beethoven’s influence.

With the informative layer braid and theme fabric, abstract compositional styles become visible: while Mozart’s melodies and forms are mostly graceful and elegant, the complexity of his music perhaps comes from his genius sensation of instruments that makes every simple micro-modules an essential contribution to the macro-texture, producing these beautiful pieces of music.

8.4 Macro-Micro Confluence

After studying the macro-relationships of layers and micro-structure of theme repetitions independently, we may analyze the interesting macro-micro confluence of layers and themes with the collapsed theme fabric (Figure 6.11). A segment of the first movement of Beethoven’s Symphony No. 5 is visualized, with arc heights indicating the respective layer roles of the theme. The display focuses on the famous “Fate knocking at the door” motif; the lyrical second theme is thus not connected with arcs. In addition, orphan theme variations that are not repeated in this segment are highlighted in yellow.

Surprisingly, the fate motif never acts as a typical principal melody that dominates the music as in other Classical symphonies. While the motif in ensembles begins and ends both sections in different variation forms, it mostly serves as a contrasting or echoing component throughout the music. In brief, there is no sole melodic line in this example, which is one of Beethoven’s pivotal compositional styles. Unlike his predecessor Haydn and Mozart, the motif is usually some forceful and marked rhythmic patterns rather than conventional melodic figures. These rhythmical materials are monumentally varied, especially in the development section where numerous new forms of the motif are heard successively. On the contrary, Mozart’s music is characterized by his elegant melodic themes with a clean and balanced structure as revealed from the bundled theme fabric.
Although the layer information are compressed in this abstract view, detailed orchestration is presented with splendid colors together with the macro-micro confluence of layers and themes. The rainbow arc patterns imply that woodwind and brass usually collaborate at the macro-level by playing the augmented fate motif at the micro-level, whereas the inverted, shortened motif in the lower strings is used to contrast the music. Also, the form B motif in a new key is specially outlined by the violins near the end of the development section, preparing for the restatement in the recapitulation section. Generally speaking, the functional roles in this Beethoven’s example are more aligned with the motif rather than the high-level layer interactions, which follows that the boldness of Beethoven’s music is attributed mainly to his extensive development of musical materials.
CHAPTER 9

EVALUATION

Our system has been evaluated by both the general audience and music experts. We performed a user study with the general audience to evaluate quantitatively the effectiveness and efficiency of comprehending musical structure with the proposed visualization. A thorough user survey is also conducted with several music students and experts to collect their qualitative feedback on the visualization framework.

9.1 User Study

20 college students without prior musical background participated in the user study, which aims to validate the effectiveness and efficiency of the proposed encoding schemes in conveying musical structure.

9.1.1 Tasks

After a 5-minute introduction, the subjects were asked 12 questions regarding musical structure similar to the exercises from music textbooks. For the layer braid, users were questioned about the layer roles over time and the relationships between functional groups, while for the theme fabric they were required to answer how, when, and where the theme appears. The macro-micro interactions between layer roles and theme variations were also tested with the collapsed theme fabric.

9.1.2 Results

The accuracy and response time are recorded in Figure 9.1. The average response time for the bundled theme fabric is the longest since frequent user-interactions are required with such an intensive display, but the noticeably high accuracy, with three quartiles above 90%, proves that thematic structure is delivered effectively despite a slightly cluttered layout. Both the layer braid and the collapsed theme fabric are fairly efficient due to simpler visual designs. The layer braid, however, shows a statistically significant difference with other modules in response time; the lowest quartile gives a wide range of three minutes. It
### 9.1.3 Discussion

In contrast to reading tiring technical essays and musical scores, our visualization prototypes are very effective and efficient for novice users to gain insights into musical structure promptly without prior extensive training. Yet, it is worth pointing out that users may not necessarily be able to connect the structural features they observe visually, or what they read from the essays, with the music they listen to. Musical training is always required for analyzing the music by ears, but the visual modules provide a friendlier pictorial guidance for music learning and appreciation by symbolizing the sophistication and beauty of the music, thus making the music listening experience more enjoyable and informative.
9.2 User Survey

In addition to the user study with the general audiences, we also conducted an in-depth survey with a group of 10 domain experts consisting of 7 music students, 2 music educators, and 1 musician for their expert feedback. Their music experience ranges from 6 years to more than 30 years in music learning, teaching, or performance.

9.2.1 Results

The survey results are summarized in Figure 9.2. The participants shows a very positive attitude to the visualization prototypes. All modules are equally valuable in terms of usefulness, effectiveness, and aesthetics. The theme fabric in bundled style is the most useful display as users are more concerned about thematic structure, whereas the collapsed style derived from the Arc Diagrams layout is the most aesthetic one. In particular, they find the rainbow arcs with reflections strongly appealing.

9.2.2 Discussion

Comments from the music experts are also very encouraging. Overall, they appreciate the familiar analogy to textile art, which intuitively expresses the sophistication and aesthetics of a composition with vivid displays. They believe braid and fabric are excellent metaphor of music since its physical structural patterns and perceptual artistic aspects are both revealed effectively.

![Figure 9.2: Survey results with music experts on a 5-point scale about the usefulness, effectiveness, and aesthetics of each prototype in visualizing musical structure.](image)
Overall

Among all prototypes, the participants like the collapsed theme fabric most because it is aesthetic and less complicated than the bundled style, while retaining sufficient information on layers and themes at a glance. They agree that the layer braid is extremely useful in understanding how different instruments relate and interact in an huge orchestra, as the layer role concept might be less definite to novice users without the visual aid. Participants having substantial musical background find the bundled theme fabric effective in presenting complicated micro-relationships, but less experienced students consider it slightly complex and prefer the collapsed style for different levels of detail. This implies that the proposed prototypes can cooperate with or complement one another in suiting a variety of needs.

Music Educators

Music educators like the intuitive color scheme and effective theme glyph very much, as they believe reading pages of words would possibly diminish a student’s interest in learning a composition. It is also easier for students to follow the music with such a real-time system, which is especially beneficial for visual learners. Moreover, they appreciate that instrumentation, that is the multi-layered property, is addressed in our framework. They find this layer information particularly important, which is however not captured by the TimeSketch program they currently use for learning and presentation.

Music Students

Music students also appreciate the layer feature as differentiating instruments in an orchestral work is difficult. By dissecting an orchestral piece into various layers, it can boost the enjoyment of orchestral works and facilitate a sensible presentation to music audience. They find that the visualization prototypes are especially useful in comparing compositions from different musical periods of distinctive characteristics, which help learners to identify the relevant period and its notable features that are otherwise abstract without the visual aid.
Musician

While the musician holds the philosophy that music should speak for itself, he agrees that these visual displays truly assist music learning and analysis. Classical music can be thus publicized with such an impressive visual guidance that exhibits the sophistication and beauty of the music, from which the general audience will find classical music more approachable and eventually be amused by the music itself. He also looks forward to attaching the visualization to the concert programme, so that audiences can better understand the performance and appreciate the genius ideas behind the music. By gaining insights into musical structure, listeners can advance from solely emotional appreciation to more conscious learning and deeper understanding, resulting in new enjoyment, excitement, and satisfaction.

Concerns

One major concern, as stated by the educators, is that problems may arise for visualizing musical works that do not have a clear form since the prototypes focus on revealing the structure, and for handling exotic instrumentations that do not take the conventional orchestra in contemporary compositions. Given the variety of musical styles, it is in fact impossible for a visualization framework to be capable of handling compositions of any style. Our design is generalized to the greatest extent to strike a balance among simplicity, scalability, effectiveness, comprehensiveness, and depth. Specifically, the layer concept we introduced is not limited to instrumental layers but can refer to any recognizable track.

Improvements

The participants also identify some possible improvements. While the proposed prototypes are well-designed and comprehensive, students are also interested in having more attributes such as the change in dynamics, tempo, and tone color in the resulting displays. This can be achieved by exploiting more visual channels like size and shading in the current prototypes, or by developing new visual modules based on the flexible layer color scheme and theme glyph grammar. Moreover, educators believe that adding some descriptive texts can guide students in discovering the aesthetic part of the music, which is currently not included due to copyright issues. Finally, the music educators conclude that they are eager to apply the visualization in music schools.
CHAPTER 10

CONCLUSION

In this thesis we proposed an innovative visualization solution to show the semantic structure in classical music works, including the macro-level layer relationships, micro-level theme variations, and macro-micro interactions between layers and themes. We first discussed several design principles for encoding the important structural units, layers and themes, with an intuitive color scheme and a sound glyph grammar founded on typographical concepts respectively. Various visualization prototypes can then be constructed flexibly with the layer colors and theme glyphs. Instead of developing some general visual displays, we specifically identified the similarities between music and textile art in their structural and aesthetical aspects. The layer braid and theme fabric prototypes are hence initiated to visualize the semantic relationships in music with an analogy to several common weaving and knitting techniques. Apart from revealing the underlying musical structure effectively, the vivid displays also intuitively portray the sophistication and beauty of the music. While the visual prototypes in essence form a comprehensive framework, they are complementary to one another as well to provide different levels of detail for users with diverse needs and backgrounds. Several representative case studies are then presented in which the musical form is clearly revealed and the complicated semantic relationships are efficiently depicted, providing visual clues to novice users about the abstract compositional styles. The merit of the visualization prototypes is further reflected by the user study that proves the effectiveness and efficiency of the encoding schemes, and by the user survey where highly positive feedback and encouraging comments from music students and experts are received.

Reflection

In summary, we demonstrated the potential of applying information visualization to another discipline that dramatically enhances the understanding of a new data type, the musical structure data. The strong human factors present in the semantic relationships in music distinguishes this work from other information visualization approaches that mainly focus on the quantitative and effective aspects of the data. The qualitative and subjective elements in musical structure poses special challenges to the visualization task
that not only the underlying structure should be conveyed efficiently, but the artistic sensation aroused from its sophistication and beauty should also be expressed intuitively. In our framework, this perceptual issue is addressed by using familiar metaphors from other art fields, including typography in graphical design and weaving in textile art, to generate visual displays for the greatest cognitive pleasure. Well-established design principles from these fields can easily be adopted to create informative and aesthetic displays once the similarities between the data and the techniques are identified. It also indicates that we may further explore other art and design analogies for applications that involve considerable perceptual concerns, such as casual information visualization and semantic data visualization. For example, typographical concepts may be exploited to formulate more comprehensive glyph grammar for web applications, and the textile art metaphor can be deployed to create information graphics for complicated semantic relationships. Therefore, the visualization prototypes presented in this thesis are not limited to revealing musical structure, but can be generalized to visualize other data types. In particular, the semantics in news visualization can be modeled as news themes that vary in context and perspective as reported by different news sources. Musical data also share some common components in story-telling visualization where a layer becomes an actor associated with various dramatic themes and scenes.

**Future Work**

There are several avenues for future work. First, more diverse compositions should be visualized to fully realize the potential of the visual design. Other types of music works from different periods would be continuously added to the database. Second, a simple user-interface that allows educators to directly mark on a digital score could dramatically reduce data processing time, and thus generalize the usage of the visualization. Lastly, we are seeking the possibility of applying the visualization prototypes as teaching tools for some introductory music courses to reach a wider audience.
REFERENCES


