A Report on Musical Structure Visualization

Winnie Wing-Yi Chan

Supervised by
Professor Huamin Qu

Department of Computer Science and Engineering
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

August 2007
Abstract

In this report we present the related work of musical structure visualization accompanied with brief discussions on its background and importance. In section 1, the background, motivations and challenges are first introduced. Some definitions and terminology are then presented in section 2 for clarification. In the core part, section 3, a variety of existing techniques from visualization, human-computer interaction and computer music research will be described and evaluated. Lastly, observations and lessons learned from these visual tools are discussed in section 4 and the report is concluded in section 5.
Contents

1 Introduction 1
   1.1 Background ........................................... 1
   1.2 Motivations ........................................... 1
   1.3 Challenges ............................................ 2

2 Definitions and Terminology 2
   2.1 Sound and Music ........................................ 2
   2.2 Music Visualization ....................................... 2

3 Visualization Techniques 3
   3.1 Arc Diagrams ........................................... 3
   3.2 Isochords ............................................. 3
   3.3 ImproViz ............................................. 4
   3.4 Graphic Scores ......................................... 6
   3.5 Simplified Scores ........................................ 6
   3.6 Performance Expression Visualization ...................... 7
       3.6.1 Vertical Bar Display .................................. 7
       3.6.2 Chernoff Faces ..................................... 8
       3.6.3 Hierarchical Approach ................................ 9
       3.6.4 Music and Emotion ................................... 9
   3.7 3D Music Visualization .................................. 9
   3.8 Commercial Products ..................................... 10
       3.8.1 TimeSketch ....................................... 10
       3.8.2 Music Animation Machine ............................... 11
   3.9 Visualization in Computer Music .............................. 11

4 Discussion 12

5 Conclusions 13

6 Acknowledgements 14

References 14
1 Introduction

We hear different types of music and sound every day unconsciously, ranging from pop songs and background music to sound effects and even noise. Yet, when we listen to the music consciously, it always requires some efforts to knowledgably understand the music, in particular the structure and the techniques applied by the composers, for appreciating the music and comprehending the thoughts behind the work. While pop music is meant to be friendly to the general public, difficulties arise when we listen to classical music. Untrained people may only be able to feel the sound elements such as pitch, rhythm, volume and speed; the actual structure and form of the music are usually, and merely, known by music lovers who have received extensive training in music theory and history. This high learning curve thus makes classical music seemingly sophisticated that many of us find it boring as we know nothing about it. In this report, we restrict the music discussed to be strictly organized compositions that are loosely known as classical music.

1.1 Background

Realizing the structure of music is the important step to appreciating the whole piece. Music expertise is in essence the ability to single out the musical elements constructing the composition by reading the musical scores or even by simply listening to the music performance. As compared to fine art and literature, it is generally harder to learn classical music due to its abstract and temporal nature. For a concrete artwork, one may quickly point out the features and viewers can hence readily identify the global structure plus detailed techniques applied on the static, definite artwork. Similarly, readers can grasp the overview by flipping through the book. Human visual system is so powerful that we need not be trained to perceive graphical patterns. In contrast, it is likely that unskilled ears cannot recognize complicated musical elements from multi-layered music easily. Some consider the lack of visual equivalence is what makes music different from other non-auditory art forms, but this is also why music appears to be more abstract. Another unique characteristic is that music is dynamic and time-varying. Audience can never project music onto one single time frame, making comparisons and appreciation more laborious.

Until now, the most common way to learn music is through studying a musical score, or sheet music, which contains the objective notations of a music composition. However, reading musical scores is demanding. Beginners have to spend considerable amount of time to learn the basics of music theory before being capable to understand the notations; and there is still a long journey to mastering a musical score for in-depth understanding of the composition in an all-round way. Even for amateurs with sufficient music background, it could be tedious to go through the original musical scores for solely listening purpose. It contains rich technical details useful for performance, but could be overwhelming for non-performers to seek structural units for appreciation. In practice, it is not convenient to obtain the full score as well.

1.2 Motivations

To lower the learning curve of classical music, visualization is one of the most promising approaches to aid the audience, thanks to our strong visual cognition ability. Although numerous efforts have been devoted to the visualization of the sonic characteristics that are available in many commercial players like Apple iTunes and Windows Media Player, few attempt to visualize the structure of the music instead of the frequency variation of sound in the physical space. Furthermore, the structure of music investigated in these few research projects usually limits to one musical element, and the multi-dimensional nature is rarely addressed.

With the advance in computer graphics, we can now rapidly generate beautiful artistic representation of sound and music. The major research issue becomes how to create informative and insightful visualization from which users without much musical background not only see the sound visually, but also understand the structure of music semantically and be able to identify the musical elements effectively.
1.3 Challenges

The major challenges of visualizing the musical structure lie in both the pre-processing of the input data and the design of a reasonable graphical representation for music.

**Defining the data structure.** Classical music comes in diverse forms having their own special features. A general data structure which can capture the essential musical elements of the largest number of compositions should be defined. We may also work on only a certain form such as symphony, concerto, sonata and fugue, but the main problem is again to decide what features to be included as variations also occur within one genre\(^1\).

**Choosing the analysis method.** Automatically analyzing MIDI (Musical Instrument Digital Interface) files could be desirable as the only concern about input is to ensure that the MIDI file, a digital musical score, is accurate. However, it will be challenging to derive algorithms to analyze the cognitive, not just the mechanical, musical elements from the score. These descriptive parameters help audience to appreciate the music qualitatively. Although most of them can be quantified, it is not known whether one could retrieve them entirely via automatic algorithms. Also, the algorithms must be validated to be held universally, which might not be possible due to the complexity and variety of music. On the other hand, if manual analysis method is applied, the challenges are shifted to how we should extract useful information from music literature and what user-interaction we should provide to facilitate these human-defined input.

**Designing effective visual encoding scheme.** It is still uncertain what the best visual mapping for music is. For each music attribute we need to find the optimal representation that is both intuitive and straightforward to the users. Moreover, these graphical elements should not interfere with each other when combined to form the complete visualization of a music composition, as in general multivariate data visualization.

2 Definitions and Terminology

2.1 Sound and Music

Strictly speaking, sound does not have specific meanings while music is a collection of organized sound that aims to convey certain messages or emotions. In other words, music is an art form consisting of sound and silence expressed through time [43]. Sound qualities include frequency (pitch), volume, speed, spatial position, etc. For music, however, musical properties like rhythms, chords, tempo\(^2\), dynamics\(^3\), timbre\(^4\), texture\(^5\) and harmony are undoubtedly more important.

2.2 Music Visualization

The term Music Visualization is often associated with the visualization of the loudness and frequency spectrum of the music being played in real time [44], which ranges from simple oscilloscope display on a radio to animated imagery rendered by media player software. Therefore, the term Musical Structure Visualization is used instead to distinguish from music visualization.

---

\(^1\) **Genre**: general term describing the standard category and overall character of a work [28]  
\(^2\) **Tempo**: rate of speed or pace of music [28]  
\(^3\) **Dynamics**: the loudness or softness of a composition [38]  
\(^4\) **Timbre**: quality of a sound that distinguishes one voice or instrument from another, also known as tone color [28]  
\(^5\) **Texture**: vertical structure of the music, i.e. how many parts or voices there are and how the voices interact [38]
Figure 1: 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 Beethoven’s *Für Elise*.

3 Visualization Techniques

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.

3.1 Arc Diagrams

Arc Diagrams [41][42] visualize complex patterns of repetition in string data by connecting a translucent arc between a pair of matching pair (Figure 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*.

3.2 Isochords

Isochords [1][2] apply Tonnetz, a two-dimensional triangular isometric coordinate grid invented by Euler (Figure 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(b)). An edge is drawn between two notes only if they are consonant\(^6\) to highlight these more important intervals. Chord progression\(^7\) is encoded with adjacent structure in the display, and color may additionally show the modulation\(^8\). The size of the triangle vertices denotes further the octave of the note with a larger dot referring to a lower octave. The inversion\(^9\) of chords is not explicitly handled though.

---

\(^6\)Consonant: harmonious combination of tones that provides a sense of stability in music [28]
\(^7\)Chord Progression: the movement from one chord to the next [38]
\(^8\)Modulation: the process of changing from one key to another [28]
\(^9\)Inversion: the position of a chord when the root position is not the lowest note [38]
While chords are important in music analysis, the more general descriptions such as phasing, orchestration\(^{10}\) 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 (in section 3.9), yet the latter, with a huge number of potential users for the purpose of music appreciation, is rarely studied in a formal context.

### 3.3 ImproViz

ImproViz\(^{37}\) is a visualization technique showing jazz improvisations\(^{11}\) 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 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.

\(^{10}\) *Orchestration*: technique of setting instruments in various combinations\(^{28}\)

\(^{11}\) *Improvisation*: creation of a musical composition while it is being performed\(^{28}\)
Figure 3: ImproViz example; melodic landscapes are shown on top, and the bottom are harmonic palettes.

Figure 4: The graphic study score of an electro-acoustic composition *Incantation* (1953) by Otto Luening and Vladimir Ussachevsky.
3.4 Graphic Scores

In [7], 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 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 spacial 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 [8][9], 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.

3.5 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 [27]. 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 visualize tempo and other specifications (Figure 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.

![Figure 5: The simplified and compressed score: (a) Initial overview of the whole piece. (b) With fisheye for displaying the original score.](image-url)
3.6 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 [21][29] 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 [12][13][14][15][26][27] 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.

3.6.1 Vertical Bar Display

In [14], quantifiable expression elements include tempo change, articulation\(^{12}\), 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 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 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\(^{13}\) passage results in a series of rectangles connected one by one, whereas a staccato\(^{14}\) 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.

---

\(^{12}\)Articulation: characteristics of the attack, duration, and decay of a given note [38]

\(^{13}\)Legato: smooth and connected notes [28]

\(^{14}\)Staccato: short and detached notes [28]
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 6(c)) additionally show us the player’s phasing.

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.

### 3.6.2 Chernoff Faces

Chernoff faces [3] were used to visualize the performance expression in [12][27]. The tempo, articulation and dynamics change of each note are mapped to the eyeball position, the contour of the face and the shape of nose of the corresponding face (Figure 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)

Figure 7: Visualizing musical expression using Chernoff faces.

![Circular form](image) ![Cone Trees layout](image) ![Viewing from the top](image)

Figure 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.
3.6.3 Hierarchical Approach

A hierarchical approach was proposed [15][25][26] towards visualizing musical structure. A circular form is first applied to display the global structure of a long sequence. Then Cone Trees [32] layout is used to visualize the musical structure of recursive hierarchies including form, section, phrase and motif (Figure 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.

3.6.4 Music and Emotion

[13] derives a graphical expression of the mood of music (Figure 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 [24]. 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 [22] for deciding the emotion of music from the underlying musical elements. [6] presented a different emotional data extracting algorithm and mapped the emotion to facial expressions.

Figure 9: An example of a graphical expression of the mood of music.

3.7 3D Music Visualization

[35][36] 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.

The comp-i (Comprehensible MIDI Player - Interactive) system [15][25][26] allows visual exploration on MIDI data in an immersive manner (Figure 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 [19][30][40]. They provide an intuitive way for users to manipulate the sound produced to achieve some music performances in the virtual environment (Figure 11). Obviously the potential of visualizing semantic musical structure and elements in the 3D space has not been well explored yet.

3.8 Commercial Products

3.8.1 TimeSketch

TimeSketch [17] 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 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 [16].
3.8.2 Music Animation Machine

In the basic bar-graph of Music Animation Machine [23], 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 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.

3.9 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 [10][11] is deployed to visualize the time structure of musical waveforms which helps comparing audio recordings for their acoustic similarity (Figure 14(a)). A composition is often self-similar which contains several repeated passages or variations of themes. A 2D array based on dotplot [5] 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 [33][34] maps tonality\(^{15}\) 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 14(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\(^{16}\) and relationships between key regions in the composition. Similarly, several analysis methods were discussed but the technical details are omitted here.

The geometry of musical structure for keys and chords is described in [20]. 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 [39]. They clearly show how the keys are related to each other (Figure 15(a)).

\(^{15}\text{Tonalit}\text{y}: \text{principle of organization around a tonic, or home, pitch, based on a major or minor scale } [28]\)

\(^{16}\text{Harmony}: \text{simultaneous combination of notes and the ensuing relationships of intervals and chords, which is central to most Western music } [28]\)
Figure 14: 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.

Figure 15: Tonal content visualization: (a) Self-organizing map. (b) Spiral layout showing tonal evolution.

MuSA.RT Opus 2 [4] is another tonal visualization of music at multiple scales that uses a spiral array model for tonality layout with dots indicating keys and traits; the closet traits will also light up as a triangle (Figure 15(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 structure may be missed. As a result, the techniques summarized in this section are good for music learners to study music theory, but general listeners may not be interested to realize how the keys are changed, or how a variety of 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 structure of music.

4 Discussion

It is seen that directly relevant work on visualizing the abstract musical structure is limited, which is likely due to the difficulty of identifying qualitative features automatically from the score, and the fact that visualization is usually treated as a tool to present complicated music details rather than an informative graph to facilitate the understanding of a composition for beginners. Most work are conducted by researchers with musical background, so the needs of the untrained ones may be overlooked. For performance expression
visualization, while their approaches are more on music cognition in a descriptive way, their results are mostly hand-crafted and have not been seriously validated with formal user studies.

Music appreciation and comprehension are rarely addressed in the previous work. Low-level structural elements like tones, chords and keys might be of interest for music experts; contrarily, high-level cognitive structure of music is useful for beginners or even amateurs to quickly recognize the overall structure of a composition. Generally speaking, the former aids music theory training while the latter visualizes abstract structure for learning and appreciating music. One may argue that musicians are familiar with the typical scores and the demand for such a visual tool is little. Yet, musical structure visualization also helps untrained people who do not necessarily acquire musical knowledge to learn music briefly without a score. One does not need any musical background to understand the visualization of musical structure, unlike the work on visualizing music theory content that unavoidably assume some background knowledge on the users.

Music scholars usually write lengthy essays on a certain composition, which may appear on a concert programme to introduce the piece to the audience, or are studied by music students when they encounter the work in music history lessons or performance classes. Huge amount of information on the musical structure can be found in the text, but it is nearly impossible to perfectly retrieve all these pieces of information from the scores using analysis algorithms. After all, the power of music is far beyond the dots and lines on a musical score; the score is not intended for reading, but for being performed. The power of music only comes when it is played. Despite the challenges in identifying these structural features in a computational way, it should not hinder us from developing an effective musical structure visualization tool.

Music information retrieval and analysis methods have been extensively studied in the field of computer music. Score analysis such as GTTM and IR [15] can be applied, but only a few qualitative attributes are tackled, and how to guarantee them to be perceptually correct is another concern. If we accept human-defined input like TimeSketch, we can constraint our work to develop an effective visualization for gaining insights into the structure of music. A promising direction would be to accept MIDI files as digital score input on which well-established music analysis algorithms can be employed, providing the users with the initial results. Then they can modify the underlying data via intuitive user-interaction to refine the visualization and to possibly add more human-defined features. It is feasible as music learners often highlight the musical structure on the scores to assist performance, which can act as the metaphor for the user-interface.

Lastly, the lessons learned from the previous work are summarized below. Most systems are implemented in Java as there are various related libraries available, and Finale [18] is used for score editing. A XML standard MusicXML [31] is commonly seen as the input data structure. Various papers also point out that animation should be included to amplify user’s perception on music and to provide clues for them to follow the display when the music is being played. Most importantly, a formal user study is demanded to judge its usefulness. However, user studies in all these papers are not well designed or are simply absent.

5 Conclusions

In this report, the background, motivations and challenges of musical structure visualization were first discussed, followed by the terminology in distinguishing sound and music. Next, related work from a variety of disciplines including visualization, human-computer interaction and computer music were compiled, with an emphasis on the visual encoding scheme and the effectiveness for knowledge discovery. Finally, some possible directions for the musical structure visualization project were presented in the discussion part.

Musical structure visualization serves as a visual reinforcement of the abstract structure of music for the audience, such that they feel less uneasy when they listen to the unfamiliar and sophisticated classical music. This graphical representation provides a good alternative to the typical music literature essays for understanding a piece of musical work perceptually with minimal music expertise. Such visualization can aid music listening in an interactive environment, be appended to posters and concert programme for visu-
ally penetrate the structure of music for music appreciation, or even demonstrate the differences between compositions from different periods and composers for learning purpose.

6 Acknowledgements

I wish to thank Professor Huamin Qu, and my group-mates Hong Zhou, Weiwei Cui, Ming-Yuen Chan, Ka-Kei Chung, Kai-Lun Chung and Wai-Ho Mak for their initial support and enthusiasms for the musical structure visualization project.

References


