WHAT YOU SEE IS WHAT YOU GET: ON VISUALIZING MUSIC

Eric Isaacson
Indiana University School of Music
Department of Music Theory
Bloomington, IN 47405 USA
isaacso@indiana.edu

ABSTRACT

Though music is fundamentally an aural phenomenon, we often communicate about music through visual means. The paper examines a number of visualization techniques developed for music, focusing especially on those developed for music analysis by specialists in the field, but also looking at some less successful approaches. It is hoped that, by presenting them in this way, those in the MIR community will develop a greater awareness of the kinds of musical problems music scholars are concerned with, and might lend a hand toward addressing them

Keywords: visualization, analysis, harmony

1 INTRODUCTION

Though music is fundamentally an aural phenomenon, we very often communicate about music through visual means. A musical picture converts the unidirectional time of a piece of music into a spatially represented dimension. This allows us to view a musical work as if it were a physical object–we can examine it in any order, at any pace, comparing temporally detached events with a simple flit of the eye.

Used in conjunction with a music-theoretically sound concept of musical structure, pictures can be effective tools for both discovering and conveying musical information. Different methods allow us to view snapshots of a musical work taken from different vantage points. No tool can provide a complete picture of a work, however, so researchers must keep in mind what their goals are, and what tools are most appropriate to achieve them.

This has implications for Music Information Retrieval (MIR), because musical visualizations are often direct reflections of an underlying musical representation, and the choice of a representation impacts directly on what musical features can be searched. It is important, therefore, that researchers consider carefully the underlying representation and its outwardly expressed visualization when implementing music IR tasks. Visual representations have great explanatory power. It is also important, however, that they be properly grounded in both musical knowledge and an understanding of human cognition. When poorly designed, or based on an incorrect understanding of the underlying musical structure, however, they can mislead.

The purpose of this paper is to consider some ways in which musicians and scholars have proposed we picture music and what role they might play in MIR tasks. The paper will focus on visualization approaches that involve music analysis, whose aim is to explain or illuminate characteristics of a musical work. The techniques discussed touch on several musical features, including pitch and rhythm, form, texture, and structural hierarchies. The focus is on Western music–and primarily on art music, though the principles generalize readily to popular musics. Some visualization techniques will be found to be less effective than others. Some of the techniques are already technology-based, others are carried out strictly manually. It is hoped that, by presenting them in this way, those in the MIR community will develop a greater awareness of the kinds of musical problems music scholars are concerned with, and might lend a hand toward addressing them.

2 COMMON MUSIC NOTATION

Although not itself analytical, because so much music analysis is derived from music notation in some way, it is worth considering first what Western notation does and does not represent. So-called Common Music Notation (CMN) traces its origins to the Middle Ages, with a number of important refinements taking place in the Renaissance, and additional incremental modifications occurring since then. It developed to more easily preserve the extensive plainchant repertoire of the Roman Catholic church, which was previously carried on through oral tradition. The primary purpose of CMN was thus, and in fact still remains, to facilitate the performance of a musical composition, by serving as a guide to performers. In much the same way that the written word allowed languages to be codified and for literature to emerge, the development of music notation facilitated the emergence of the concept of a musical work–an artefact that could be reliably passed

Figure 1: Score reduction of Beethoven, Symphony No. 5, Mvt. 1, measures 1-24.

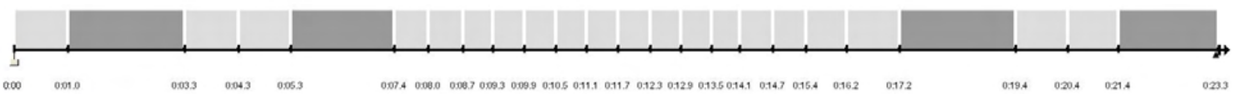

Figure 2: Timeplot showing relative lengths of notated measures in a recorded performance of Beethoven's Symphony No. 5, Mvt. 1, measures 1-24.

along to later generations.

Our notation system embodies a number of metaphors that both reflect and shape how we understand music. Primary among these is the notion that pitches are discrete objects that exist in spatial relationships: one note is "higher" and "longer" than another. These basic spatial metaphors lead to others: scale degrees are arranged in "steps"; melodies "ascend" and "descend"; we speak of "big sounds," "thick textures," of motives being "stretched" or "compressed"; we speak of not just "voice leading," but "smooth" voice leading; we speak of "soft" dynamics, "hard" attacks, "harsh" dissonances, melodic "shape," and musical "form." (See Hatten, 1995, and Zbikowski, 2002, for recent discussions of metaphor in music.)

The musical score horizontally represents the temporal order of events left-to-right. The vertical dimension expresses multiple variables, particularly in a complex orchestral score. Within a part, pitch height is determined by the height of a notehead on the staff (relative to the clef). Staves are grouped together in order of instrumental range, first within instrument families (e.g., piccolo, flutes, alto flute) and then between instrument families (flutes, oboes, clarinets, bassoons). Finally, the families themselves are grouped together (woodwinds, brass, percussion, strings). The ordering is designed to optimize reading by the conductor. (The musical score itself was a later development; many vocal works in the Renaissance existed only as individual part books–one for soprano, one for alto, and so on.)

Music notation reflects many centuries of accumulated user feedback and collective wisdom. Properly interpreting the intention of a score is more than a matter of reading the notation, however. In addition to understanding the symbols, there are a host of performance conventions that affect the interpretation of those symbols. Some of these include the addition of improvised ornamentation, determining when notated repeat signs are to be followed, knowing when the seventh note of the scale should be raised a half step in Renaissance polyphony, what tempo to play, and when and how much to deviate from the strictly metronomic tempos.

Although CMN is a remarkably adaptable system, it is largely optimized for performance. It is therefore insufficient by itself for music analysis. It doesn't show us harmonic analysis, motivic relations, musical form, etc. To illustrate this with a single example, consider the problem of depicting musical time. Though music is generally assumed to have a preferred tempo and note values seem to be defined in a strictly hierarchical manner in which elements at one level are grouped into twos or threes at the next level. In practice, however, most music is anything but regular.

Figure 1 shows a score reduction of the first 24 measures of Beethoven's fifth symphony. Each measure consists of two beats which, in the absence of other factors, would be expected to be roughly the same length. Figure 2 shows the timing in a recording of this passage with Pierre Monteux conducting the London Symphony. Each block represents one measure of notated music. (The timing points were set by manually clicking a button on each downbeat and then carefully checking the placement and adjusting as needed.) Of course, the fermatas in the measures shown in darker gray would be expected to be longer than the others, but note that the measures preceding these are also longer than those in the fourteen measures in the middle that are uninterrupted by fermatas.

An MIR system needs to decide whether to focus on the "musical" time as notated in the score or on the real time reflected in the performance. In the case real-time information is desired, then another type of visualization might be more appropriate. Similar decisions relating to the representation of pitch, timbre, and other musical features must be made.

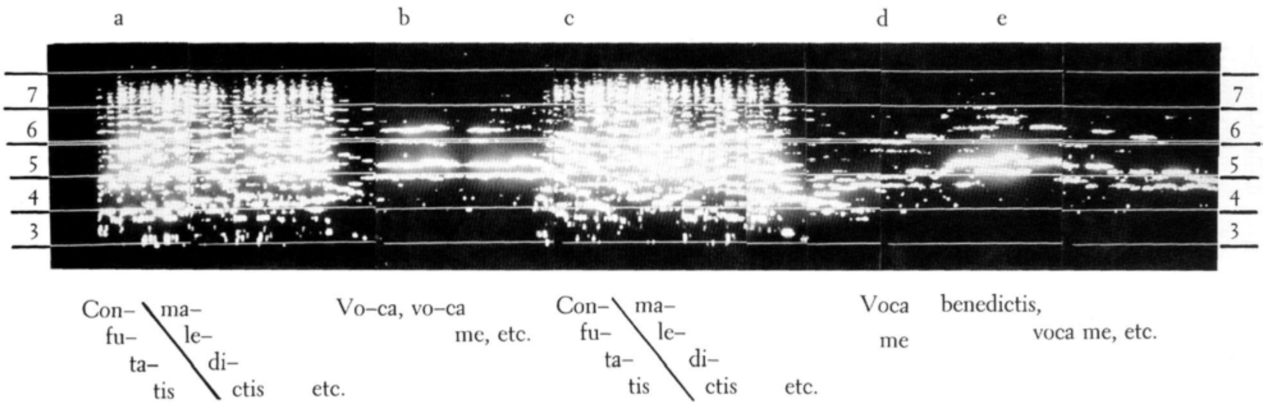

Figure 3: Spectrogram of a recorded excerpt of W.A. Mozart, *Requiem* K. 626, “Confutatis,” as performed by the chorus and orchestra of the Gulbenkian Foundation of Lisbon, Michael Corboz, Conductor. (Cogan, 1984)

3 SEEING SOUND

The value we place on the musical score notwithstanding, we experience music primarily through sound, usually in the form of an intricate combination of complex waveforms representing (potentially) dozens or even hundreds of different sound sources. An acoustic signal can be represented visually with a spectrogram, which graphs time (x) vs. pitch frequency (often on a logarithmic scale on the y axis). Cogan (1984) devotes a book to the analysis of a wide range of pieces based on spectrograms taken of performances of them. Figure 3, taken from that book, represents a section from Mozart’s Requiem. Numbered bands along the y-axis represent octave regions. The image depicts clearly the dramatic musical contrast between *Confutatis maledictis, flammis acribus addictis* (“When the accursed have been confounded And given over to the bitter flames,” sung by male voices and accompanied by brass and low strings) and *Voca me cum benedictis* (“Call me with the blessed,” sung by female voices). Cogan observes that the high partials present in the Confutatis sections are due in part to the frequently occurring high-pitched [i] vowel. The spectrogram is particularly useful for conveying the broad sonorous contrasts created by changes in orchestration, in musical texture, and dynamics–more so than traditional notation.

Although music in audio form is comparatively ubiquitous and, for computer-based applications, is easier to obtain than music in symbolic form, it has many limitations as a representation for analysis. Humans have the (remarkable) ability to recognize individual components in a sound source, including identifying specific instruments, or instrument families, as well as melodic lines and rhythmic patterns within each, and to translate that information into a mental symbolic form that is more reminiscent of the musical score than of a spectrogram. As those carrying out research in polyphonic transcription know, it is exceedingly difficult to extract this information from an audio signal, and hardly more visible in a picture of that signal. In fact, except when spectral (i.e., timbral) information is specifically the focus, the visual “noise” that the overtone structures add to the image masks much of the information that is traditionally of interest in music analysis.

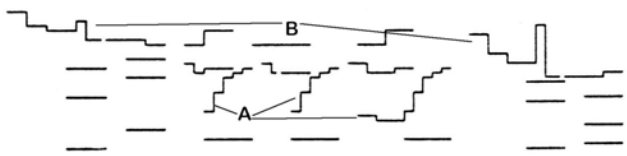

Figure 4: Brinkman and Mesiti (1991) graphic rendition of score of Schoenberg, *Drei Klavierstücke*, Op. 11, No. 1, measures 1-11. Annotations added.

4 SPECTROGRAM ANALOGS

It is possible to derive some of the same benefits of the musical spectrogram, but without the messiness of the timbral information using symbolic data, which can easily depict just fundamental pitches. Figure 4 is a “part plot” (Brinkman and Mesiti, 1991) of the first eleven measures of Schoenberg’s Piano Piece, Op. 11, No. 1. As in the spectrogram, time proceeds left to right, though now by notated time, not in real performance time, while notated pitch follows the y axis as in a spectrogram. Pitches that belong to the same musical voice and are not interrupted by rests are connected with vertical lines, so that melodic gestures can be readily seen. The format makes it easy to see recurrences of the same motive (A) or a variation of the opening melodic gesture (B), as well as where the moving melodic lines occur in relation to the sustained accompanying chords.

This type of visualization has a number of potential benefits relative to standard notation. In a chamber or orchestral score, for instance, it flattens the contents of the various staves into a single coordinate system. (A symphonic score can easily include 15 staves or more.) It also eliminates the visual clutter of staves, barlines, as well as note heads and associated stems, flags, and beams, allowing one to focus on basic melodic shape. The tradeoff is that specific pitch and rhythmic/metric information is missing, as is the timbral information present in a spectrogram.

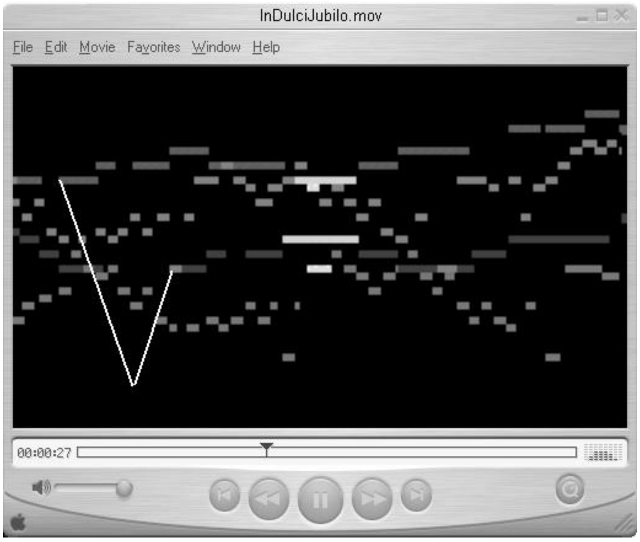

Figure 5: Video excerpt from Music Animation Machine (Malinowski, 2005) realization of J. S. Bach, “In dulci jubilo” from *Das Orgelbüchlein*. V-shaped line added.

A related type of notation is used in several animations produced by Malinowski (2005). In Figure 5, time and pitch are plotted on the same axes as in Brinkman and Mesitis graphic, with two differences: pitches are not connected with vertical lines to show larger gestures, and the input is a MIDI file, which means the graph is based on performance data rather than notated durations. (Of course Brinkman and Mesiti’s graphs could also use performance timing, and Malinowski’s MIDI files could be generated with strictly quantized data, so this distinction is not particularly meaningful.)

Malinowski uses color effectively to represent different parts of the four-part musical texture: the chorale melody being played in the first and third parts are in darker colors. I have added lines to show how the lower part imitates the upper in this excerpt. The faster-moving accompanying parts are depicted with a lighter color. (White is used for the currently sounded notes.)

This type of line graph can be generalized to show features other than pitch on the y axis. Another graph by Brinkman and Mesiti (Figure 6) shows, for instance, the dynamic levels notated in the score of the first 24 measures of Bartók’s fourth string quartet, first for each of the four parts, and then in composite. When only a solid line is shown, it indicates that the instrument is not playing at that time. Those familiar with the opening of Bartók’s quartet will recognize the characteristic dynamic contour of the movements opening gestures, as well as the loud-then-soft contour of the passage as a whole.

5 MUSICAL FORM AND TONALITY

It is common to represent musical form in a graphical format. The purpose of a form diagram is to show the recurrence of previous themes and the introduction of new ones. Relatively simple music can be diagrammed quite minimally. For instance, a large number of American popular songs (“Autumn Leaves,” “Over the Rainbow”) is in the form $aaba$ or a close variant.

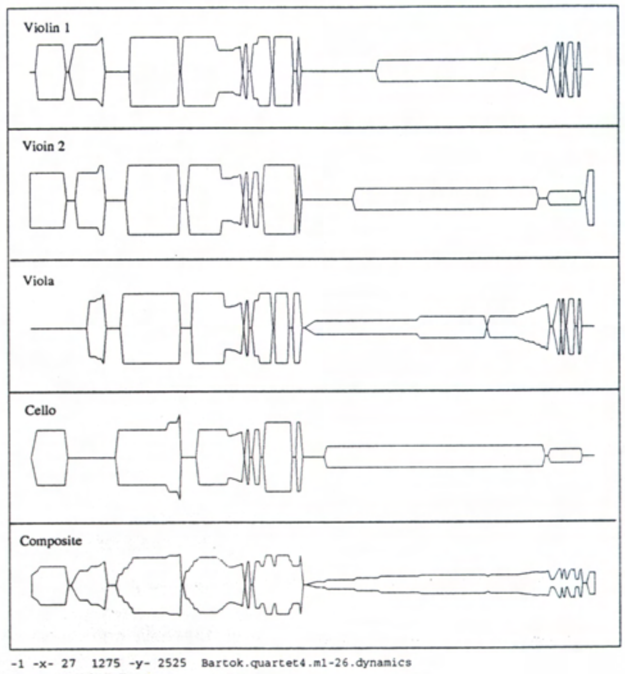

Figure 6: Brinkman and Mesiti (1991) plot of Bartók’s *String Quartet* no. 4, mvt. 1, measures 1-24, mapping time (x) vs. notated dynamic level (y).

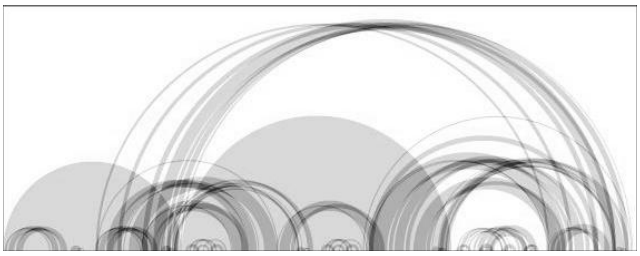

Figure 8: Wattenberg (2005) diagram showing repeated musical fragments in an unspecified Mazurka in F# minor by Chopin.

Ultimately, the complexity of a formal diagram is based on the complexity of the music and the desired level of granularity. Figure 7, for instance, shows a diagram of a typical minuet and trio movement from a piano sonata. Thematic similarities are depicted with both similar text labels and similar colors. The large ABA structure depicts the overall arrangement, minuet, trio, and shortened minuet repeat. Internal bubbles reveal the essential bipartite division of each of those sections, while the ||: a :||: b a :|| structure of the A and B sections are apparent at the lowest level. Derived from a recorded performance, the sections in this diagram are shown proportional to real, not musical time. (Numbers below the diagram are measure numbers.)

Figure 8 represents musical recurrence in a rather different way. Designed by digital artist Martin Wattenberg, the diagram uses arches to connect repetitions of musical material. Whereas Figure 7 shows thematic repeti-

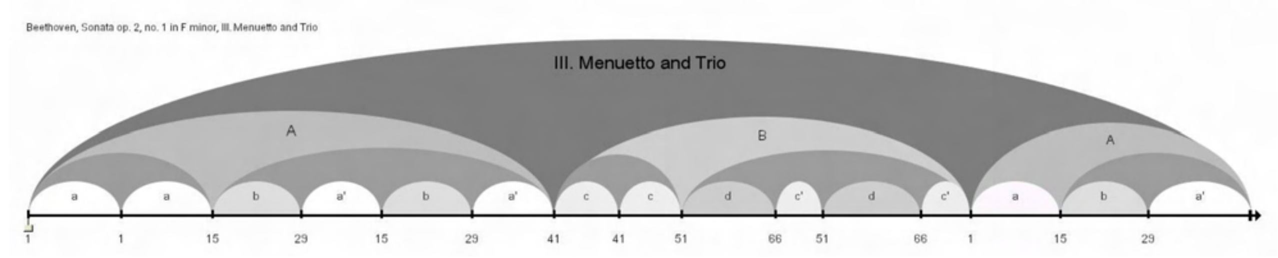

Figure 7: Formal diagram of Beethoven, Piano Sonata, Op 2, No. 1, Mvt. 3, produced using Variations2 timeline tool (`http://variations2.music.indiana.edu/`). Sectional proportions based on performance by Richard Goode.

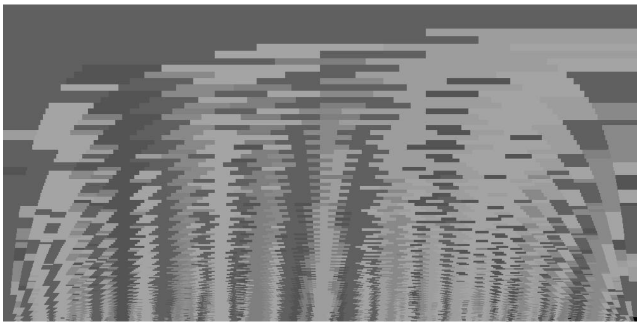

Figure 9: Craig Sapp's Tonal Landscape (type 1 plot) of J. S. Bach, *Well-Tempered Clavier*, Book 1, Prelude in C-sharp minor. From `http://ccrma.stanford.edu/~craig/keyscape/`.

tions only at the sectional level, this one connects repeated events wherever they first occur. The thicker the band, the more extensive the material that is repeated. For example, two large immediately repeated sections are apparent in the first quarter and the central half of the example, whereas a number of short elements from the end of that first repetition recur at the end of the piece (see the series of tall, thin arches spanning most of the length of the figure).

Though the diagram is visually appealing (even more so in the translucent pastel blues in the color version), it fails as an effective depiction of musical design in several respects. First, if the first section is repeated (as suggested by the first solid grey arch), then why do the materials at the end of the piece refer to the repetition and not to their first instance? This leads to a strangely non-hierarchical view of the piece that is surely at odds with its structure (compare the orderliness of the previous Beethoven example). Also, because the height of an arch is related only to the distance of the events it connects, it gives a sense of importance to repetitions that are far apart in the music that may or may not be justified musically. An MIR system that contains (or automatically generates) form diagrams of musical pieces should support the formal model, not the latter.

A different type of musical structure is depicted in a "tonal landscape" by Sapp (2001). Figure 9 depicts the tonal structure of the C-sharp minor prelude from Bach's *Well-Tempered Clavier*, Book 1. Again, time proceeds along the x axis, but in this case, as one moves downward

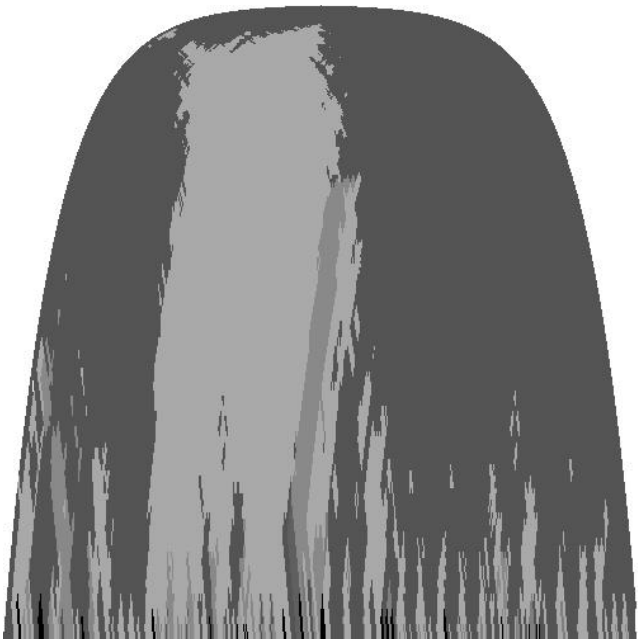

Figure 10: Craig Sapp's Tonal Landscape (type 2 plot) of Mozart's *Viennese Sonatina* No. 1 in C, Mvt. 1. From `http://ccrma.stanford.edu/~craig/keyscape`.

from the top of the figure, each row divides the piece into $n+1$ segments of equal length (the first row 1, the second 2, and so on). Each segment is then assigned a color based on an estimate of the overall key that is characteristic for that segment. By the bottom of the graph, key estimate are being made for very short segments of music. The estimated keys are displayed using a color scheme that maps each note around the circle of fifths to adjacent colors in the rainbow (E = red . . . C = green . . . A = violet).

A second type of graph (Figure 10) also depicts increasingly local key estimates as one moves from top to bottom. Rather than using $n + l$ discrete segments for layer l, the figure uses a continually sliding window that grows smaller as it moves toward the musical surface at the bottom. This figure has been further modified by the application logarithmic scaling that squashes the top of the image, allowing the features near the bottom to extend more visibly toward the upper part of the image.

Sapp's plots are intriguing in that they depict the occasional ambiguity of tonal orientation experienced as one listens to a piece. Several details of his approach are problematic, however. One is that it is highly doubtful that we perceive tonality on as many levels as suggested by the diagrams. More critically, the perception of tonality is bound up closely with the perception of form. In particular, phrases usually begin and end stably in a key. Only certain kinds of musical events trigger ambiguity of key center, and this ambiguity exists with nowhere near the frequency implied by the diagrams. Third, though the images supposedly convey a sense of key distance, the use of colors to characterize this distance is of little help visually, for we do not generally conceive of colors as being a certain distance from each other.

6 OTHER MUSIC STRUCTURE VIEWS

A final pair of analytical approaches relies on music notation, albeit in untraditional ways. These approaches require a considerably more sophisticated conception of musical syntax to understand properly.

The theories of Heinrich Schenker (1969) are among the most commonly used in the analysis of tonal music. Schenker posits a hierarchical view of music that resembles the construction of buildings: upon a basic foundation common to all music is built an inner framework, to which is added wall and floor members, then paint and carpeting, and finally the furnishings. (This oversimplification will suffice for present purposes.) Figure 11 depicts one layer of Schenker's analysis of a chorale by Bach. As is characteristic of Schenker's analyses, emphasis is given to the counterpoint between the outermost voices of the music. In the layer shown in the figure, many of the decorative pitches from the musical surface (passing and neighboring tones, for instance), along with the inner parts (alto and tenor) have been removed. Pitches that belong to deeper structural levels are shown with open noteheads and beamed together. Stemless noteheads are least significant structurally and are generally considered decorative at this level. Schenker's complete analysis includes two further stages of reduction, plus one version that is more elaborated than the one shown here. The last stage of reduction shows only the notes in open noteheads here– the foundation of the piece. The more elaborated version simply resembles the actual score, but with the structural pitches from this layer shown.

The final analytical approach to be discussed is that proposed by Lerdahl and Jackendoff (1981). Based on theories of generative linguistics, Lerdahl and Jakendoff use various means to describe metrical organization in a piece, as well as grouping and prolongation structures. Each of these is depicted in Figure 12. Dots directly below the music are used to show relative metrical strength of each beat. Brackets below that depict a hierarchy of groupings of musical events. The tree diagram above the score depicts either prolongations from one structure to the same or a related structure, or progressions from one structure to a different one, and these are further characterized as tensing or relaxing. The events linked to the highest-level branches are considered the most important structurally in the passage. Each symbol is determined by well-formedness and preference rules, derived in spirit from generative linguistics.

The theories both of Schenker and of Lerdahl and Jackendoff are highly sophisticated and, because they rely on artistic interpretations on the part of the analyst, their application in explaining musical works seem yet to be beyond the capabilities of automated retrieval systems.

7 CONCLUSIONS

Many of the visualization techniques described here can reveal interesting, musically relevant, and at times highly sophisticated information about a musical work, information that would be hard to depict in another way. The questions that led to these visualization techniques have been posed by music analysts who think deeply about musical structure and musical meaning. They represent only a small sampling of the rich literature that awaits discovery by those in the MIR community who might wish to address similar questions.

REFERENCES

A. Brinkman and M. Mesiti. Graphic modeling of musical structure. *Computers in Music Research*, 3:1–42, 1991.

R. Cogan. *New Images of Musical Sound*. Harvard University Press, Cambridge, Mass., 1984.

R. S. Hatten. Metaphor in music. In E. Tarasti, editor, *Musical Signification: Essays in the Semiotic Theory and Analysis of Music*, number 121 in Approaches to Semiotics, pages 373–391. Mouton de Gruyter, Berlin, 1995.

F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, Cambridge, Mass., 1981.

S. Malinowski. Music animation machine, 2005. URL `http://www.well.com/user/smalin/mam.html`.

C. Sapp. Harmonic visualizations of tonal music. In *Proceedings*, pages 423–430, Havana, Cuba, 2001. International Computer Music Conference.

H. Schenker. *Five Graphical Music Analyses*. Dover Publications, New York, 1969.

M. Wattenberg. The shape of song, 2005. URL `http://turbulence.org/Works/song/`.

L. Zbikowski. *Conceptualizing Music*. Oxford University Press, New York, 2002.

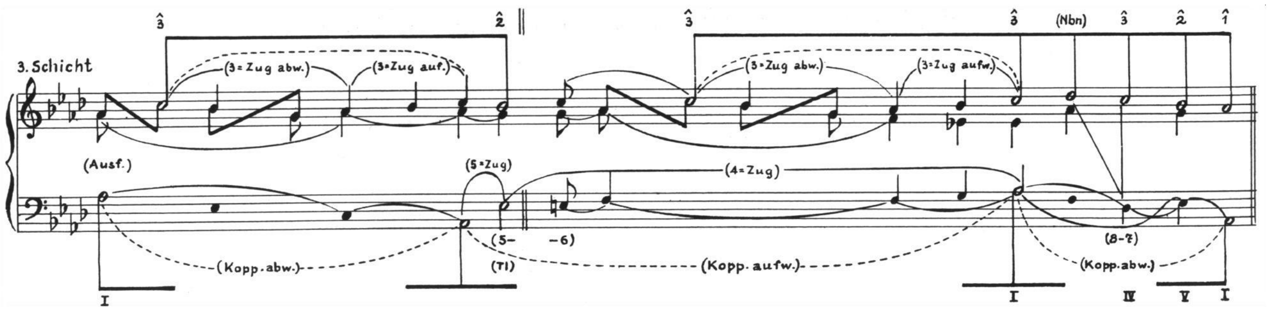

Figure 11: Third (foreground) layer from Heinrich Schenker's analysis of J. S. Bach's setting of "Ich bin's, ich sollte büssen." (Schenker, 1969).

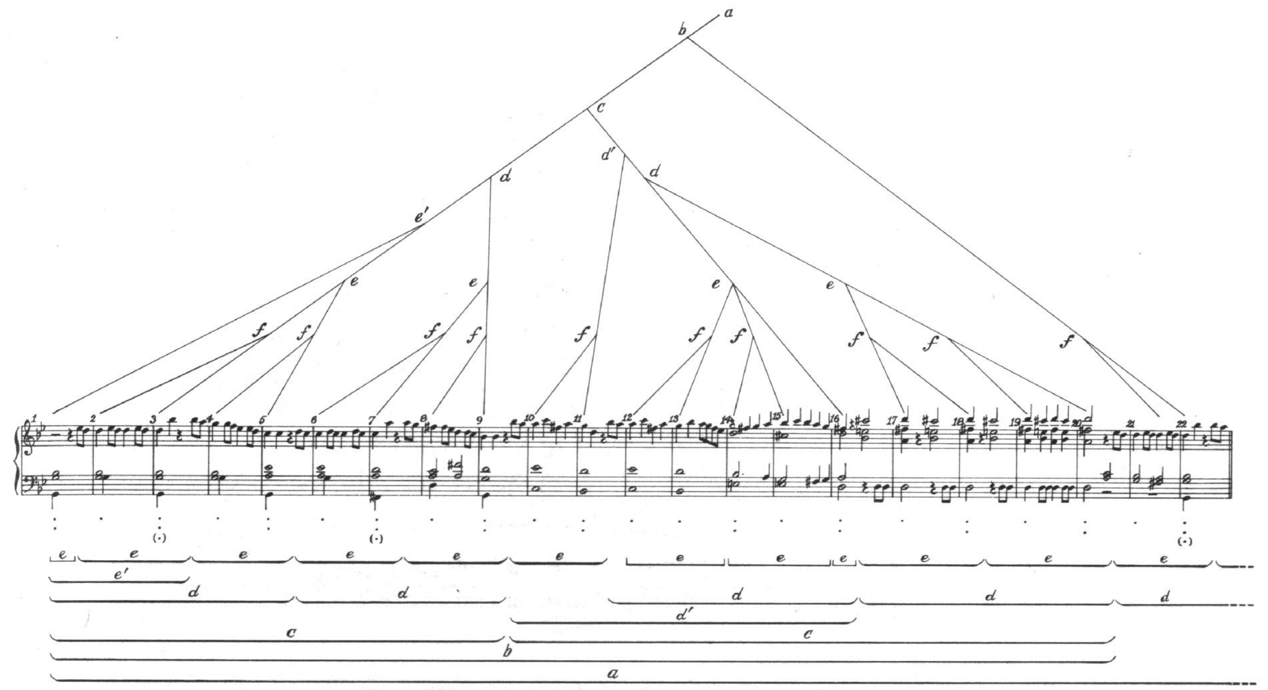

Figure 12: Excerpt from Lerdahl and Jackendoff's diagram of Mozart, Symphony No. 40 in G minor, measures 1-22. (Lerdahl and Jackendoff (1981), p. 259).