Scoregram: Displaying Gross Timbre Information from a Score

Rodrigo Segnini and Craig Sapp

Center for Computer Research in Music and Acoustics (CCRMA),
Center for Computer Assisted Research in the Humanities (CCARH)
Stanford University
660 Lomita Drive, Stanford, CA 94305, USA
{rsegnini,craig}@ccrma.stanford.edu

Abstract. This paper introduces a visualization technique for music similar to that of spectrograms which display time and frequency content, but with the addition of a multi-timescale aggregation that offers at-a-glance visual structures which are interpretable as the global timbre resulting from a normative performance of a score.

1 Introduction

A musical score using common music notation (CMN) does not convey a literal representation of the sound it notates; rather, it contains the instructions necessary to produce the sound. This realization from score to sound is a convoluted process which may be simplified as follows: (i) pitch and duration are read from the vertically and horizontally positions of symbols on a staff; (ii) associated markings—not always aligned to the symbols they modify—inform us about loudness or articulation-dependent onsets; and finally, (iii) other standard symbols and editorial practices—such as placing the name of an instrument next to staves—complete what is needed to produce or imagine the notated sound. This is repeated for all instrumental parts in order to obtain a broad mental picture of the sound from the score.

With sufficient knowledge of CMN, one is thus able to aggregate these raw graphical symbols on the event level into higher level structures that can be described in terms of phrases, melodic contour, harmonic complexity, tonality, event density, intensity, *etc*. Arriving at this representation is particularly useful in obtaining an idea of the overall form of a piece and its characterization.

However, despite the standardization of CMN, various constraints may affect the layout of this data, affecting the speed at which we can parse it. Space limitations are an example of such constraints, which may force changes in clef, octave markings or in the spacing between symbols, all of which hinder the spatial relationship between a notated event and its audible correlate.

We denominate this kind of mental picture of a score the ‘gross timbre information’ because it represents the compounded result of the actions by the performer(s) producing the notated sound. This paper introduces an approach for displaying this information directly from the score using computational methods.

1.1 Timbre Information Display

One way to simplify the display of gross timbre information is to use a spectrogram. A spectrogram displays on the vertical axis frequency content in bands of width relative to the sampling resolution—with the amount of energy in a band depicted by grayscale or color values against time on the horizontal axis. The spectrogram's axes are more regularized than a musical score; however, larger musical structures other than the instantaneous surface features are difficult to identify when viewinwhen viewing a spectrogram.

Also, spectrograms display timbre in a non-intuitive way by giving too much literal information about frequency content rather than more perceptual measures of timbre. The physical parameters of timbre are usually reduced to a more compact set of features which still describe the acoustical signal, some of them with perceptual relevance. A partial list of such features which can be obtained from the time and/or spectral domain would include: root-mean-square amplitude (power), bandwidth (spread of the spectral energy), centroid (amplitude-weighted average of energy distribution), harmonicity (how much does that energy falls along harmonic partials), density (how much energy per critical band), skew (tilt toward low or high end of the spectrum), roll-off (decay of high frequency partials), flux (between frames), among others.

Grey [2] worked with listeners in similarity experiments so as to determine the perceptual correlate with some of these features, and he produced a timbral space displaying the perceptual distance among notes produced by different instruments. Most recent work, as exemplified by Fujinaga [3], Brown [1], and others, uses a host of those features to categorize timbre in an attempt to have computers recognize specific instruments.

1.2 Acoustic v. Symbolic

All of the approaches for timbral description, however, are derived from the acoustic representation of a musical sound, therefore their results are somewhat different from what can be specified by its symbolic representation, namely, the musical score.

Assuming that a score is the closest there is to the original compositional idea, then we have to count every step from there to our ears as potentially transforming factors. There are two major such steps in this path: performers and performance space; performers add vibrato, tremolo, rubato, plus their 'mistakes', and the performance space adds reverberation, and background noise. While many of these factors can be desirable, we sometimes end up with very different acoustic renditions of the same piece. As with listening, whatever structural information that can be derived from this approach becomes biased by the specific performance.

On the other hand, information derived from the symbolic representation is performance agnostic and is a time-honored way of generating gross conceptualizations of timbral content. However, this human-based approach is expertise-dependent and is time-consuming. This presents issues of consistency and speed given variabilities in CMN layouts, but it is very good to obtain information using different time-scales. In other words, humans are able to change their analysis window-lengths ranging from a single time event to the whole duration of the piece. The visualization techniques presented below attempt to keep the advantages of the human-based approach, while dealing with the shortcomings through a computer-based approach.

1.3 Previous Work

Recent visualizations of timbre include *Timbregram* and *Timbrespace* [11]. Timbregram is based on a time domain arrangement of the music (can be superimposed to a waveform display), with colors according to spectral features added to variable-size slices. Timbrespace maps features to objects with different shapes, texture and color in a 2D or 3D virtual space. Their goal is to facilitate browsing of a large number of sound files; the latter also suggests groupings among different pieces. For an experimental study on cognitive associations between auditory and color dimensions see [4].

The most direct predecessor of scoregrams are Craig Sapp's *Keyscapes*, which show tonality structure of a piece. In Keyscapes, the horizontal axis represents time in the score, while the vertical axis represents the duration of an analysis window used to select music for a key-finding algorithm; each analysis window result is shaded according to the output key. Independent analysis group together according to the relative strength of key regions the composition. A more detailed description of the visualization approach is given in [9] and [10].

Scoregrams are also closely related to *Dynagrams* used by Jörg Langer, *et al.*, to study loudness changes on multiple-time resolutions graphs [7]. Both plot axes are similar to keyscapes, but the vertical axis is inverted and the windowing method is slightly different. Dynagrams are used to plot the change in loudness of a recording over time. Crescendos are shown in shades of red, and decrescendos are shown in shades of green. Local dynamic changes display rapid changes in loudness and global dynamic changes can be seen emerging from this low level description of the loudness. Dynamic arches are displayed visually from the interaction of the local and global dynamic descriptions in the plot.

2 Implementation

To introduce the potential of scoregram we will display a single feature from the score—pitch height—according different subdivisions. In these examples, images were automatically generated from CMN data encoded in the Humdrum file format and analyzed using command line programs from the Humdrum Toolkit [6] as well as custom-built programs. Other symbolic representations would be just as good, such as MIDI files.

Meaningful visualizations are accomplished by mapping perceptually relevant features into an equivalent dimension in an alternate domain. Visual elements, for example, have a number of perceptually significant characteristics, such as shape, texture, and color, which can be linked in the auditory domain; some of them, like timbre, are also multidimensional. In this work we mostly explore color which has three perceptual dimensions of hue, saturation, and intensity, and focus on the first of them: hue.

Mapping According to Register A common association to the concept of timbre in a single instrument is register. The pitch range of most orchestral instruments can be summarily subdivided into three timbral zones each covering about a third of their range (i.e. low, medium, and high). We can determine these thresholds manually (i.e. setting a fixed note value at the boundary), or automatically (i.e.: at the 1/3 and 2/3 percentiles in the events histogram). For the following scoregrams, activity in each gross timbral range is indicated by the colors red, green, and blue, respectively, and it is proportional to the number of tokens from that class in the histogram, normalized by the largest token value of either: (i) all colors across the time-window, (ii) all values of a single color, or (iii) among the three values in that window. Finally, the normalized value becomes a number in the Red-Green-Blue color space. Therefore, a piece with activity only in the mid register would yield a green picture, while simultaneous activity in the extreme registers, would yield magenta resulting from the combination of red (low register) and blue (high register).

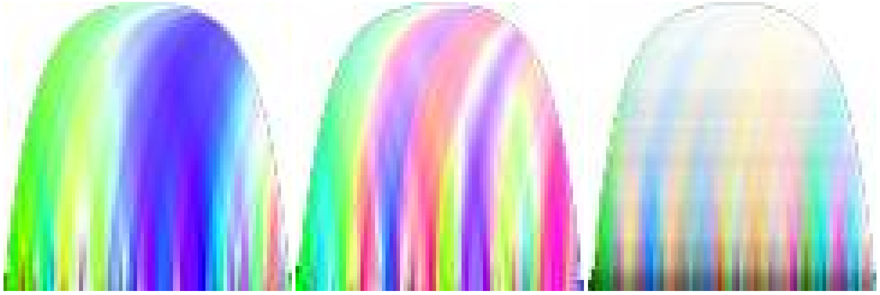

Fig. 1. Three scoregrams using range data. They illustrate a progression from strongly segmented and contrasting range-derived structures to a more more homogeneous structure. These examples are taken from J.S. Bach's fugues (Nos. 14, 1, and 20 from left to right, respectively) in the Well-Tempered Clavier, Book I. No.14 (*left*) has three clear sections where the medium and high registers appear most prominently; No.1 (*middle*) shows more boundaries with no color in particular becoming emphasized; No.20 (*right*) shows all colors with almost equal presence, resulting in an early aggregation toward white at the top of the scoregram

The images in Figure 1 show at-a-glance aspects about pitch distribution—by extension, register-dependent timbre quality— that are not obvious to the naked eye in a musical score. At the bottom is the event level, quantized to include every 16th-note duration on the score; this is done to keep equal score time for each token. Time goes from left to right, from the beginning to the end. The size of the analysis window increases from bottom to top, so that local features are shown below and global features at the top, which represents the entire duration of the piece. The progression from bottom to top is done in a logarithmic scale to match the way our perception of time works. Each row is the same fraction larger/smaller than the previous row. It can be suggested that the color at the tip of the dome is the characteristic gross timbre of the complete composition.

Another useful piece of information displayed in the scoregram are the color boundaries where register changes occur. For example, the rightmost plot in Figure 1 suggests that the resulting timbre is more uniform since no color becomes emphasized, whereas in the first plot, the movement from mid to high register becomes a distinctive characteristic of the piece.

Other Mappings Any arbitrary subdivision of the instrumental range is possible. For example, in a microtonal context, fine subdivisions may be necessary to augment the contrast of auditory variations. We have implemented subdivision into octaves—suggested to be a general bandwidth for timbre invariance [5]—and into critical bands for the note pitches (see Figure 2), a more perceptually uniform measure of frequency with a width of about 1/3 octave each; it is generally assumed that timbre can be characterized by the energy contents in each critical band [8]. Since these subdivisions produce more than the three regions which could be conveniently mapped one of the three RGB colors, we used a 2-D interpretation of the color space commonly known as the color wheel, and assigned an angle equivalent to a distinct color wavelength to each one of the 10 (v.g. octaves) or 24 (v.g. critical bands) tokens.

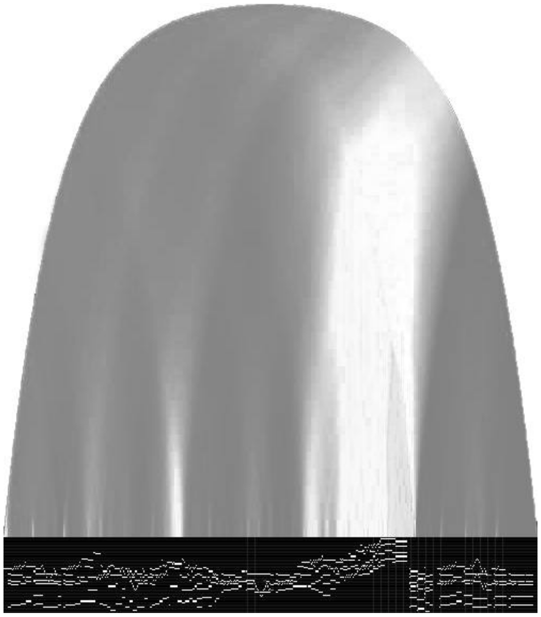

Fig. 2. A scoregram using critical band data from Barber's Adagio for strings. A piano-roll representation is appended to the bottom of the picture to depict the position of musical events. There is a clear boundary at the point were the music reaches a climax in the high register, before returning to the broad low and medium registers

Figure 2 also demonstrate how more striking structural features will rise higher in the scoregram plot. For example, in this plot the extremely high registration of all instruments about 75% of the way through the piece generate a strong band of contrasting color to the other regions of the piece.

3 Discussion

A scoregram can have various interpretations. For example, a piece whose event distribution is homogeneous across the dimension in which it is measured (e.g. register) may be perceived to be less dramatic than one with marked changes. The idea is that if at the top of the scoregram we can see boundaries preserved from the bottom, or the event-level, it means that the piece has contrasting sections.

Scoregram is extensible to any other types of musical features. We are considering the mapping of multiple features to unused color dimensions. The basic strategy we used is to plot three states in independent RGB values. Interpolating these values in the Hue-Saturation-Intensity (HSI) space can be used to map dynamics, for example, to saturation (*e.g.* how vibrant the color is), and articulation to intensity (*e.g.* how bright the color is).

In the sample of music examined thus far, scoregrams proved useful for detecting basic musical structures based on the musical features being examined. It may also useful for establishing measures of similarity between repertoires and forms, or comparisons between the precisely observable acoustic event and its notated counterpart, which would help to quantify a performer's contribution to the music.

References

1. Brown, J. C.: "Computer identification of musical instruments using pattern recognition with cepstral coefficients as features". Journal of Acoustic Society of America 105 (1999) 1933–1941
2. Grey, J. and G. Gordon: "Perceptual effects of spectral modifications on musical timbres". Journal of the Acoustical Society of America Vol. 63(5) (1978) 1493–1500
3. Fujinaga, I.: "Machine recognition of timbre using steady-state tone of acoustic musical instruments". Proceedings of the international Computer Music Conference (1998) 207-210
4. Giannakis, K. and M. Smith: "Imaging Soundscapes: Identifying Cognitive Associations between Auditory and Visual Dimensions" in Godoy, R. I., Jorgensen, H. (eds.): Musical Imagery. Swets & Zeitlinger (2001) 161–179
5. Handel, S. and M.L. Erickson: "A Rule of Thumb: The Bandwidth for Timbre Invariance Is One Octave". Music Perception 19 (2001) 121–126
6. Huron, D.: "Music Information Processing Using the Humdrum Toolkit: Concepts, Examples, and Lessons". Computer Music Journal 26 (2002) 11–26
7. Langer, J., R. Kopiez, C. Stoffel and M. Wilz. "Real Time Analysis of Dynamic Shaping" in the Proceedings of the 6th International Conference on Music Perception and Cognition, Keele, United Kingdom, August 2000.
8. Moore, B.: An Introduction to the Psychology of Hearing. Academic Press (2003)
9. Sapp, C.: Harmonic Visualizations of Tonal Music. Proceedings of the International Computer Music Conference (2001) 423–430
10. Sapp, C.: "Visual Hierarchical Key Analysis". Association for Computing Machinery: Computers in Entertainment, 3(4) (Fall 2005).
11. Tzanetakis, G.: Manipulation, Analysis, and Retrieval Systems for Audio Signals. Ph.D. Dissertation. Princeton University (2002)