COMPARATIVE ANALYSIS OF MULTIPLE MUSICAL PERFORMANCES

Craig Stuart Sapp
Royal Holloway, University of London
Centre for the History and Analysis of Recorded Music (CHARM)

ABSTRACT

A technique for comparing numerous performances of an identical selection of music is described. The basic methodology is to split a one-dimensional sequence into all possible sequential sub-sequences, perform some operation on these sequences, and then display a summary of the results as a two-dimensional plot; the horizontal axis being time and the vertical axis being sub-sequence length (longer lengths on top by convention). Most types of timewise data extracted from performances can be compared with this technique, although the current focus is on beat-level information for tempo and dynamics as well as commixtures of the two. The primary operation used on each sub-sequence is correlation between a reference performance and analogous segments of other performances, then selecting the best correlated performances for the summary display. The result is a useful navigational aid for coping with large numbers of performances of the same piece of music and for searching for possible influence between performances.

1 INTRODUCTION

In the Mazurka Project [1] conducted at CHARM during the past two years along with Nicholas Cook and Andrew Earis, we have collected over 2,500 recorded performances for 49 of Chopin's mazurkas—on average over 50 performances for each mazurka. Keeping track of differences and similarities between numerous performances is difficult when comparing recordings heard weeks, months or even years apart. And remembering the distinguishing features of 50 individual performances of a composition would be taxing on anyone's memory. Often the surface acoustics of a performance (such as reverb, microphone placement, piano model, recording/playback noise) are more noticeable and memorable than the actual performance, so identifying related performances solely by ear can sometimes be difficult.

A written score contains only the most basic of expressive instructions. The composer relies on the performer to interpret the work according to implicit rules as well as the written instructions. The unwritten rules of a composition are transmitted aurally between performers as well as passed down from teacher to student. These performance conventions can apply to specific pieces, genres, composers or entire time periods. Performances may involve combining interpretations from several sources, such as teachers or other admired pianists; or conversely, it could be a reaction against convention.

To help in the exploration of influences between performances, basic descriptions of tempo and dynamics are extracted from each performance of a work which can then be correlated against each other. A single global similarity measurement for this data could miss interesting smaller-scale structures. Therefore, the following plots were developed which display the closest performance to the reference at all possible timescales.

In the most interesting variation of the plot, each performance is assigned a color, and when a particular performance is most similar to the reference, its color is filled in the corresponding point in the plot. As a result of looking at all time spans, patterns of color emerge which can give clues to the relative importance of other performances to the reference performance of the plot.

2 RAW DATA

Two types of data are used for comparative analysis: beat duration and loudness. There are many other facets of performance which are being ignored, such as individual note timings, voicing, pedaling, and articulation. However, tempo and overall loudness level at the beats are easier to extract from audio data than many other expressive features and form a reasonable expressive baseline.

Both tempo and loudness data are extracted beat by beat throughout a performance, and the data can be plotted against the sequence of beats as illustrated in Figure 1. While the data is extracted by beat from the performances for this paper, we are also working on extracting individual note times and dynamics (including off-beats as well as hand synchrony). Such fine-grained performance information may prove useful in characterizing similarities or differences between performances.

Beat durations are extracted by first recording taps in real-time while listening to a performance in an audio editor called Sonic Visualiser developed at the Centre for

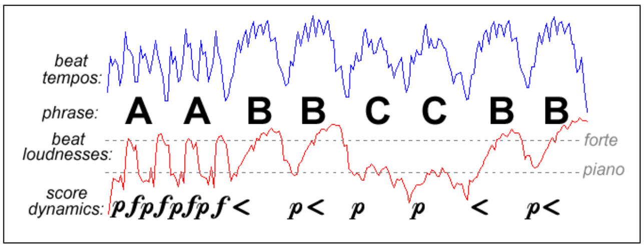

Figure 1. Average tempo and dynamic graphs for 35 performances of mazurka in B minor, 30/2.

[1] `http://mazurka.org.uk`

Digital Music at Queen Mary, University of London. [2] The resulting taps are not aligned precisely to true beat onsets in mazurkas due to a lag in response by the listener—typically with a standard deviation of 60–80 ms (compared to about 30 ms for following a steady tempo). Therefore, audio analysis plugins are used to assist in adjusting the taps onto the exact attack times of notes played on the beats. [3] By repeating data entry for the same performance in an independent manner, the alignment error is reduced to a standard deviation of around 11 ms. Defining a data error as a difference in beat localization by more than 50 ms, the measured data-entry error rate was about 1% for recordings made after 1980 and 3% for recordings in good condition from the early 1920's.

At timing resolutions around 10 ms, defining beat location can become difficult in piano music, particularly due to attack-time differences between the left and right hands (hand synchrony). In these cases, the best procedure is to define the beat location in a consistent manner in the analogous places in each performance. Since the melody usually contains more expressive timing, it is useful to define the beat as the time at which the melody note is played rather than using the less-expressive accompaniment.

For comparisons of musical dynamics between performances, a smoothed version of the raw power calculated for the audio signal every 10 ms is sampled at each beat location. The raw power in decibels in a sample of audio is given by the equation:

$$\text{raw power} = 10\ \log_{10}\left(\frac{1}{N}\sum_n x_n^2\right) \tag{1}$$

where N is the number of audio-samples in sequence x being considered. The raw power measurements are then smoothed with an exponential smoothing filter described by the following difference equation:

$$y[n] = \alpha\, x[n] + (1-\alpha)\, y[n-1] \tag{2}$$

where α is a constant set to 0.2 in the case of 44100 Hz audio data with power measurements made every 10 ms. The exponential smoothing filter is applied twice to the raw power data: once in the forward direction and once in the time-reversed direction. This keeps the smoothed data centered at its original time location. To extract a loudness level for a particular beat in the audio, the smoothed power value about 70 ms after that onset is used—to compensate for a loss of high-frequency information in the smoothed data which delays the maximum amplitude location of note attacks.

3 ANALYSIS TOOLS

3.1 Correlation

Normalized correlation, or *Pearson* correlation, is defined in Equation 3. This form of correlation yields values in the range from -1.0 to $+1.0$, with 1.0 being an exact match, and 0.0 indicating no predictable relation between the sequences being compared.

$$r(x,y) = \frac{\sum_n (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_n (x_n - \bar{x})^2 \ \sum_n (y_n - \bar{y})^2}} \tag{3}$$

[2] http://www.sonicvisualiser.org

[3] http://sv.mazurka.org.uk

where x and y are number sequences of the same length; $\bar{x}$ and $\bar{y}$ are average values of each number sequences x and y.

Correlation is a useful way to measure the similarity between two shapes such as comparing different performers tempo and dynamic curves as shown in Figure 1.

3.2 Scape plot

Correlation values are difficult to interpret in isolation, so the following plotting method is one way of presenting the data in a more human-readable format. Scape plots take their name from the word *landscape* since they show small-scale features analogous to the foreground in a picture, as well as large-scale features similar to the background. And like a painting, the interesting parts of the scape plot usually lie somewhere in the middle-ground.

Consider a simple example illustrated in Figure 2. A musical performance consists of six beats which are labeled: A, B, C, D, E, and F. These six beats can be chopped up into 21 unique sub-sequences (n-grams). Firstly, the elements can be considered in isolation. Next they can be grouped by sequential pairs: AB, BC, CD, DE, EF. Then by threes: ABC, BCD, CDE, DEF; by fours: ABCD, BCDE, CDEF; by fives: ABCDE, BCDEF; and finally one sequence covering the entire performance: ABCDEF. All of these possible sub-sequences of the basic six-beat performance, can be arranged on top of each other to form the arrangement shown in Figure 2.

Originally the scape plotting method was designed for structural analysis of harmony in musical scores ([2] and [3]). However, it has also been applied to audio-based harmony analysis[1] and timbral analysis[4].

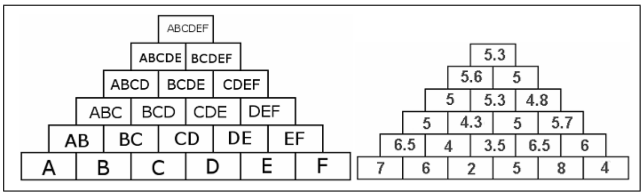

Figure 2. Scape plotting domain (left) and an example application of averaging in each cell (right), where the original data sequence is (7,6,2,5,8,4).

4 COMPARATIVE PERFORMANCE SCAPES

What operation is done in each cell of a scape plot is arbitrary. The plot on the right in Figure 2 shows the application of averaging in each cell. In the following subsections, the calculation for each cell is done using the following steps:

- Choose one performance to be the reference for a particular plot.
- For each cell in the scape plot, measure the correlation between the reference performance and all other performances, then make note of the performance which yields the highest correlation value.
- Color the cell with a unique hue assigned to that highest-correlating performance.

Note that the actual correlation values are thrown away in this variation of the scape plot. This is primarily because

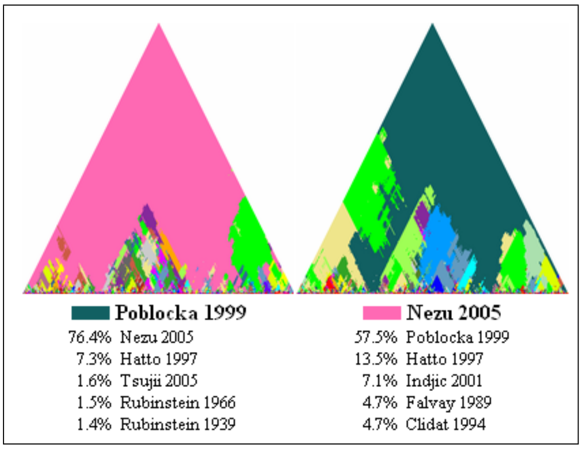

Figure 3. Timescapes for two performances of mazurka in C major, 24/2, showing teacher/student pairing, each showing large regions of best-correlation to each other (out of 35 performances).

the plots would become too complex and confusing if it were kept (for example displayed as gray-scale mask on the indexed performance colors). Other plot variants may display raw correlation values such as one that correlates half-sine arches to performance data for identifying phrasing structure.

4.1 Timescapes

Figure 3 demonstrates a pair of similar performances found in the set for mazurka in C major, Op. 24, No. 2. Mutual best matching seen in this figure indicates a strong link between two performances and is less likely to be caused by chance. However, other structures seen in this figure are more likely to be random links to other performances with no interesting relationships. The total area covered in a plot by a particular performance is also an indication of significance, but less so than mutual similarity between two particular performances. In this case the performance on the left contains an area of 76% from another particular performance, and that performance in turn contains 58% by area of the original performance. Who was influenced by whom cannot be deduced from the plots. They only show that there is a strong relationship between the two performances in this case. Clues as to what is going on can be gleaned from the fact that the performance on the left was recorded in 1999 and the one on the right in 2005; also the performer on the right did post-graduate studies with the performer represented on the left.

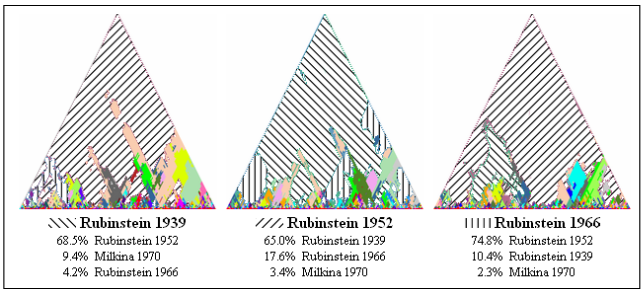

Figure 5. Timescapes for three performances of mazurka in B minor, Op. 30, No. 2. showing early, middle and late career performances by Arthur Rubinstein.

It is often useful to include the average of all performances in the collection of a piece of music being analyzed so that minor and random relationships between performances are hidden by the similarity to the average performance which is usually quite strong. Figure 4 demonstrates the effect of including the average performance along with the other real performances (compare to Figure 3).

In all five mazurkas examined comprehensively so far, all performers for which we have multiple recordings of show very strong relations to each other, regardless of the amount of time between the recordings. In Figure 5, three recordings of Arthur Rubinstein are displayed—an early, middle and late career sampling covering a time period of 25 years. In each case, the closest performance to the reference is another Rubinstein performance.

4.2 Dynascapes

Beat-level tempo is fairly unique to each performer, and when there is a strong mutual similarity between performers, it is usually not likely to be a coincidence. For dynamics (beat-level amplitude measurements in this case), the uniqueness is less pronounced due in part to the composer writing basic loudness guides such as *forte* or *piano* in the composition or data extraction accuracy. Dynamics (as extracted in this study) are less unique to a single individual performer, and a greater likelihood of random patterns

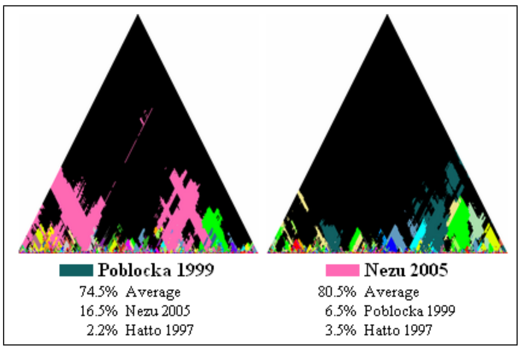

Figure 4. Same performances as in Figure 3, but with the average of all performances included (black).

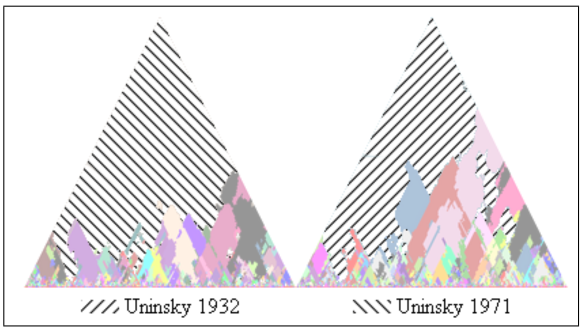

Figure 6. Two dynascapes of mazurka in C♯ minor, 63/3, showing early/late career pairing of performers.

make the plots more difficult to interpret than when using tempo data. Also, it is possible that tempo expressivity is more static between performances, while loudness is easier to consciously control.

However, Figure 6 shows some nice mutually similar dynascapes for the same performer, recorded almost 40 years apart. In this case, the performer is closest to his dynamic interpretations in these two performance than to any of the other 58 performance of the same work which were examined. Also consider that the performances were recorded in very different technological eras, the first in the time of 78 rpm records, while the later one in the 33.3 rpm era.

4.3 Scape plots of parallel feature sequences

For Pearson correlation calculations, the ordering of the data is not significant as long as the sequence order is identical for both performances. But to generate multi-feature scape plots with a structure similar to the single-data forms, the independent values are interleaved in the correct time order so that the structure in the scape plot remains analogous to the single-sequence plots. To combine tempo and dynamics for comparison between performers, the time series of each feature are interleaved. Here are examples of two data sequences for tempo and dynamics to be mixed:

$$t = (t_1, t_2, t_2, t_4, ..., t_n) \tag{4}$$

$$d = (d_1, d_2, d_2, d_4, ..., d_n) \tag{5}$$

To mix them together with equal strength, create another sequence of joint features which interleaves tempo and dynamic values:

$$J = (J_{t,1}, J_{d,1}, J_{t,2}, J_{d,2}, ..., J_{t,n}, J_{d,n}) \tag{6}$$

To minimize the effect of mixing unrelated data in such a manner for the correlation calculations, the standard deviation and mean of the two sets of data should be equivalent. In this case the tempo values are left unchanged since they contain more performance information to start with:

$$J_{t,n} = t_n \tag{7}$$

while the loudness sequence's standard deviation and mean are adjusted to match that of the tempo sequence:

$$J_{d,n} = s_t \left(\frac{d_n - \bar{d}}{s_d} \right) + \bar{t} \tag{8}$$

where s_x means the standard deviation of a sequence x, and $\bar{x}$ represents the mean value of a sequence x. The joint sequence can either be created globally, or locally based on the sub-sequence data (the latter would not work well at small timescales).

Figure 7 demonstrates the benefit of finding a performance match which is probably not random. When only time data is compared, there is little direct matching between the two performances. Comparing dynamics alone gives a stronger match between the performances, but is difficult to ascertain if the match is relevant due to the limited range for dynamics between performances. However, when both time and dynamic data are processed in parallel into a scape plot, the match between the performance becomes clear, and is likely to show a direct relation between the performance rather than a random occurrence.

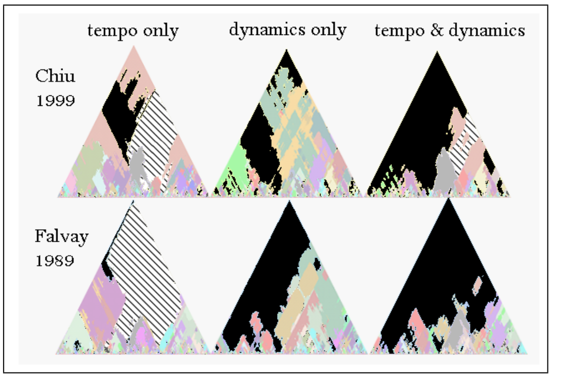

Figure 7. Tempo, dynamics and joint data plots. Black regions indicate mutual best matches. Striped region indicates a third performer common to both.

5 CONCLUSIONS AND FUTURE WORK

Significance of correlation measurements is difficult to assess in performance data since it is hard to statistically model a performer. So the precise meanings of the color patterns which emerge are not easy to pin down. Scape plots are a step towards identifying significant relations and can show where in a performance similarities are occurring.

The most difficult aspect of the plots is determining how relevant the best matches between performances are. Large patches of color do seem to be more significant, but not always. In particular, if a patch of color starts from a point and widens as it rises in a plot, it is most likely due to chance. Mutual best-matches between performers seems to be a good indication of significance, and sharp boundaries between color regions also tend to indicate more significant matches.

Tempo data in particular can be a superposition of several types of performance features. In mazurkas, for example, the low-frequency tempo component (phrasing) can be controlled independently by the performer from the high-frequency mazurka metrical pattern (where the first beat is typically shorter than the other two) and time accentuation of notes. Thus, it would be useful to identify and extract single performance features and compare them in isolation as well as in composite.

6 REFERENCES

[1] Gómez, Emilia and Jordi Bonada. "Tonality visualisation of polyphonic audio", *Proceedings of the International Computer Music Conference*, Barcelona, Spain, 2005.

[2] Sapp, Craig. "Harmonic visualizations of tonal music", *Proceedings of the International Computer Music Conference*, Havana, Cuba, 2001. pp. 423-430.

[3] Sapp, Craig. "Visual hierarchical key analysis", *Computers in Entertainment* 3/4 (October 2005). ACM Press; New York.

[4] Segnini, Rodrigo and Craig Sapp. "Scoregram: Displaying gross timbre information from a score", *Proceedings of the International Symposium on Computer Modeling and Retrieval*, Pisa, Italy, 2005.