Extracting Musical Structure from Multi-Modal Performance Analysis

Jennifer MacRitchie
Centre for Music Technology, University of Glasgow, UK
j.macritchie@elec.gla.ac.uk

Professor Graham Hair
Department of Music, University of Glasgow, UK
graham@n-ism.org.uk

Dr. Nicholas J.Bailey
Centre for Music Technology, University of Glasgow, UK
nick@n-ism.org.uk

Stuart Pullinger
Centre for Music Technology, University of Glasgow, UK
s.pullinger@elec.gla.ac.uk

Proceedings of the fourth Conference on Interdisciplinary Musicology (CIM08)

Background in computing/performance analysis. Past analyses of performance has utilised solely audio [Lindstedt,2004] or MIDI recordings, but further research into the importance of bodily movements and physical gestures [Davidson,2005] has since exposed a wealth of untapped information. The use of multi-dimensional recordings has only recently become prominent [Camurri,2003], transcending the shortcomings of each individual data stream. Performance information can be used to emphasise higher-level compositional issues that may not be obvious through traditional analysis methods.

Background in music theory. The finale of Chopin's B flat minor piano sonata Op.35 has long proved problematic for traditional musical analysts and recently for computer-assisted analysis [Lindstedt,2004]. In this composition, the search for immanent structural elements such as the initiation of new musical themes, arrival points and climax points is not immediately obvious from simply looking at the score. The quantitative methods and multi-modal view of this proposed performance analysis provides tools to view the music beyond the score and so enable an extended analysis of the piece.

Aims. We aim to analyse a number of live professional piano performances of first Chopin's Prelude in A Major Op.28 No.7 and secondly the finale of the B Flat minor sonata Op.35, the former providing control results enabling accurate analysis of the latter. These performances are recorded through audio, MIDI and video, making use of data storage technologies developed at Glasgow University’s Centre for Music Technology, to investigate issues in thematic/harmonic structure in an otherwise structurally ambiguous composition.

Main Contribution Method: Each performance is recorded using microphones, an optical motion capture system and a Moog Piano bar. The bar rests on top of any 88-key piano and using infra-red beams, converts movements of keys into MIDI. Timing information is retrieved from MIDI using a note-matching algorithm [McGilvray,2008], dynamics from the spectrum of the audio and movements from image-processing software. The exact duration of notes are calculated from the combination of audio and MIDI. The information is stored as performance data in PML [McGilvray,2008] – an extension of MusicXML enabling the inclusion of performance information - and movement data in the GMS(Gesture,Motion,Signal) format [Luciani,2006]. These files are stored alongside the score in a PostgreSQL database [Pullinger,2007] within which queries are performed based on points of agreement and change within the performances. The results are then displayed over the score in a format readable to musical analysts.

Implications This work will provide further insight into the correlation of performance parameters with structural form and enable more extensive musical analysis of structurally ambiguous pieces. The intricate measurements of each performance require expertise in such areas as signal processing, image processing and physical modelling, whilst storing and manipulating the gathered data requires extensive database and computer systems knowledge. Expertise in music theory and performance is crucial to the design of all aforementioned systems and analysis of the results, making inter-disciplinary collaborations essential to project.

Musical Analysis

Music analysis attempts to describe the melodic and harmonic relationships that occur within a particular composition, using a series of traditional methods which vary from analyst to analyst. One of the first tasks of a traditional music analysis is the segmentation
of the composition on the basis of structural features and the different characters and functions of the different sections. Important aspects to note are points of change such as the initiation of new musical themes, arrival points and climax points, which can indicate the conclusion of a preceding harmonic or melodic progression. In some compositions, it is not immediately obvious from simply looking at the score where many of these points are. In cases of ambiguity, the comparison of different performances is crucial in order to identify not just the correspondences between points of change in different performances, but the extent and limits of the degrees of change. In these instances, using computers to complement musical analysis has been the next logical step. These methods do not attempt to implement the processes of a traditional music analyst, but are used to assist and in some cases, extend existing analyses.

Lindstedt's work on computer-assisted analysis of the finale of Chopin's B Flat minor sonata [Lindstedt,2004] using score-processing program Humdrum [Humdrum], searched for melodic and harmonic patterns in an attempt to clarify structural form. Lindstedt considers formal analyses such as those by Rosen [Rosen,1999], Tuchowski, Kholopov and Leichentrett [Tuchowski,1996] which diverge widely in their view of the function of the first four bars. Some place these measures as an introduction to a theme beginning at bar 5 whilst for other readings, the initial theme begins at bar 1. As the results of the computer analysis disclose only a general indication of the form, and no more detail than the formal analyses discussed previously, Lindstedt suggests that a thorough analysis of the musical structure may be acquired through combining the score analysis with performance analysis.

Performance Analysis

It is suggested [Cook,1987; Lester,1995; Barolsky,2007] that just as musical analysis informs performance, a performer acts as a musical analyst. The performer's “analysis” occurs during practice [Rink,2002], where each part of the music is re-considered and re-shaped as the performer's appreciation of each cadence in the context of the whole composition develops. This suggests that the analysis of performance information can emphasise higher-level compositional issues that may not be obvious through traditional analysis methods.

Only recently has performance analysis research extended to analysing recordings of multi-dimensional information in live performances, thus transcending the shortcomings outlined by Eric Clarke [Clarke,2004] of work utilising solely audio or MIDI recordings. Further research into the importance of bodily movements and physical gestures [Davidson,2007] has since exposed the wealth of untapped information existing in the visual dimension of a musical performance. Existing work on extracting meaningful physical gestures particularly from piano performance concludes that the most information is in the head sway and the hand movements of the performer [Clarke,1998; Davidson,2005].

This research will initially focus on the hand movements of the performer in combination with tempo, dynamics and MIDI information. Further work may extend to using head movements and other bodily gestures. Camurri et al [Camurri,2003] record multi-dimensional parameters from a piano performance, and concentrate on the auditory and visual cues which communicate the performer's emotion. The EyesWeb image analysis tools for expressive gestures [EyesWeb] used can process measurements such as the quantity of motion in general body movements, such as head sway and upper torso movement. For this particular project, we also want to acquire extremely detailed data about the precise position of the hands in piano performance, which will require a more thorough image processing system, detailed in the section on Multi-Modal Acquisition, which accurately tracks each part of the fingers whilst also handling occlusion.

Multi-modal performance analysis produces an immense amount of data for each performance and so the existing traditional analyses of the piece being performed will be used as starting points to determine the manipulation of performance parameters at significant structural boundaries, whether it be at the end of a phrase or concluding a section of the piece.
Method

This paper proposes to use the combination of multi-modal performance analysis alongside more traditional methods to extend analyses of more structurally ambiguous pieces. In order to correctly identify the parameters correlating with the motivic/harmonic structure, a control experiment is designed which uses performances of a structurally explicit piece to determine the style of the performer and attempt to identify how they are manipulating parameters at certain points in the composition.

Chopin’s Prelude in A major Op.28 No.7 has been chosen for such a purpose due to its fairly obvious musical structure. This 16 bar piece shown in Figure 1, can be segmented into two main sections each containing four rhythmically identical two-bar phrases. For both sections, these four phrases may be considered to form an antecedent-consequent structure. Two major structural boundaries therefore occur at the ends of bars 8 and 16, bar 16 being the more significant as it is the end to the piece. A simple analysis of the piece [Kresky,1994] suggests the harmonic progression in the bass line which changes between the tonic and dominant chords each phrase, particularly emphasises the climax point at bar 12.

The correlation between the performance nuances and the motivic/harmonic structure of the prelude for each performer will serve as a set of control results. When searching for evidence of structural elements as a result of performance information in a structurally ambiguous piece, the particular performance style i.e. how they convey phrase endings and arrival points with manipulation of time, dynamics and movement will be used as a basis for comparison.

The overall objective is to enable extensive analysis of compositions with ambiguous structures. The finale of Chopin’s B flat minor sonata Op.35. has been chosen as such a piece. Occurring immediately after the famous funeral march, analyses of this piece have been widely divergent. As can be seen from the excerpt in Figure 2 showing the first eight bars of the piece, each note is of quaver (or eighth-note) length and this continues throughout until the final fortissimo chord.

Figure 1. Chopin’s Prelude in A Major Op.28 No.7

Figure 2. First eight measures of Chopin’s B Flat Minor Sonata Finale Op.35

Despite the ambiguity in its structure, this is a piece performed regularly by professional pianists, suggesting the possibility of underlying performance traditions.
Multi-Modal Acquisition

Several professional piano performances of Chopin's A major prelude Op.28 No.7 and the finale of Chopin's B Flat minor sonata Op.35 are recorded using microphones, an optical motion capture system using a high frame rate camera, and a Moog MIDI bar to capture audio, video and MIDI streams. Using multi-streams of data transcends the shortcomings of each individual mode and means that the performance is not limited to simply audio information, but instead attempts to capture the "essence" of the music being played.

Audio. Each performance is recorded using the multitrack recording application Ardour [Ardour]. This audio data can provide information that is unavailable on the MIDI stream, e.g. MIDI can provide note onsets and offsets but only when combined with the audio stream can the actual length of the note be determined. This can be due to factors such as the use of sustain pedal and psychoacoustic masking effects. An example of the psychoacoustic effect of temporal masking is when a sudden stimulus loud sound masks other quieter sounds which occur either immediately before or after the stimulus.

MIDI. MIDI information is captured using the Moog Piano bar [MOOG] which consists of a scanner bar that rests on top of any 88-key piano and a magnetic pedal sensor that rests beneath the pedals. The scanner bar sits against the fall board of the piano and is designed to be un-intrusive to the performance. The scanner bar's "teeth" are positioned between the black keys and lie just above the white keys. Key depressions are sensed by the detection of reflected infrared beams projected by these “teeth” directly onto the white keys and through the black keys. A MIDI note-on at a white key is triggered by an infrared beam being broken and for a black key by an infrared beam being detected. The device also outputs the note-on velocity information. The additional magnetic sensor which lies beneath the pedals, detects the depression of the una corda and sustain pedals. The sensors feed the note information to the Control Module where it is transformed into MIDI information. It is then recorded through the open-source sequencer Rosegarden [Rosegarden]. The MIDI data will provide information on what key is pressed, its onset time, its offset time, onset velocity and also which pedal is depressed and its onset time.

Video. It is necessary to use a high frame rate camera when recording the visual aspect of a piano performance, as a skilled pianist can play up to 30 sequential notes per second [Rumelhart,1982]. The AVT Guppy F-046 camera has a Region of Interest facility, allowing a smaller size of frame to be transmitted at a higher frame rate, and so can reach up to 60 frames per second. Video data is recorded through the application Coriander [Coriander] as a raw video file. The camera is placed with a birds-eye view, 80cm above the keyboard, producing a 780x216 pixel frame, allowing coverage of 75% of the keys. The far upper and lower registers of the keyboard are not covered as this is sufficient coverage for the two pieces being examined and increasing it would decrease the picture resolution. UV reflective paint markers are placed on each joint of each finger as seen in Figure 3 and a blacklight sits level with the camera in order to make the markers fluoresce. UV paint has been used as the pixels representing the markers in a frame will peak at a certain colour (in this case, yellow) making it easier to subtract the background image from each frame.

Figure 3. Video frame showing UV blobs

The image processing software is written in C++ using the Intel OpenCV Image Processing library and the bolt-on OpenCV blob extraction library [OpenCV].

The software reads in the AVI lossless encoded video and processes it frame by frame. Each frame is passed through colour filters and submitted to the blob detection algorithm. The algorithm scans each raster image frame line by line and records connecting regions of similar colour. Once the detected marker set is split into two 'hand'
objects using K-means clustering [Hartigan, 1979], the detected position of each marker is then submitted for comparison against a physical model of the hand. These constraints are taken from Rijpkema's model [Rijpkema, 1991] but also account for the extra constraints in the context of piano performance. Each "hand" point set is split into metacarpophalangeal joints and interphalangeal joints between the proximal and middle phalanx and between the middle and distal phalanx for the four fingers of the hand. The thumb of each hand will only have a point for the metacarpophalangeal joint and the interphalangeal joint between the proximal and distal phalanx as it lacks a middle phalanx. From the diagram in Figure 4 we can see that the markers on each of the interphalangeal joints of the fingers all lie on a curved line and this will be true with the majority of hand shapes encountered in piano performance.

Using the general shape of the hand, the markers can be grouped, thresholded by calculated distances and angles, into a group of points each for the metacarpophalangeal joints and the two sets of interphalangeal joints. This is demonstrated by the screenshot in Figure 5 where the points are grouped in yellow for metacarpophalangeal joints, purple for the joints between the proximal and middle phalanx and blue for the joints between the middle and distal phalanx. Figure 6 shows the same point set redrawn so the points for each finger are grouped together.

The software can also estimate the position of any "lost" markers by calculating the average transformation of the other points in the point set between each frame. These transformations are calculated using affine transforms.

The image in Figure 7 shows the metacarpophalangeal joints. Four markers from the left hand and five from the right hand have been detected. The wrist point is denoted as a white cross, with coloured crosses corresponding to the first, second, third etc. fingers of each hand. Yellow lines are drawn connecting the markers for each hand for observation of its shape.

This system requires the full detection of all markers in a test frame before tracking can begin, as the estimation algorithm calculates the new position based on the marker's last tracked position as well as the transformations between frames for the other markers. The consecutive frame from these results shown in Figure 8, shows such an estimation (denoted by the red circle) as one of the left-hand markers becomes occluded.
These resulting positions of markers can be analysed for velocity and positional information against the context of the music being performed. This software also enables further investigation of parameters such as finger curvature and its direct relationship to the resultant sound. Fingering patterns may also be analysed through encoding the positional information of the keys and correlating this with the MIDI data to show exactly which finger played which note.

**Data Storage and Display**

The processed visual data is written as a set of coordinates per frame to standard GMS (Gesture, Motion, Signal) files [Luciani,2006] structured to represent each marker of each finger of each hand. Distances and velocities of markers can be calculated from these files and plotted against the corresponding frame numbers. The performance data is stored in Performance Markup Language (PML) [McGilvray,2008], an extension of MusicXML [Good,2001] which enables the inclusion of performance information. The performance notes are stored in a separate section at the bottom of the MusicXML file as onset and offset times, MIDI note number and a specific ID for each note. The file is then processed using a note-matching algorithm [McGilvray,2008] which identifies successfully matched performance notes by including a reference in the file to its corresponding score notated note.

These files are parsed to populate a PostgreSQL database [Pullinger,2007] placing the performance information alongside the notated score information. Queries are sent to the database based on points of agreement and change at suggested structural boundaries. These queries are formed in such a way that the calculated performance information is returned alongside the score note ID, allowing the output to be displayed in the form of a score with the performance information annotated above each note. Among the performance parameters evaluated are hand movement, timing and dynamics, which can be displayed alongside and cross-referenced with the musical information of the notated score. Further work will incorporate the GMS files into the database to enable movement information to be displayed alongside the score.

**Preliminary Results**

Multi-modal analysis of the Chopin A major prelude Op.28 No.7 provides a basis of the correlation between performance parameters and musical structure, which can then be used as a standard to search for structural boundaries of other musical compositions. In this study, multiple performances will be used to analyse the points of agreement and difference between performers, to minimise the search for arrival points and phrase/section endings within the composition.

Two pianists were recorded performing Chopin’s A major prelude Op.28 No.7. For the purposes of these results they shall be referred to as pianist A and pianist B. Their performance data was stored in a database and queries were performed to produce results of each performance and a comparison between the two for measurements such as note onsets and keypress durations.

The images in Figures 9 and 10 annotate the score of the prelude with each pianists' keypress durations for the notes of the treble clef. From keypress duration and impressionistic observations from the audio we can examine the articulation pianists have used to emphasise particular notes. What is evident from first observations, is the clear separation of the prelude into two sections by pianist A. The durations at measure 8 indicate a large pause before commencing with the second section of the piece. Pianist A makes small pauses at the end of each phrase, whilst also denoting large durations to the anacrusis before measure 1 and also the anacrusis before measure 9. Pianist B exhibits large durations for the anacrusis of every phrase. This act of holding back at the anacruses is considered to be a recurrent style in Chopin's compositions [Thieffrey,1937].

Pianist A and B’s keypress durations follow the same general shape for the entire
prelude. Both pianists make an emphasis of the F sharp major seventh chord in measure 12, considered to be an arrival point.

Figure 9. Keypress durations for Pianist A's performance of the Chopin prelude.

Figure 10. Keypress durations for pianist A's performance of the Chopin prelude.

Note onsets and particularly the times between onsets i.e. inter-onset intervals can be used to indicate the rhythm and tempo of a performance. As each phrase of the prelude is rhythmically identical, we can also see the performer’s manipulation of rhythm more clearly. In this case, the comparison of the two performers, as shown in Figure 11, revealed the most information. The absolute measure of the bars are equal to the subtraction of pianist B's note onsets from pianist A's plus an offset factor of the difference in start time between the two performances. The point of interest here is the gradient between the bars displaying the difference between onset times. When the gradient is zero, the two performers are playing at the same tempo. Where a positive gradient occurs, pianist A is playing at a faster rate than pianist B and vice versa for when a negative gradient occurs. The comparison of note onsets reveals an agreement from measures 10-12 in terms of tempo. In every other respect, pianist A is playing at a faster tempo than pianist B.

Figure 11. Comparison of note onsets for pianist A and B's performances of the Chopin prelude.

Looking at inter-onset intervals further explained the tempo differences for each performer. From Figure 12 we can see that pianist A plays the first three phrases at roughly the same tempo, slowing slightly at the end of the fourth phrase. The second section is assigned with a slowing of tempo for each phrase, exaggerated particularly for measures 11-12 where the antecedent presents more to resolve in the consequent in terms of harmony. Pianist B assigns a slight slowing of tempo to each phrase, exaggerating this also around measures 11-12.
What can be suggested in terms of performance parameters from these exemplar results is that both performers assign longer note durations and tempo variations to points of harmonic arrival and major sectional breaks. Smaller phrase endings are characterised by a short pause and anacruses employ longer note lengths to produce a Chopin-like hesitation before the start of a phrase. Harmonic arrival points can be characterised by longer note durations and the performers appear to be in agreement for the tempo leading up to this point in the piece. These results suggest that performers convey medium-level structural information such as phrase endings with parameters such as articulation and tempo, much along the lines of Parnicutt’s investigation into accents and expression in piano performance [Parnicutt, 2003] which links musical structure with the position and strength of several types of accents in a local scale. The full multi-modal results can be analysed more extensively to further clarify this relationship between medium level structures and multi-modal performance parameters.

Further work will incorporate the image files into the database allowing full multi-modal analysis of each recording. At the time of publication, the physical gestures of the two performers had been written to GMS files. The results of these movements, in distance moved, position, velocity and acceleration information could be displayed against the corresponding frame number, but to derive the most information from this data, it is absolutely necessary to display movements aligned with the score. In combination with the audio and MIDI information, these movements, can identify phrase shaping and emphasise structural features illuminated by the audio and MIDI streams. Further extensive queries covering the range of parameters varied in piano performance will be performed. These multi-modal results will then be used as a basis to search for structural features in each pianist’s performance of Chopin’s B flat minor sonata finale.

Conclusions

A system has been presented to accurately acquire detailed performance gestures from a multi-dimensional perspective. These are stored meaningfully in a computer database which presents the results of analytical queries based on the musical structure of the piece in a form useful to musical analysts. Preliminary results correlating long keypress durations with sectional breaks and with harmonic arrival points for each performer are presented for a control experiment using a simple and structurally explicit prelude. The full multi-modal results of which can be analysed more extensively to provide a basis of relationships between performance parameters and structure for each performer. This is then used as a basis to search for
structural boundaries of other ambiguous compositions, particularly the finale of Chopin's B flat minor sonata.

References