Elucidating Musical Structure Through Empirical Measurement of Performance Parameters

Jennifer MacRitchie

16th February 2011

This thesis is submitted in fulfilment of the requirements of the degree of Doctor of Philosophy

Department of Electronics
& Electrical Engineering

University of Glasgow

Abstract

The differences between a musical score and an instance of that music in a performance, communicates a performer's view of the information contained in that score.

The main hypothesis in this thesis is that by measuring quantifiable parameters such as tempo, dynamics and motion from live performance, performer's interpretation of musical structure can be detected. This will be tested for pieces for which the structure is explicit and obvious, and then used to discover musical structure from looking at patterns of aural and visual performance parameters in performances of more ambiguously structured pieces.

This thesis is in two strands. The first part covers the acquisition of multimodal parameters in piano performance. This will explore current technologies in acquiring MIDI information such as accurate onset timings and key velocities as well as motion tracking systems for measuring general body movements. A new cheap, portable and accurate system for tracking the intricacies of pianists' finger movement is described as well as methods and tools available for analysis and visualisation of musical data. The second strand of this thesis will explore uses of these capture systems in empirically measuring performance parameters to elucidate musical structure. Two experiments follow which test the hypothesis of detecting musical structure from parameters such as tempo, dynamics and movement, before using these patterns as a basis for discovering structure in performances of the finale of Chopin's B flat minor sonata.

Body movement is discovered as an indicator of phrasing boundaries, which when combined with the measured aural parameters provides interpretations of the performed music. Phrasing boundaries are identified correctly for the control piece (Chopin's Prelude in A major Op.28, No.7) and consequently for the first test piece (Chopin's Prelude in B minor Op.28 No.6). The proceeding experiment identifies performers' style of phrase endings through performances of the control piece and tests them against patterns found in the second test piece (Chopin's B Flat minor Sonata Finale). Five out of the six performers confirm the musicological hypothesis that bar 5 is not the entry of a new theme but the continuation of the the theme beginning in bar 1.

Acknowledgements

I would like to thank my supervisors Dr Nick Bailey and Prof Graham Hair and Dr John Williamson for their invaluable assistance, guidance and support. Special thanks also go to Stuart Pullinger and Douglas McGilvray, whose specially designed tools and software were pivotal in conducting this research and whose help in the design of these experiments was crucial. Special thanks also to Bryony Buck, who provided an opportunity for us both to conduct some truly interdisciplinary research and also to Tom O'Hara for his technical assistance over the years. I would also like to thank my colleagues at the Centre for Music Technology, Bill Evans, Ben Hillman and Graham Percival for their help. This research has been EPSRC funded.

To Annie, Doug and Elisa for their endless coffees and support as well as a different perspective on musical analysis, to Lynsey and Maria for lunches of reprieve and finally to my close friends outwith the department and my family without whose support, none of this would have been possible.

"*I can do all things through Christ who strengthens me.*" Philippians 4:13

"The function of music is to release us from the tyranny of conscious thought."
Sir Thomas Beecham (1879-1961)

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

To understand what is going on in a piece of music, and the different characters and functions of each group of phrases, musical analysts will use several techniques to break down the rhythmic, melodic and harmonic relationships between each set of notes. The first step of this is referred to as segmentation, breaking down the piece into different sections, and using smaller units such as motifs and phrases to build up a hierarchical picture of how the piece works. This agrees with Lerdahl and Jackendoff's theory [70] that music is made up of perceptually discrete elements organised into hierarchical structures. Can the boundaries of these discrete elements be measured quantifiably? Traditional detection measures of these structures are performed through methods that focus on looking at the score. These traditional analyses attempt to describe a piece of music and as different analysts have different perspectives and techniques, there are often cases where these analyses diverge. Some classical pieces are so ambiguous that analyses cannot agree on the entry of the main theme. An example of this is the finale movement of Chopin's B Flat Minor Sonata op.35. Despite this, there exists a large number of audio recordings of these pieces suggesting they are still widely performed regardless of their ambiguity.

The most interesting information in a Western classical music performance seems to reside in the measurable differences that exist between what information is contained in the notated score and what is actually performed. A performer instills their interpretation and understanding of structure into a composed piece of music by manipulating parameters such as expressive timing, changes in intensity or dynamics, choices of timbral sound and tempi creating different moods and characters for each part of the music. Taking for an example the opening phrase of Beethoven's Fifth Symphony, these bars tells us implicitly a lot of information about what is to come in the following music. The opening motif as seen in Figure 1.1 is repeated and changed several times throughout the first movement. This is information that can be seen by just looking at the score. However, there exist many different performances of this symphony by several different orchestras. Conductors can spend many hours in rehearsals focusing on the opening bars, changing the stress of the rhythm, the tempo, the dynamics, the balance of instruments and many other parameters as the performance of these opening bars sets the tone for the entire performance. This suggests that performance nuances carry certain information about the music being performed. However, as performers can use many varied expressive features to express essentially the same structural feature, this means the relationship from performance to analysis of a piece of music is not always straightforward. Referring back to performances of the Beethoven symphony, these opening few bars can be varied quite entirely across orchestras and conductors depending on their own personal interpretation.

Figure 1.1: The opening of Beethoven's Symphony No.5 First Movement

So by measuring the quantifiable differences between score and performance in the use of parameters such as tempo and dynamics, the structural information being communicated through the performance could essentially be measured. The question then arises as to whether audio recordings can be used to measure these communicated features accurately as opposed to a live performance.

Despite most music being listened to on mp3 players and i-pods in recent times, live performances of music regardless of genre are still widely in demand and well attended. Reasons for this have been documented in a survey of listeners' habits with the results showing that audiences prefer live performances when possible. The most popular answers were 'atmosphere' or 'ambiance' whilst in second place was the response 'communicating with performers' [120]. Research exploring the visual element of music suggests that performers' physical movements have an impact on what is communicated to the audience [123]. Classical pianist Glenn Gould was renowned for his strange posture and erratic movements during performances, both in front of audiences and in more secluded environments, suggesting that his gestures were not simply used for communicative purposes but contained an entirely expressive purpose related to the music [38]. In other musical genres, for example in jazz performances, musicians' movements are related to a wide variety of musical causes such as 'groove', classified as relating to the beat of the music. Classification of meaningful movements i.e. physical gestures and exploring current research in the area is discussed in Chapter 2.3. Another example of the link between music and gestures lies with Bobby McFerin, now a famous improvising beat-boxer, who uses his voice to emulate a number of different instruments when he performs. When watching videos of his performances you can see the movements he makes relate to the strumming of a guitar, or the movement of his fingers on the microphone look like a trumpet player pressing down valves. Recently at the World Science Festival 2009 as part of the Notes and Neurons talks [9], he made the audience intuitively sing notes of a pentatonic scale by simply jumping at different points on the stage. This link between movement and music is also explored in Chapter 2.3. Seeing as visual movements can be important in conveying the meaning behind the music being performed, when measuring quantifiable parameters to detect the performer's interpretation, motion is a factor which should be included, particularly when examining a live performance.

1.1 The performer as analyst

In Western classical music, the performer provides the medium through which the composer's ideas can be conveyed to a live audience. A performance therefore requires a demonstration of understanding of the piece by the performer. Whilst learning new pieces, performers will refer to previous recordings as well as looking at the score and various existing analyses, to determine the function of each section and how best to convey what is going on in the piece. Throughout practice sessions, this 'analysis' of the piece by the performer will be refined by experimenting with different sounds and different uses of dynamics and tempi [104]. The idea that a performer acts as an analyst in this situation is echoed by Cook, Lester and Barolsky [29, 71, 14].

Berry states that

> ...there can be divergent , reasonable concepts of structure in any given piece is a fundamental rule of existence for the analyst unfettered by bias

so many 'correct' or 'authentic' performances can exist even though they may be completely different from each other. Based on this, it becomes interesting to look at places where performers agree in their interpretations of a given piece and equally interesting to examine the places where performers disagree or diverge.

Performance traditions, or places where performers agree on certain aspects of the music can change over time. Bach is played in a completely different fashion to the way it was over a hundred years ago. Performance interpretations can change completely from generation to generation despite the notated music remaining the same. This is different for music that is not based on notated scores e.g. some folk music is passed down aurally, and although the structure remains the same the notes and rhythms can be entirely different. For Western classical music, the score becomes useful as a starting point for each performer as the differences between interpretations and the notated score can be examined. Although it is not an entirely explicit document, the score contains information on structure and arguably emotive qualities of the music [119].

1.2 Aims

Concerning performers as analysts, the work discussed in this thesis will be focussed on the main research question

- Can musical phrasing structure be detected from multi-modal performance parameters?

This research uses performance analysis of piano performances to detect musical structure by analysing patterns in tempo, dynamics and movement. These patterns will be used to identify phrasing boundaries. Our definition of a phrase is a small structural element that acts as an upbeat to a cadence, which at some point reaches a climax. Phrasing boundaries tend to be characterised by a slowing of tempo and shaping of dynamics and so patterns are searched for in the recorded aural cues that correspond to such boundaries. Patterns within performers' movements during performances are investigated to also establish how their visual gestures coincide with phrasing. This model is then used to discover structure in pieces where analyses seem to be widely divergent.

In answering this main question, a set of sub-questions arise:

- Is movement related to phrasing structure?
- Can this phrasing structure be detected through patterns of tempo, dynamics and movement for pieces where the score-analysis of phrasing is explicit?
- Can these patterns of tempo, dynamics and movement be used to look at a performer's interpretation of structure in a piece where the score-analysis of phrasing is ambiguous?

To be able to answer these questions, multi-modal capture systems must be designed to record and analyse the data. Therefore, this thesis works with two main aims:

Aim 1: To design capture systems, storage and visualisation formats that allow accurate and robust methods of recording live performances and display the results useful to musicological analysis

Aim 2: To determine whether structure can be deduced from the empirical analysis of multi-modal performance parameters

This is split into the following steps:

(a) Designing acquisition systems for piano performance that record as much information as possible from a performance, are relatively un-intrusive and provide comfortable surroundings for the performer so they can accurately recreate a typical concert-setting performance.

(b) Using these different types of systems in experiments which analyse the relationship between body movement and phrasing structure. This will be achieved by recording pianists performing Chopin preludes with differing structural layouts.

(c) Analysing these gestural cues in conjunction with audible parameters such as tempo and dynamics again in relation to the phrasing structure of the pieces.

(d) Using these analyses as a basis to discover musical structure, particularly in pieces where traditional musical analyses seem divergent. This is based on the theory that performers are constructing analyses themselves through practice and performance, and that structure can be deduced from the differences between the score and performance.

These experiments will focus on piano performance, particularly as when dealing with movement and recording and analysing meaningful physical gestures, pianists are restricted in how much they can actually move. They cannot move the instrument, unlike clarinettists or violinists. Therefore, expressive movement is all relative to the position of the keyboard. Pianists also do not require to breathe in particular places to shape phrases such as wind and brass players do. When considering the parameters of the mechanical action of the piano that can be manipulated by performers, the key velocity is the only known variable. Whether this is changed by the force acting on the piano by the hand, and whether different forces with the same key velocities would produce a different timbre is a question still under investigation.

Studies by Suzuki [114] show that pianists' 'hard' and 'soft' touches make a slight difference in the spectrum of high pitched notes, but not to the extent that pianists may expect. Research into the elasticity of certain parameters in hammer-string contact has been physically modelled by [113] and further notes on the acoustical properties that could lead to different 'touches' having an effect on the produced note are recorded in [28]. Other variables in piano playing include the use of the pedals. Investigations into the effect of the sustain pedal also show that its use in the mid-range of the piano introduces an element of distortion in the two-step decay of the note [68] and 'half-pedalling' (where the pedal is partially depressed) shortens the decay time as opposed to notes with a fully depressed sustain pedal [67].

This research requires a system for multi-modal performance capture that will be un-intrusive to performers, robust and accurate. There is a multitude of data that can be recorded in a performance and so to avoid an overload in data processing, we must choose which data is necessary to record and how. This influences which systems are used in the overall design.

1.3 Overview of capturing and storage technologies

In order to design multi-modal capture systems, various existing capture technologies are examined for measuring timing, dynamics and visual information. Frustratingly, no one solution exists for recording all aspects of a performance. The advantage of recording pianists is the availability of many MIDI capturing appliances ranging from portable external retrofit devices such as the Moog Piano Bar, to factory installed MIDI pianos such as the Yamaha Disklavier series. These devices are explained in more detail in Section 3.2. These devices record key velocities as well as accurate onset and offset times of each note played, with the factory installed pianos even able to record the angle of key depression through time. This represents a definitive advance on the extraction of note onsets from audio recordings. This former method proves difficult when trying to distinguish between notes in a chord. This is more of a problem for piano recordings than wind or string instruments (even in the case of double stopping).

Recording video in performance is becoming increasingly important as perceptual studies show visual information is improving communication between performers and audiences (see Sections 2.3), and technologies have been developed to allow in-depth analysis of movements made during a performance. Motion capture systems vary from stationary infra-red camera arrays detecting passive retro-reflective markers placed on the performer, to simple image-processing cameras working with purely the image of the performer. Other devices using active markers such as accelerometers with gyroscopes or head tracking devices with a portable battery pack to be worn on the performer's waist can be placed on the performer's body. With motion tracking there is no single solution that will be appropriate for every situation as seen in the review of technologies by Welch [127].

This thesis explores these different technologies and designs two different systems, one of which incorporates a specially designed portable motion capture system for tracking pianists' fingers.

1.4 Summary of chapters

To clarify how these two main areas of research will be addressed in this thesis,there are three main parts: Background, Developing multi-modal capture tools and systems, and Experiments and Results.

Theory and Background

Chapter 2 provides a review of the relevant performance analysis theory and research to date, focussing on how it influences the design of these multi-modal capture systems and the following experiments.

Chapter 3 then reviews the available technology to capture performance data as well as storage and visualisation methods. The best of these commercially available products in terms of accuracy, price and portability are incorporated into the system design.

These chapters present the motivation and need for new multi-modal capture systems and storage methods.

Developing multi-modal capture tools and systems

Chapter 4 describes in detail the specifically designed motion capture system for tracking pianists' fingers. The system setup, 3D estimation and occlusion algorithms are explained in depth.

Elements of the list of commercially available capture products are combined with the specially designed systems described in the previous chapters to create separate two methodologies. Chapter 5 lists these two separate methods used to collect data. The first method looks at general body movement in tandem with MIDI and audio data whilst the second method looks at finger movement again in tandem with MIDI and audio data.

Once these data streams are recorded from performance they need to be stored accurately and presented in such a way that is useful to musicological analysis. Chapter 6 describes the storage and visualisation methods developed by colleagues at the Centre for Music Technology in order to be part of these multi-modal capture and analysis systems.

These two sets of methodologies are then used to perform experiments to answer the larger research questions on performance and musical structure.

Experiments and Results

Chapter 7 analyses the music being used as stimuli for the experiments in terms of their phrasing structure. The music for the first experiment is chosen specifically to test whether explicit structures can be detected through performance analysis and the second set of music tests for being able to discover structure from more ambiguous pieces.

Chapter 8 outlines the experiment for detecting musical structure. Results are analysed in terms of relating body movement to musical structure and then looking at multi-modal cues for phrasing boundaries.

Chapter 9 outlines the experiment for discovering musical structure. Results are analysed in terms of low-level parameters like inter-onset intervals, keypress durations, finger curvature and sound amplitude in relation to accents and phrasing.

Chapter 10 presents a discussion of results along with recommendations for further work both in designing performance recording systems as well as analysing performance data. The main conclusions of this work are presented in Chapter 11.

Part I

Theory and Background

Chapter 2

Performance Analysis

Performance analysis techniques are used for a variety of different purposes. These can be large scale studies of parameters, such as examining expressive timing across several performances of a particular piece to make comments on the general usage of tempo fluctuations for expressive purposes. Other studies involve analysing particular interpretations of a composition to determine how one performer has created this interpretation by manipulating factors such as timing, dynamics, articulation and timbre.

This next section will cover uses of performance analysis by examining various pieces of theoretical and empirical research, identifying the requirements and considerations necessary for a system designed to discover musical structure from performance data.

2.1 Music Performance Theory

By using performance parameters to elucidate musical structure, the relationship between the score and the performance must be considered as well as the role that musical analysis plays in these interpretative decisions. Music analysis attempts to describe the certain melodic and harmonic relationships that occur within a particular composition, using a series of traditional methods which vary from analyst to analyst. The first step of a traditional music analysis is the segmentation of the piece 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. Identifying the manipulation of performance nuances across several performances of the selected Chopin pieces in Chapter 7 will provide the necessary tools to highlight the communication of structural boundaries in more ambiguous compositions.

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 [74] using score-processing program Humdrum [60], searched for melodic and harmonic patterns in an attempt to clarify structural form. Lindstedt considers formal analyses such as those by Rosen [106], Tuchowski, Kholopov and Leichentrett [122] which diverge widely in their views of the function of the first four bars. Some place these bars 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.

The proposed research aims to do just this, by comparing traditional analyses to the suggested analyses provided by measurement of certain performance parameters. Combinations of tempo, loudness and movement will supply a potential segmentation of each piece performed, as is required as an initial step in traditional analyses.

It is suggested [29, 71, 14] that just as musical analysis informs performance, a performer acts as a musical analyst. The performer's "analysis" occurs during practice [104], 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. What is interesting in performance analysis is the deviations or differences between the notated score and the actual performances. Early research found that performers did not reproduce the notations on the score mechanically but that there was a deliberate manipulation of timing and dynamics added to what was explicitly written [108]. These were found not to be completely unrelated to the score but instead appeared to emphasise certain points. Todd [121] provided theoretical support describing a model of expressive timing which linked expressive devices such as rubato to key features of the musical structure e.g. cadences. As well as structural information, scores can contain implied emotions or moods as suggested by Thompson and Robitaille [119]. This research suggests that it is not just deviations from the score that should be considered, as this is expected from a human performance, but the similarities and differences between several performances.

Performance in itself then does contain a mixture of structural information and implied information about moods evident in the music. These parameters are not entirely separable, just as musical parameters such as pitch, rhythm, timing etc. should be considered as interacting and not entirely separable as Clark [26] states:

> Analysis of performance and also perception tends to treat musical parameters individually as if they are processed relatively independently by specialised psychological mechanisms but this is untrue

So in performance analysis, the context (being the score) must be examined when considering the audio, the audio when considering the visual and the performer's views on structure when considering their analysed interpretation.

Palmer's review on music performance research [86] expands on Kendall and Carterette's model of performance which encompasses the coding of the composer's ideas (the score), the recoding of these ideas by the performer (the interpretation) and the decoding of these ideas by the audience. The score can represent pitch and duration quite explicitly but information on structure, such as groupings is only implied and instruction as to precise articulation is often virtually absent. These ambiguities allow the performer a certain amount of interpretative freedom and this interpretation includes the performer's ideas on the musical composition.

The encoding part of this communication process is modelled for the performer by involving the production of audio and visual cues from the origins of the notated score. This can be seen in Figure 2.1. The performer uses movement to play the instrument and produce these sounds. Feedback is used by the performer to constantly monitor what is being produced in terms of audio and to an extent visual content.

Parncutt believes that expression in piano performance can be explained by immanent accents present in the score [90]resulting in performed accents such as

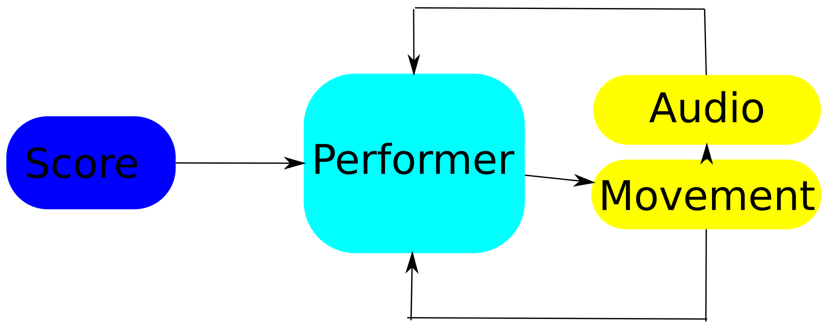


Figure 2.1: My illustration of the feedback loop of a performer

expressive differences in timing and dynamic stress. He labels certain accents belonging to time such as grouping and metrical accents as well as ones dependent on pitch e.g. melodic or harmonic accents and reductional accents which fall along the lines of Schenkerian reductions of the score. This lower level accent structure is something that will be investigated after detecting higher level phrasing structures, or as Parncutt defines them, grouping accents. Drake and Palmer investigated the interaction and independence of these accents in the presence of other accents [42]. Rhythmic and grouping accents remain constant whereas melodic accents tend to change in the presence of other accents. These represent the lowest level of a hierarchical structure [70]. Taking the hierarchical importance of the phrase as a factor, the relationship between expressive timing and musical structure has been documented such that the amount of rubato used reflects the hierarchical importance of the phrasing boundary. From this we expect that the more important the boundary e.g. the most important being the end of the piece itself, the larger the rubato will be. This phrase-final lengthening [121] is an example of how mid-level parameters such as tempo can provide clues as to the structure of the music. Establishing that theoretically, the score implies certain expressiveness by the performer, I aim to examine how we can use performance parameters resulting from the expressive interpretation to locate or suggest structure.

2.2 Performance Analysis Studies

In Eric Clarke's experiment, [27] measuring a professional pianist playing a Chopin prelude six times on a Yamaha MIDI grand piano, tempo and dynamics are plotted with respect to time i.e. the place in the score. The performer was given no directions to vary his interpretation of the piece or to stick to one interpretation. The six resultant performances therefore differ where the performer has picked out different passages of interest. He says of the expressive performance parameters:

> the force of musical expression must be understood by interpreting the function of any expressive features within the specific structural context that they occur. What may appear to be the same expressive element - an acceleration for instance - may have quite opposed functions depending on the structural context in which it occurs

[27]. This demonstrates how structural context is extremely important when considering the different performance parameters across various performances.

Another example of examining in-depth a single interpretation of a piece is the study of Martha Argerich's distinct performance of Chopin's E minor prelude op.28 no.4 [109]. Senn studied the initial four bars of the piece, attempting to discover which structural features in the score inspired Argerich's particular interpretation. A particular point of interest is at the end of the first four bar phrase, where one would expect a traditional ritardando, Argerich instead produces a mid-bar ritardando and then gains speed at the end of the bar. This is explained as instead of the last note belonging to the first phrase, it instead marks the beginning of the next phrase, hence the acceleration. This is one example of using performance data in an effort to provide a segmentation of the score. However, while much information can be gleaned from single interpretations, it is also necessary to examine large numbers of performers to suggest patterns of timing or dynamics in relation to structure.

An example of larger scale studies involving a number of performers comes from Repp's analysis of expressive timing patterns in graduate piano performances of Schumann's Traumerei [98]. This study compared those patterns of students to previously collected timing patterns of professional performers. The patterns were largely comparable across the two groups, however, principal components analysis showed the student timing patterns to be largely undeviating from each other whereas the professionals had the more divergent patterns of expressive timing. Timing profiles across the group were largely repeatable on repeated recordings when performers were asked for the same interpretation each time. That the students played with remarkably similar timing profiles as the experts is interesting. Despite differences between pianists' profiles suggesting that individuality plays a part in each performance, it is proposed that there is also a high similarity between performances. Other conclusions from the timing data concern the accelerations in the lead up to the melodic peak in each phrase, which are noted as sharing a certain parabolic fit to the shaping of each melodic gesture.

In a similar study, this time with performances of a Debussy Prelude [99], similar results were found on the whole, suggesting that the similarity between student and expert pianists' timing profiles is the result of trained musicians being able to easily interpret the timing implied by the structure of the notated music. Repp's studies argue that when evaluating expressive timing, it is not the absolute deviation from the score that should be considered but the deviation from a performance norm. Some amount of expressive timing and dynamics will always be expected in any performance. Points of agreement and departure between the timing of individual performances would therefore be more interesting to examine.

Repp's extensive study of over 100 audio recordings of performances of Chopin's Etude in E major examined expressive timing and dynamics respectively in the initial measures of the piece [102, 100]. Principal components analysis was used in both cases to determine timing strategies and profiles of expressive dynamics. Repp discovered that although there is the infinite potential for different performances, actual performances tend to be realised within constraints of what is accurate or authentic for the piece. Sampling such a large number of performances, it was found that within these limits, clusters of performances do not exist suggesting that different timing profiles are not necessarily the result of different structural interpretations. The produced principal components were therefore considered as ways of expressing the same structural features through different timing profiles. No significant relationship between timing and dynamics was found suggesting that a greater level of freedom is found by performers when forming their expressive shape of each phrase. The correlation between the grand average profiles of timing and dynamics produced an unexpected positive correlation but this was mainly due to the nature of the composition where the accompaniment is played fast and softly. This is another case where Eric Clarke's consideration of the structure of the music i.e. the context is particularly important. Correlating with just the melodic notes, the negative correlation produced was extremely low and not significant suggesting timing and dynamics are relatively independent. The main conclusions from these studies implies that performers may have more freedom in their use of dynamics than the use of expressive timing, as this is governed by certain constraints in defining what is acceptable. The different uses of these expressive parameters also implies that instead of different structural interpretations, these different profiles are ways of expressing the same structure.

Investigations concerning other keyboard instruments include Gingras and colleagues' studies recording 16 organists performing a Bach fugue on a MIDI prepared organ exploring the emphasis of phrasing through expressive timing [45]. The performers' traditional analyses of the piece were also used as a point for comparison. The largest measured tempo fluctuations coincided with the agreed structural boundaries whilst others coincided with features that were not relative to the phrasing subdivisions. Again a high similarity between timing profiles was found. An interesting point arising from this study was the non-significant correlation between the performers' formal written analyses and the analyses resulting from their timing profiles. The author acknowledges that this may be that the written task encouraged the performers to note structural analysis as they had been taught through formal music analysis classes instead of the phrasing they performed in this particular piece. This study provides a point to note when collecting performers' ideas on phrasing segmentation as their written analyses may not be exactly the same as what they perform.

The individuality of performers through different timing profiles can be measured by looking purely at the expressive timing data in studies such as those by Grachten and Widmer [57]. By measuring the final ritardandi through inter-onset interval deviations from a performance norm, a classifier determines whether the residual data can supply clues to the performer identification in performer pairs. This theme of identifying clues about performance from measured performance data is extended to searching for clues about musical structure through patterns in aural performance parameters. Examining repeated timing patterns in performances of Chopin's Etude Op.10 No.3 [112] through pattern matching and Functional Data Analysis, Spiro *et al.*. suggest a number of motivations including structural and motivic features. However, they note that repetitions expected by looking at the score are not necessarily echoed in a performance. Also, timing patterns seem to be more salient when the performer uses a range of expressive timing during the piece. Full phrasing reconstruction is attempted through pattern finding in the tempo and loudness curves [56]. Repeated musical structures are searched for in unsegmented data of audio recordings of Schumann's Traumerei. Correlations between tempo and dynamic values are used as a basis for the pattern finding algorithm. Similar musical structures are identified with some success for this one piece representing a first step in phrase reconstruction.

The final ritardando in performances of the same piece is examined in terms of visualising the implied motion from expressive timing. First and second-order phase-plane representations are used to visualise the changes in timing across three performances. The segmentation of the final ritardando into three motifs is clear from the curves in the plot. There exist many kinematic models of expressive timing in performance [59, 43] which work on the basis of music (and tempo) being closely related to motion. For a complete review on the studies involving keyboard and other instruments, analysis of aural parameters and the study of motor programs and kinematic models see [44, 86]. The role that motion plays in a musical performance is examined next, explaining why this thesis looks at physical motion as a visual performance parameter, equivalent in informing studies on musical structure as aural parameters of timing and dynamics.

2.3 Music and Movement

The production of sound from an acoustic musical instrument requires movement whether it be to pluck a string, press a note on a keyboard or blow air into a wind instrument. Music and movement are therefore completely interrelated in the production and perception of music. Otto Ortmann, whose work explores the physics of piano playing states simply that music *is* movement [85]. Movement is also inherent in the interpretation of music. Performance directions governing tempos are described in terms of walking and running, fast and slow. Movements are not just found in sound-production but can be seen as an expressive visual factor in performance. The purpose of these non-sounding movements has been compared with non-verbal gestures which accompany speech, reinforcing or negating what is being said at the time [83]. Movement in performance can therefore contain more musical meaning than simply motor movements required for the production of sound.

The importance of movement is also reflected in music education. Teaching theories from Dalcroze [13] work on the basis of teaching music through movement (eurhythmics) which may involve games and exercises which encourage children to link rhythmic properties in music with movement. Many musicians also utilise the writings of Frederick Alexander which speak of the mind and the body being an inseparable unit of 'self' and teach performers how to use their full body effectively [37]. Performers are directed to focus on the coordination of the neck and the spine as the base control unit of the body, and to increase their awareness of posture and any self-limiting habits. Optimizing movements is encouraged for the achievement of both control and expression. Of course these teachings are not just limited to the performance of music but also used in areas such as sports performance. Ethnomusicology contains examples of music cultures where movement can also be useful in terms of emphasising structure, particularly in genres of music where notated forms of music are not common. An example of this is African folk music where songs are traditionally passed down aurally. In this case, the role of the body is emphasised, particularly in communication with other performers and particularly highlighting rhythmic properties of the song.

The body can also be used as a way for performers and audiences to 'feel' the music. Embodiment cognition theory, particularly when applied to music [69] considers the full body as having an important role in the experience of music. This falls into line with Alexander's view of the mind and body being inseparably one unit. Embodied music cognition regards both performer and perceiver as subjects as audiences have been seen to respond through movement to the music being performed [48] and highly associate sonorous objects with movement.

After determining various reasons why movement is produced in performance, it is interesting to look at how this is manifested corporeally in musical examples. Movement in classical piano performance appears to be completely personal and there is a range of famous performers who incorporate different physical styles when playing. Arthur Rubinstein is one example of a performer who plays with such visible effortlessness and barely moves from the centre of the piano. Glen Gould on the other hand has been characatured almost as an ogre over the piano, hunched over the keys and moving around with vigour and energy. One question to ask when considering physical gesture and its relationship to music is whether different styles of movement can be attributed to the same musical feature, much like the differences in performers' use of parameters such as tempo and dynamics can convey the same musical feature.

So with movement coming into the foreground of theories to do with how performers play and audiences perceive music, examining motion in performance becomes crucial when researching how performers encode information from the score. The next section looks at current empirical studies involving motion and its relationship to the audible parameters produced in performance.

2.4 Research into Gesture

So although a certain amount of movement is necessary to play a musical instrument, it appears that movement is not solely for this purpose. Before diving into gestural studies, it is necessary to define exactly what is meant by gesture in musical performance, and the different functions gestures may have. From the studies of performance gestures of the pianist Glenn Gould, Delalande proposes a three level structure of gestures ranging from functional to abstract[38]. The first level are effective gestures, which are necessary for playing the instrument i.e. bowing, blowing, pressing keys etc. Accompanist gestures are those movements which are associated to effective gestures i.e. elbow and chest movements which are used to help the performer articulate a particular sound. The final level is figurative gestures which are visually perceived by the audience but seem to have no correlation with the actual production of the sound. Existing gesture taxonomies for music are based on this three-tiered structure [21]. Several classifications on gesture are also listed in [62]. The definition of gesture used in this thesis alludes to physical motions made by the performer that carry meaning. The research in this thesis aims to explore how musical structure factors into these gestures, whether this information is produced visibly in accompanying gestures.

Davidson and colleagues [35] have explored various purposes for physical gesture in performance, mainly the communication between performers, conveying personal issues. It was found that between performers, features such as accents are used to communicate with each other and physical gestures provide the anticipation to these accents. This may be a reason as to why performers watch each other for visual cues. Jane Ginsborg also investigated the use of gestures and movements in the rehearsal of singer-pianist duos [46]. Gestures were used for keeping time, coordinating entries and also highlighting particular expressive points. Familiarity between the duo and similar levels of expertise showed a wider range of gestures being used than in unfamiliar or unbalanced partnerships. From the many different functions and purposes gestures in performance may have, this thesis focusses on those made in solo performance, eliminating the communicative purpose between other performers. I aim to discover gestures in piano performance from full body movements to intricate fingering details which provide some information or link to expressive features of the music.

On studying expression in musical performance Eric Clarke states

> ...body movement associated with the production of expressive musical performances is directly perceivable, can communicate differences of performance intention even in the absence of accompanying sound, and is strongly related to the timing and dynamic profiles of the resulting sound

[27].Clarke and Davidson's study into movement in piano performance [25] identified different types of head movement in their relationship to the aural parame-

ters measured from the recorded MIDI. Although body sway was not clarified as being directly related to phrasing structure, the authors acknowledged that neither is it random. Further exploring and quantifying these relationships between movement and sound in performance through multi-modal recordings, Camurri *et al.* recorded information from repeated performances of a Scriabin etude by a professional pianist to see how movement, tempo and dynamics conveyed emotional information [22]. Movement was measured using the EyesWeb software (explained further in Section 3.3.5) in terms of openness or contractedness of the performer's posture. Ratings of emotional intensity by audience judges were also gathered to assess the communication of this emotional intention. Correlations between pairs of parameters were used to judge which agreed for each bar. The results highlighted inter-onset intervals, key velocities, movement velocity and the openness and contractedness of posture. This study highlighted specific parameters that may contain emotional information on the music, that is relayed effectively to an audience. Jane Davidson's extensive work on body movement in musical performance has also established that information about intent and structure amongst other cues such as communication between performers, can be conveyed from performer to audience [32]. Point-light displays of 'deadpan', 'standard' and 'exaggerated' performances were presented to audience judges who were asked to rate the level of intended expressivity. When varying the level of expressive intent from 'deadpan' through to 'expressive', pianists changed the amplitude of their movements suggesting a link between movement and expression. This was also perceived accurately by audiences judging the expressive intent from videos of the performances. Subjects who were not given the visual information performed poorer than those with the visual stimuli suggesting that the presence of these visual gestures enable communication of information on intent accurately. It was also discovered that performance intentions were more detectable from the upper torso movements than those of the hands [33] implying that audiences use the full body gesture to make their judgements rather than more localised gestures from the hands. Further work used 2D tracking of such movements made during piano performance to quantify the relationship between movement size and expression [34]. Results showed that the more exaggerated the performance intention, the more exaggerated the amplitude of movement. Other studies on the visual communication of intent include [111, 31]. Establishing a link between performance intent and performer gesture, we now look to see if more intricate details of structure can be contained in such movements and aim to quantify more deeply the relationship between visual and aural gestures.

Wanderley states that although we are not entirely certain as to why accompanist gestures are performed, it is evident that they exist frequently in performances [126] and are repeatable at the same points in the score across several performances by the same performer. The three-level topology in this paper is based again on Delalande's theory as stated earlier. Performers played a selection of pieces, including Stravinsky's three pieces for clarinet and the first movement of the Brahms 1st clarinet sonata in standard, expressive and immobilised performances. The Optotrak system was used (see Section 3.3) to collect data from the bell and mouthpiece of the clarinet as well as the performer's head, arms and legs. Further analysis of this movement data was conducted for the opening of a solo clarinet piece by Stravinsky, that lacked certain rhythmic accents that may influence movement such as in the Brahms sonata[125]. Movement data analysis was influenced by recordings taken from a digital video camera and was calculated and registered as a Total Amount of Movement value by using frame by frame subtraction. These results were time-warped to allow comparison across performers. An interesting result from this research was the influence of performer movement on keeping rhythm and timing which led to hypotheses about the role of continual movement in phrasing/musical motion. However, it was also clear that movements became restricted in very fast, technical passages whilst increasing their movements at easier passages. It was noted in particular that performers moved a lot at phrasing boundaries. There were many different performer styles of movement and although there were some similarities, there were significant differences in what parts of the body they used to move. Some would sway their heads whilst others moved their waist and shoulders. From observational analysis they concluded that these movements were related to patterns of tension and release in phrasing. Bell movements were not always related to phrasing but in this case appeared to be more rhythmical. Other performers who hardly moved within a phrase would use large movements to perform a phrase-end gesture. Other correlations between movement and the musical properties of the score were investigated in Rodger [105] where performers were recorded through motion capture and audio from different stages in learning a piece of music. Principal components analysis was used to analyse body motion and this movement was correlated with both melodic contour and dynamics. Results found that the further through the learning process of the piece, performers' correlations between movement and melody increased. This suggests that as the performers develop the interpretation of the music through practice, movement becomes a more integral part of the performance, becoming more highly related to what is being played. These studies of a few performers suggested some generalities about how clarinettists use gesture. Different patterns of movement were found across the performers although at phrase endings, most would perform some sort of phrase-end gesture. Movements also appear to be correlated to the melodic contour of the piece being performed. However, as these are instrumentalists who have the freedom to move not only themselves but their instrument, it is interesting to analyse the movements of the actual clarinet further. Bell movement in clarinet performance is further explored regarding its relationship with rubato [89]. Intensity values taken from the audio were found to be correlated with the melodic contour of the piece (in this case it was the Adagio from Mozart's A major clarinet concerto) and not with bell height as might be expected. The bell movement however, was related to sound properties and appeared to be related to phrasing.

From these extensive studies on clarinet performance it is inferred that movement of both the full body and the movement of the clarinet is related to what is being performed in cases of melody and phrasing. Movement in clarinet performances has also been used to study the effect these gestures have on the perception of certain aspects of the music. It has also been shown that visual information appears to aid perception of musical information from performances. An example of this is seen in the work of Vines *et al.* [124]. This study used one of the clarinet performances recorded for Wanderley's experiments, showing them to thirty musically trained audience participants. Audience judges in a between subjects design were shown different modes of presentation (audio-only, visual-only and audio-visual) and asked to make real-time judgements on the phrasing structure and emotional intensity of performances. Functional data analysis techniques were used to examine the underlying factors changing over time. The combination of aural and visual information appeared to be the most accurate when determining phrasing and intensity. Further analysis on bodily gestures concluded that motion sequences were approximately slightly longer than the duration of the musical phrase being performed. They proposed that the contour of this movement over time might also correlate with the phrasing contour. Further work into the cross-modal interaction in perception used two different performances and repeated the experiment [123]. Again they saw gestures extending the sense of phrasing for participants. Also, the visual modality proved to be full of information relating to the phrasing structure, as much as the auditory stream. They also found anticipatory movements at the beginning of phrases which cued the beginning of a phrase for the perceivers earlier than in the purely audio presentations. This effect is also true for co-articulation gestures in speech which precede sounds [83]. Extending the analysis even further in [84], performers were asked to play an excerpt from the Brahms sonata without piano accompaniment and were given no performance directions this time. Motion capture was performed through the Vicon motion capture system. The kinematic displays produced from the Vicon captures were modified by 'freezing' certain parts of the body, changing the movement amplitudes or showing the movement in reverse order. These performances were used to analyse how ancillary gestures affected perceiver's views on tension, intensity, fluency and professionalism. Results proposed that freezing the motions of body parts did not affect the perception of these musical values and so it could be suggested that general body movement communicates more information efficiently.

These multiple studies in clarinet performance concerning gesture production and perception identify areas where this research could benefit from in terms of analysing motion in pianists for musical structure. High differences in movement between performers are evident, however, there are similarities in the points within the performance that these movements occur. We return to studies on piano performance but now particularly with an emphasis on relating movement to sound and structure.

Thompson and Luck's recordings of movement in piano performance noted that having subjects repeat performances in multiple recording sessions had little real-world effect on the amount of movement used in the performance. They look at the head and shoulders as ancillary gestures as they are more removed from physically producing the sound, whereas data from the fingers and wrists are more involved in sound production. When asking performers to vary their levels of expression from 'deadpan' to 'expressive' they noticed a change in amplitude of movement much like the previous clarinet studies and Davidson's piano studies. On further examining the link between movement and audio, it was discovered that movements sometimes predicted features of the audio stream [117, 118]. Work by Shoda [110] looked into this temporal relationship between body movement and temporal expression, finding that in fast tempi pieces movement appeared to be in synchronisation with expressive timing, whereas slow tempi pieces experienced lags between the movement and audible expression.

Delving deeper into more intricate movements in both clarinet and piano performance, finger motion is examined in reference to its relationship to the acoustical outcome of the sound. Palmer and Dalla Bella's investigation into the motor movements of pianists' fingers as they played fast passages saw a surprising result in higher amplitude movements for faster repetitions of the musical excerpt [88]. These motor movements were then analysed for the effect they have on the way the musical passage is performed [87]. Clarinettists were used in this particular study as they do not use finger height to change the loudness of the sound produced. This eliminates the possibility of increased tempo passages in piano performance requiring a louder dynamic and therefore possibly higher amplitude finger movements. Again a relationship between faster passages of music and higher finger height was seen despite this not having an affect on the loudness of the sound produced. They propose that these movements are governed by biomechanical constraints in finger movement as well as musical considerations. However, more studies across different instruments are suggested as ways to separate which movements are a response to music instead of biomechanics. Other studies have investigated the use of tactile information at the fingertips to enable performers to control the accuracy of their timing [51, 53]. Differences in pianists' touch at different tempi were evaluated by extracting landmarks in motion such as key-bottom landmark and the maximum finger height preceding performed notes [52]. Results showed that a different 'touch' was used at faster tempi than at slower tempi. The musical extracts used in these studies were designed specifically for 'fast' or 'moderate' performances and in order to manipulate certain fingering combinations. Although these results show differences in pianists' touch, for the research executed in this thesis, I aim to look at fast passages of music where particular notes may be accented for structural reasons. Differences of pianists' touch within a certain passage such as this may provide clues as to how the music is being interpreted.

2.5 Summary

From the various theories and empirical studies examined in this chapter, it is suggested that performers are free, within certain constraints, to use expressive parameters such as tempo, dynamics, timbre, articulation and motion to emphasise structural and emotional aspects of the music. Timing and dynamics profiles across groups of pianists are remarkably similar, however, different strategies for these parameters can be used to express the same structural features. Gestures within performance appear to be largely idiosyncratic although some similarities are evident. Finger motion also seems to have a relationship to certain properties of the music being performed although this needs to be more widely investigated across instrumentation.

The results of these studies pose a number of questions which the experiments in this thesis will aim to answer:

1. Despite the idiosyncratic nature of motion profiles across pianists, are there commonalities which exist in occurrence with features in the phrasing structure?

2. Are these motion profiles repeatable across different pieces?

3. Are there commonalities, therefore, between the tempo, dynamics and motion patterns of performers which suggest a link to phrasing?

Implications for the following research include evaluation methods which do not depend on one interpretation of the music or one particular profile concerning timing, dynamics or motion. It would be beneficial to investigate the commonalities behind each of these interpretations and parameter profiles for several pianists, and whether these are consistent across performances of different pieces, particularly as many studies focus on several performances of just one piece, or one performance of a few pieces.

To accurately capture each of these parameters for several performers, systems need to be in place for multi-modal recordings. The next chapter outlines the different techniques available for recordings such as this.

Chapter 3

Review of Capture and Analysis Techniques

This chapter will detail some of the available capture technologies and analysis techniques for audio, MIDI and video with reference to piano performance. This is by no means a comprehensive list of all the available technologies but means to serve as an example of the range of products and applications that exist, identifying the advantages and disadvantages of each. A further review of data acquisition techniques in music performance can be seen in Goebl *et al.*[50].

3.1 Audio

Performance analysis up until recently, has been mainly concerned with the analysis of audio recordings from famous pianists, due in part to the wide availability of data. Measuring parameters such as dynamics and expressive timing can be beneficial in this way, but when comparing two performances together, the differences in how they were recorded become a factor, particularly for the intensity of the sound wave.

In the experiments detailed in Part III, audio is recorded through a stereo microphone setup, connected to a laptop computer via a Tascam Audio Interface. This data is transported to the application Ardour via the Jack Audio Client which has a low I/O latency of around 46.4ms in this particular case.

Once the audio has been recorded, there exist many tools for audio analysis. Examples of these are libxtract [18], aubio [16] and other audio feature extraction libraries that attempt to estimate note onsets, tempo and other lower level features such as Mel Frequency Capstral Coefficients and spectral densities etc. The disadvantage with using only audio recordings is particularly in the accuracy of these note onset and tempo estimations. For instruments such as the piano, onsets within a chord cannot be separated. This is something that can be overcome by recording MIDI information. However, the audio data provides information that cannot be recorded from simply MIDI information alone such as the effects of pedalling, the exact duration of notes (also influenced by pedalling and the acoustic effects of the performance space) and also for instruments in which onset information cannot be measured any other way.

3.2 MIDI

Information transported through the MIDI protocol can be collected in various ways, particularly when concerning keyboard instruments. There are a number of devices which can be used as external retrofits including the Moog Piano Bar [5], which has a recommended retail price of approximately $1495 [1]. This device uses infra-red beams to detect depression of the keys. Internal retrofits such as the TFT Midi Record system place a strip of carbon coated plastic underneath the keys to record the onsets and velocities by changes in resistance and also use a sensor detecting the onset and offset of the sustain pedal. Retail prices for an internal retrofit such as this start from 1130 Euros [2] however, extra cost must be accounted for the installation of the device. There are also factory installed pianos from Yamaha and Bosendorfer that include the optical sensors for the keys and pedals. Retail prices for the Disklavier range from the basic system at £25,000 to the more advanced system at £35,000 [3].

The factory installed series has limits in its price and portability, issues expected to be solved by the internal retrofit optical devices. These however still require modification to the actual piano which involves specialised installation, reducing the portability somewhat, whereas the external retrofit devices are the most portable and sit slightly above the keys of any piano. However, internal retrofits would combat issues arising from the space the external device takes up at the back of the keyboard. Interviews from the professional pianists in the experiments in Chapter 9 highlighted opinions that the factory installed pianos had a different 'feel'. It is possible that regardless of whether the response of the piano

[1]RRP taken from `http://www.moogmusic.com/newsarch.phpcat_id=24` on 05/04/10

[2]RRP taken from a local distributor in London on 23/06/10

[3]RRP taken from a local distributor in Glasgow on 05/04/10

was changed by these optical devices, it may be a psychological issue for performers and that seeing a device such as the MOOG piano bar sitting above a normal piano they had used in previous concerts helped make them more comfortable.

Table 3.1 shows a direct comparison between these three types of device. The most portable of the three kinds of devices, which is the MOOG piano bar, will be used for the experiments in Section III.

3.2.1 MOOG piano bar

This device 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 infra-red 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 infra-red beam being broken and for a black key by an infra-red 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 [8]. The MIDI data will provide information on what key is pressed, its onset time, its offset time, key velocity and also which pedal is depressed and its onset time.

The piano bar needs to be calibrated by playing each note on the piano, making sure the hammer does not bounce back. The red and green lights on the scanner bar indicate whether the height of the bar needs to be adjusted upwards or downwards. The accuracy of the piano bar for recording purposes is good but disadvantages are that there is no pedal recognition apart from when the pedal is fully depressed (a huge issue for pianists as there are several degrees of 'on' for the sustain pedal). There is also no release velocity measured. Also as this is an infra-red system it cannot be used in conjunction with motion capture systems that use infra-red, see Section 3.3. For more detailed measurements such as key angles and pedal depression angles, the Bosendorfer factory installed series is ideal, if rather expensive and unable to be moved easily.

However, in terms of low cost, high accuracy in terms of timing and onset velocity, being portable and the least amount of disturbance to a performer, the

Table 3.1: Comparison of MIDI Capturing Devices

Device	Type	Measurements	Price	Advantages/Disadvantages
MOOG piano bar	External retrofit	Key onset, offset, onset velocity, pedal depression	$1495	Adv: very portable and cheap, Disadv: no release velocity or pedal angle, bar sits above keys so reduces playing space by 1cm
TFT MIDI record strip	Internal retrofit	Key onset, offset, onset velocity, pedal depression	1130 euros	Adv: cheap, could be portable, and sits underneath the keys, Disadv: no release velocity or pedal angle and requires specialised setup
Yamaha Disklavier	Factory installed	Key onset, offset, onset velocity, release velocity, pedal angle	£25000-35000	Adv: All measurements required, technology underneath the keys, Disadv: Expensive and stationary

Moog Piano Bar is the best device for use in the multi-modal capture experiments.

3.3 Motion Capture Methods

When choosing a motion capture method there are a number of options [127]. Active markers can consist of accelerometers and gyroscopes or magnetic, inertial, optical, acoustic, radio frequency or ultrasound sensors. Passive marker systems can involve the use of infra-red camera systems with retro-reflective markers or image processing systems with coloured markers. Image processing can help to eliminate the use of markers altogether. Continued research into human movement and gait has produced a number of motion tracking algorithms for full body motion. A number of solutions exist particularly for capturing hand gestures designed specifically for applications in sign language recognition or finger detection for multi-touch surfaces. The best choice for accurate tracking is heavily context-dependent. For example, musicians will not be entirely comfortable with heavy electronics balancing on their wrists and so our choice for pianists must be based primarily on the need to be as un-intrusive as possible to the performance.

This section describes a few systems and devices available including active markers like accelerometers, passive markers such as retro-reflective dots and other image processing systems focussing on their advantages and disadvantages in the context of measuring movement in musical performance.

3.3.1 Accelerometers

Active markers such as accelerometers can be used to determine acceleration patterns in body movement. Available on chips with additional gyroscopes, positional information can be calculated by integrating the measured acceleration vectors. An example of such a device is the IMU 6 Degrees of Freedom v2. This device consists of three iMEMS gyroscopes with a Freescale three axes accelerometer and costs approximately $124.95 [4]. Prices increase with rises in bandwidth and sensitivity. Accelerometers can now be bought with wireless capability, but as portable as these small devices are, there are still limitations in placing them on pianists' fingers without causing interference. Therefore, these devices are more suited to measuring general body movement. Other disadvantages include errors that can

[4] RRP from Sparfun Electronics http://www.sparkfun.com/commerce/product_info.phpproducts_id=9184 on 23/06/10

arise from bias and sensitivity drifts with temperature. If using accelerometers to integrate for positional information, this can give rise to errors with a magnitude to the power of 2.

3.3.2 Optotrak System

A real-time capture system such as the Optotrak Certus Motion Capture System [12] has its advantages for being able to interact with the performer. Being highly accurate with a resolution of 0.01mm, error of 0.1mm and a marker frequency of 4600Hz, the system also benefits on not relying on retro-reflective markers and so is not subjected to noise due to reflections of light. This achieves a higher level of portability, helped by the setup of the capture cameras on moveable stands. However, as the markers are connected by wires, this system may not be suitable for every performance situation. The development of the Optotrak Smart Markers which can be connected to a portable pack instead of the capture system allow slightly more freedom of movement for the capture subject, however, the wires may still cause discomfort or disturbance to the performer and would be particularly unsuited for tracking individual finger motion. Prices start for a basic Optotrak system at £70000 [5] capturing a 6.5 metre volume.

3.3.3 Vicon System

The Vicon range of motion capture systems combine infra-red camera arrays, control modules and specialised proprietary software for capture and analysis. Vicon cameras allow fast frame rate capture of 3D motion by detecting retro-reflective markers in a capture volume through several cameras, and triangulating their position. Vicon is an example of many existing infra-red tracking systems and has been designed primarily to capture human gait. This is one in a range of such systems offering full 3D motion capture by using infra-red technology.

The particular Vicon system used in the experiments in Section III employs a 12 camera array of the F-series MX F40 cameras with 4 MegaPixel resolution recording at 120 frames per second. This system costs approximately £110,000 for the cameras, master computer, markers, velcro jacket, software and maintenance [6]. There is a 0.5/0.6 pixel error in the cameras and a resolution of 3mm. Small changes in temperature and lighting as well as floor vibrations from walking re-

[5] RRP taken from UK distributor on 24/06/10

[6] Estimated price for system at Glasgow University

quire the cameras to be calibrated for noise before each use. The Vicon system allows creation of body models which specify how each marker connects to the others and once the Vicon VST model is created it can be used several times. For recording each subject, a stationary capture must be taken for the system to recognise the programmed model.

Post-recording reconstruction involves frame by frame viewing although there is functionality for filling in gaps later. However, as it is proprietary software, these estimation algorithms are unavailable to view and so cannot be evaluated. As it was impossible to extract the data from the Vicon program any other way apart from through the proprietary format, it was necessary export to ASCII files and then process text files with huge lists of numbers. Other software packages exist to allow full analysis of recordings made including biomechanical calculations, however, these come at an additional price.

The highly accurate measurements of several markers for the human body makes this system highly desirable for use in performance analysis experiments, despite its limitations in price and portability. The experiment in Chapter 8 analysing upper body movement in piano performance utilises this particular capture system.

3.3.4 Image Processing

Several image processing algorithms have been designed for tracking hand and body gestures for a variety of purposes. Research into tracking hand gestures for sign language detection has produced algorithms using model-based detection [72, 73] and crevice detection [55]. Visual detection using robotics theory captures the curvature of each finger but sacrifices the detection of the horizontal positional information of the hand [81]. Fingertip detection for guitarist's fingering can track the position of the hand over the fret but has problems with finger-finger occlusion [20]. Although these algorithms help to detect certain hand postures from 2D images, they fail to track accurately the position of each point of each finger. A particular interface for using such image tracking algorithms is discussed next.

3.3.5 Eyesweb

Eyesweb [23] is a graphical user interface image processing system which allows users to create their own analysis of captured video images using various algorithms found in the OpenCv image processing library [11]. Designed for full body motion in music and dance (particularly dance) it contains a number of analysis techniques such as motion history images which provide a visualisation of motion in time in a single snapshot [36]. EyesWeb XMI also provides the functionality to convert between several layers and data types. Users can select various functions as blocks and connect the input and output to other functions, as well as being able to write their own processing blocks. It is a free and open source software (FOSS) application and requires only video images input from video or live camera. As well as functions for overall body motion, algorithms for finger tracking have been assessed [19] from Hough transforms to tracking with coloured markers. The accuracy and ability of coloured markers to work with complex backgrounds, such as a piano keyboard with changing light conditions, far outweighs the benefits of the other algorithms assessed. The application however, only runs on the Windows operating system so far.

3.3.6 A portable, low cost, accurate motion capture system for pianists' fingers

Although many of these motion capture methods detailed in Table 3.2 offer advantages of portability and high accuracy, few allow this in combination with being low cost and being designed considering the distraction caused to the performer. A solution to this lies in image processing systems. The image processing techniques explored above offer solutions in tracking general hand shapes in performance although will not allow the intricate measurement of each joint of each finger. A specially designed image processing system with passive coloured markers for each joint of each finger is described in Chapter 4.

3.4 Data Storage

The trial of recording multiple streams of data arises when attempting to computationally store the data in a way that makes sense in combination with the musical score information.

3.4.1 Storing Musical Data

Storing musical data, particularly for performance analysis purposes, has to consider future extraction for purposes such as comparing a number of different performances against certain positions in the score. Amongst the many data for-

Table 3.2: Comparison of Motion Capture Devices

System	Active/Passive	Wireless	Error	Sampling Freq	Resolution	Price	Advantages/Disadvantages
Accelerometer and Gyroscope	Active	Not fully	Sensitivity Drift $\pm 0.03\%$ /°C, Output signal at zero ± 2mg/°C and gyro noise $0.05°/s/\sqrt{Hz}$	11kHz internal	± 3g	$124.95	Adv: direct application to body, Disadv: drift and bias error
Optotrak	Active	No	0.1mm	4600Hz	0.01mm	$70000	Adv: portable, never loses position of markers, Disadv: wired markers
Vicon	Passive	Yes	0.5/0.6 pixel	120Hz	0.3mm	$110000	Adv: 3D motion capture, wireless markers, Disadv: stationary, can experience noise and occlusion
Eyesweb	Passive	Yes	Depends On Camera	Depends On Camera	Depends On Camera	Free Open Source Software	Adv: used with any input camera, Disadv: can experience noise and occlusion, graphical user interface not good for overcomplicated programs

mats for representing score information, MusicXML [54], an XML based tool, has proved the most popular. Storing precise performance data such as timing alongside recorded audio is straight forward enough through the use of a simple audio editor such as Audacity[2], which supports tagging audio files and can read and write these into text files. Storing performance data in alignment with score information ,however, requires a fully integrated infrastructure that can support a more sophisticated level of data processing. Amongst existing solutions are the Music Encoding Initiative (MEI)[93] and Performance Mark-up Language (PML) [7]. MEI's main aim is to "a) provide a standardised universal XML encoding format for music content (and its accompanying metadata) and b) facilitate interchange of the encoded data". MEI represents score as well as analytical data, and also has the ability to time-stamp objects in various time codes. However, these time-stamp objects' associated semantics are fairly trivial, and the performance data is not given an explicit, separate representation.

A solution to this lies in the development of Performance Markup Language (PML) which stores the performance data in a separate hierarchy to the musical score data, linked to each other by note IDs (see Chapter 6).

3.4.2 Storing Gestural Information

Gesture Motion Signal (GMS) files [75] are one of the few open source representations for gesture available. Many of the other gesture capture files are associated with proprietary software, and consist of a set of two files, one for the marker data and one for containing the skeleton model. These are generally un-usable outwith the said proprietary software.

GMS is a binary format for storing and streaming low-level movement data for a variety of applications. These files were chosen as an appropriate format to store the coordinates of the tracked positions of markers in the motion capture system described later in Chapter 4 whilst preserving the hand object structure for each frame. Each frame consists of a gesture "scene" which in turn can consist of numerous units, channels and tracks stored at any user required frequency. Each channel allows storage of performed gestures such as position or force in either 1, 2 or 3 dimensions in the subsequent 'tracks'. The structure of the GMS file is shown in Figure 3.1.

Despite being able to store these motion values in a meaningful manner, it does not allow storage of analysed gesture in terms of phrasing etc. The gestural data is also stored in a completely separate format to any other musical information. PML

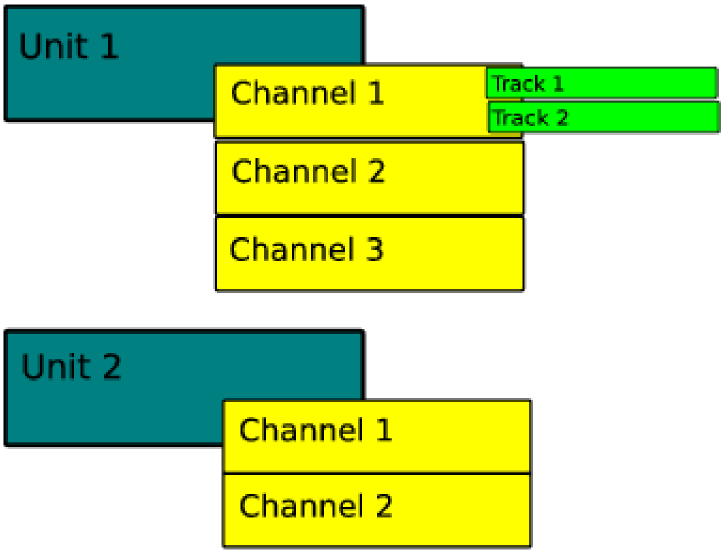


Figure 3.1: Example Structure of Units and Channels in a GMS file

is currently under development to allow gestural data to be stored in alignment with score information (see Section 6.1.1).

GDIF (Gesture Description Interchange Format) [64] is a another gesture format based on SDIF (Sound Description Interchange Format) originally designed to describe properties of audio signals. GDIF is being developed at the University of Oslo based on already existing formats such as XML and OSC (Open Sound Control - a message based format for communicating between software and hardware systems). This is developed on the need to be able to store performer-instrument qualities and other mid and high level gesture descriptors alongside lower level descriptors such as velocity and position. It is intended for both storage and streaming but is not available for use at present. A review of these existing formats can be found in [63].

3.5 Visualising Data

Several tools exist for visualising mid-level gestural information in both audio and video such as tempo, dynamics and bodily gestures as well as low-level descriptors such as velocities, accelerations and spectral densities. This section provides a few examples of these and highlights the need for effective visualisation of this information alongside the score.

3.5.1 Performance Worm

The Performance Worm created by Langner and Goebl [66] and later used by Dixon [40], plots a 2D graph of dynamics versus tempo in the form of an animation for each performance. Using an audio signal as input, the dynamics are measured by taking the sound pressure level and the pulse is extracted using the beat-tracking system Beat-root [39]. The musical timing of the notes relative to their expected time and duration can then be calculated. A circle is plotted for each point in time (depending on the frequency of occurrence of notes within the excerpt) with the colour fading as time progresses, plotting a path of these circles to give the user an idea of how the tempo/dynamics change over a period of time. In the most recent version of the application, the bar number of the music being played is displayed within the most recently plotted circle and major boundaries such as the end of an excerpt are identified by large black circles within the plotted path. This is an extremely useful tool for comparing patterns of performers' use of tempo and dynamics within a piece and users can see distinct styles of performance producing different paths. Unfortunately, there is no direct visualisation of the music they are playing or a continuous feeling of time, except that the picture of the worm moves about the screen in synchrony with the audio output. The resultant graphs of dynamics versus tempo allow easy comparison of two different performers playing the same piece and so it is a good visualisation and analysis tool for comparing the performance styles of famous artists. However, it does not provide useful information about the particular performance itself in terms of the musical score.

3.5.2 Sonic Visualiser

Sonic Visualiser is an application allowing the user to analyse and view audio files, developed at Queen Mary University in London [24]. Along with a series of VAMP Plugins based on audio analysis libraries (see section 3.1), users can look at tempo estimations, note onsets and other low-level audio measures such as Mel Frequency Capstral Coefficients. This application allows direct visualisation of spectrum graphs/line graphs for tempo estimations against the audio wave and even comparisons of different performances. There also exists a MATCH algorithm [41] providing the ability to directly compare different audio performances of the same piece at the same point in the music. The program also allows the user to export the results of these analyses (known as annotation layers) into text files so they can be used as raw data. This application is beneficial to audio analysts particularly in the Music Information Retrieval community. However, there is no direct view of the actual notes.

3.5.3 Summarising Video

A tool for summarising video images is referred to as key-frame displays [62]. Based on the theory of gestures having certain key-frames that are structurally important and inter-frames in between, the programs allow the users to visualise certain frames along the progression of time. The video images are sampled at an interval of 2 seconds. The development of salient key-frame displays removes the information least likely to be perceptually salient, thus reducing the amount of images needing to be displayed. However, it is still a very static representation of continuous movement sampled at 2Hz and so does not provide clues as to what movement led to the particular position of the key-frame.

3.5.4 Motion History Key-frame Displays

To investigate motion along the path of time, Motion History Images [36] are developed by using a running time window to record the trajectories of movements between images. These have been combined with the idea of key-frame displays to produce motion history key-frame displays [62]. These show the trajectories of movement leading up to each key-frame, however, this visualisation is still limited when attempting to represent long movement sequences.

3.5.5 Motiongrams

Motion data is always too much to plot in one two dimensional graph, and so in an effort to visualise overall motion, and particularly longer sequences of motion, Jensenius has created a number of tools that can be used much like spectrograms are used to look at audio files [61]. Motiongrams analyse the differences between frames and take the mean of the rows and columns, displaying the results on a continuous graph. This can be visualised in synchronisation with the spectrum of the recorded audio. This particular tool allows the user to identify particular points of interest in the audio and video spectrum for further analysis.

3.5.6 Visualisation with the score

Although these visualisation methods detailed in Table 3.3 all help to give an instant impression of the audio or video performances such that they are distinguishable between performers, they all lack a direct relation to the score or a representation of the notes being played. A representation involving both the score notes and the performance data would be of great use to performance analysts. The specially developed Pullinger Database (see Section 6.2) presents a method for displaying performance metadata of any kind above the notes on a score, allowing direct analysis and obvious relationships to be determined between the performance measurements and the notes being performed.

Table 3.3: Comparison of Visualisation Tools

Application	Input	Measurements	Advantages
Performance Worm	Audio	Tempo and Dynamics	Good for quickly comparing performance styles
Sonic Visualiser	Audio	Low-level audio descriptors and tempo estimations	Good for comparing details of differences in two performances of same piece
Key-Frame Displays	Video	Displays images of video	Good for overview of video
Motion History Key-Frame Displays	Video	Motion trajectories in time	Good for analysis of motion and overview of video
Motiongrams	Audio and Video	Audio spectrogram and Video motiongram	Good for comparison of motion details compared to details in audio stream

Part II

Developing Multi-Modal Capture Technologies and Tools

Chapter 4

FingerDance

From the review of available motion capture technologies in Chapter 3 a need has been identified for a system specifically designed for tracking finger movements in musical performance. This system should be cheap and portable as well as being as un-intrusive as possible to the performance. FingerDance is a specially designed, open source, image-processing-based motion capture system for tracking pianists' fingers. It is designed for use with a single, fast frame rate camera, placed with an aerial view of the keyboard of the piano. This camera captures images containing passive paint markers applied directly to the performer's fingers. This chapter explains the setup of the FingerDance system and the algorithms behind the identification and tracking of the hands.

4.1 System Setup

It is necessary to use a fast frame rate camera when recording the performer's hands in a piano performance, as a skilled pianist can play up to 30 sequential notes per second [107]. The AVT Guppy F-046 FireWire camera has a Region of Interest facility, allowing a smaller size of frame to be transmitted at a higher frame rate, and so under the current settings, can reach up to 60 frames per second. More expensive cameras are capable of reaching higher frame rates which would produce a better accuracy when measuring movement, if expense is flexible within the project. The camera is placed with an aerial view, 83cm above the keyboard, producing a 780x216 pixel frame and 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 pieces of music being examined in the experiments detailed in Chapter 9 and increasing the coverage would decrease the picture resolution.

The width of a pixel at this height is 1.1mm and so the error in detection is approximately 0.55mm. The calculated angular resolution of the camera is 0.076°, as seen in Figure 4.1. The black box represents the camera and the light squares represent the pixels at 1.1mm width. The angle is calculated by the simple equation

$$\tan^{-1}\left(\frac{1.1}{830}\right) = 0.076^{\circ} \tag{4.1}$$

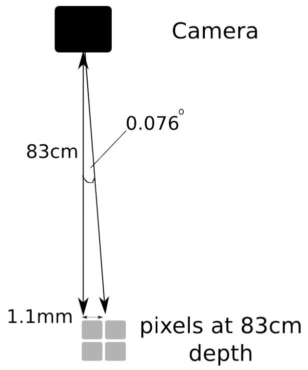


Figure 4.1: Monocular setup showing calculated angular resolution of camera

A stereo setup of these same cameras would allow for depth detection in the image, however, a change in one step of the angular resolution in each camera at the mid-point between the stereo pair would result in a change in depth of as much as approximately 9.2mm, as seen in Figure 4.2. The darker square represents a pixel closer in depth to the camera than the lighter square. This calculated error does not account for extra error that would occur if the cameras do not have an external sync.

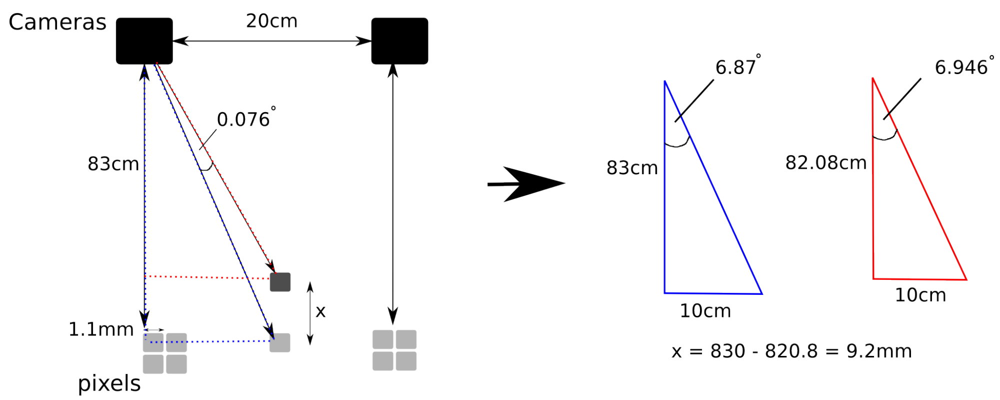


Figure 4.2: Stereo camera setup showing difference in height for one step of angular resolution

As the error in depth calculation is so high in a stereo setup, particularly when considering the small differences in height between each joint of the finger, a monocular system is preferred. Using a monocular setup is cheaper and uses less processing power to capture the raw images, hence making the system more portable. Depth can be estimated from the 2D image reference markers as seen in Section 4.5.

The raw images from the camera are captured through the open-source application Coriander [10], which allows manipulation of the image parameters including frame size, gain and packet size, and stores these images appended as a raw video file. The raw video files are encoded using mencoder [4] and dumped into an avi container with the video coded as lossless jpeg frames. This format is chosen so that the videos obtained are compatible with the image processing library of functions used to program the detection software, the Intel OpenCV library [11]. There is capability for the system to be real-time, as the OpenCV functions can also grab images live from a connected camera. However, to avoid stressing the laptop with high processing requirements during capture and to ensure the system is as portable as possible, all image processing is done post-recording.

Once the markers have been tracked, the output data is stored as a GMS (Gesture Motion Signal) file. The structure of these storage files were explained in the Section 3.4.2.

4.2 Marker Detection

UV reflective paint markers are painted on each joint of each finger, as seen in Figure 4.3, and a black-light sits level with the camera in order to make the markers fluoresce. Each hand's 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. UV paint has been used as the pixels' RGB values will peak at a certain colour, making it easier to subtract the background image from each frame. A different colour of paint is used for each hand, yellow for the left hand and cyan for the right hand, for ease of tracking in cases where either the pianists' hands cross over one another, or even the thumbs of each hand cross over.

The image processing software is written in C++ using the Intel OpenCV Im-

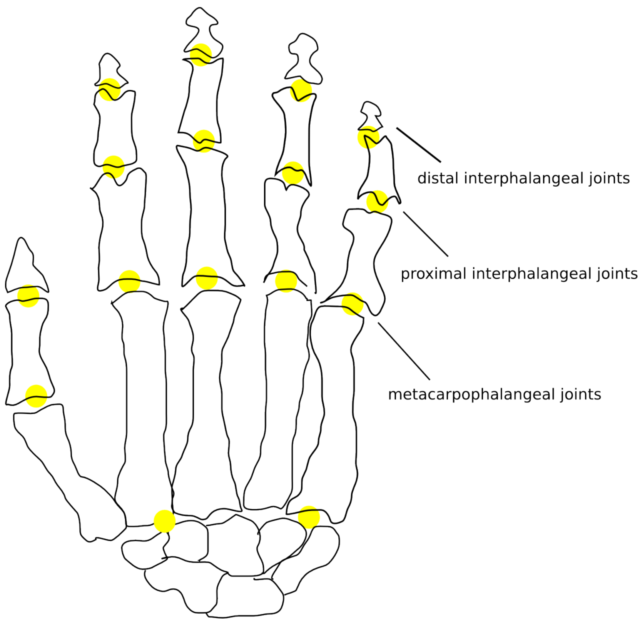


Figure 4.3: Placement of Hand Markers, Plotted as Yellow Dots

age Processing library and the bolt-on OpenCV blob extraction library [11]. The software reads in the avi video files and processes them frame by frame. The first frame requires the user to click on the markers in order of the structure of the hand model to allow a reference frame to be stored before tracking commences. Each frame is passed through colour threshold filters, yellow for the left hand and cyan for the right hand. These two sets of binary images are then submitted to the blob detection algorithm. This algorithm scans each raster image frame line by line and records connecting regions of similar colour. This process can be seen starting from the captured image in Figure 4.4, which is passed through colour thresholding for the left hand markers, which are yellow. This thresholded image is seen in Figure 4.5. This binary image is then submitted for blob detection, the results of which are presented in Figure 4.6. The blob detection algorithm searches for blobs of a certain area to minimise noise. This process is repeated for the right hand markers. Each detected set of blobs are stored in a C++ vector to be compared with the coordinates of the detected markers from the previous frame. A simple correlation algorithm determines which detected blobs are likely to be the new position of each of the hand markers. The thresholding and blob detection functions on an average frame tend to split the average sized 67 pixel marker into two or three distinct blobs and calculates the centre of each. It is this centre which is recorded as the blob's location in the frame. An extra function is included which calculates the distance between each registered blob, combining blobs which are less than 10 pixels distance away from each other's centre. This is in effort to cancel out the effect of the previous functions which split the blob into several other blobs. The error introduced by these image processing functions of thresholding and blob detection in an average frame in calculating the centre of each blob is one pixel in both the x and y direction i.e. 1.1mm in each direction. As this function to calculate the centre of the blob discretizes to approximately 1 pixel, the worst case error is calculated by simply adding the blob and camera errors together. This gives a total error of 1.65mm.

Figure 4.4: Raw Captured Image

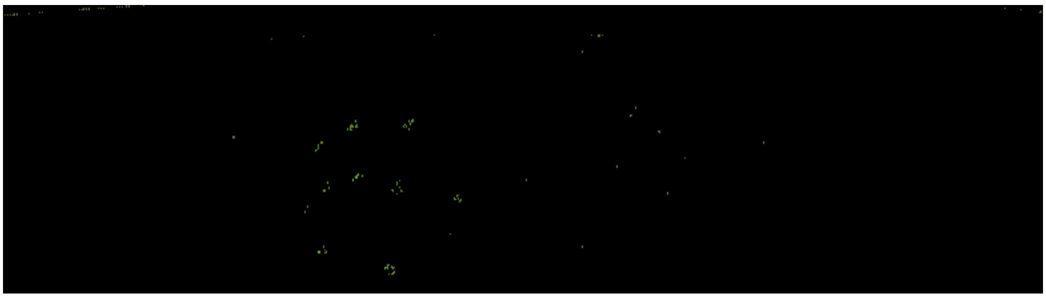


Figure 4.5: Thresholded Image

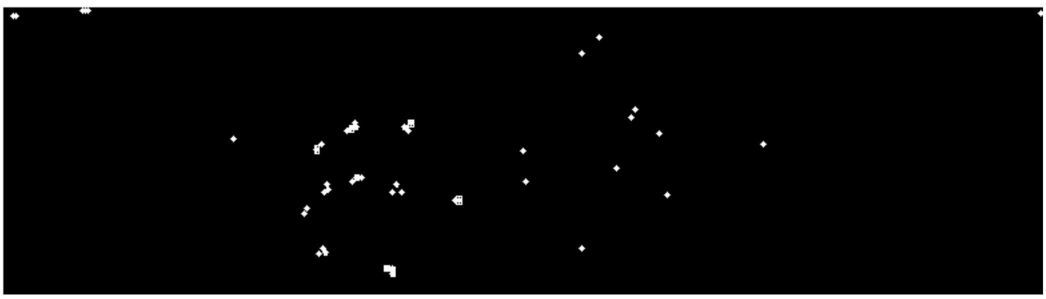


Figure 4.6: Blob Detection Results

Even at frame rates above 50 frames per second, pianists' finger movements are rapid enough to require further remedial action over and above the basic blob

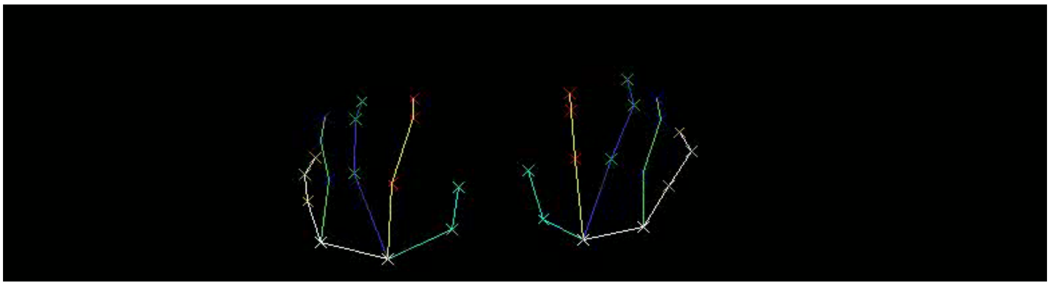


Figure 4.7: Tracked Markers Results

tracking described above. Algorithmic improvements include the incorporation of a skeletal model of the hand as a set of heuristics. This also renders the need for a user-defined reference frame obsolete. The tracked results from the final system processing the original captured image (Figure 4.4) can be viewed in Figure 4.7. The benefits of adding this physical model to the system are assessed in the next section.

4.3 Heuristics

These heuristics are programmed from a list of constraints, advised by Rijpkema's model of human hand constraints [103] with some additions to account for the extra constraints in the context of piano performance.

Basic constraints that are incorporated into the program include the position in x and y coordinates of each finger on each hand, where x is the distance along the width of the keyboard and y is the distance from the top of the frame. Calculating distances between the base wrist points and each of the other markers can also be used to group the markers each for the metacarpophalangeal joints and the two sets of interphalangeal joints. Two examples of basic constraints are therefore as follows:

1. The distances between the metacarpophalangeal points and the wrists are unlikely to be smaller than the distances between the proximal interphalangeal joints and the wrists. These are again unlikely to be smaller than the distances between the distal and the proximal interphalangeal joints. Using this simple rule, the points can be easily separated into groups of joints. This rule is set out in pseudo code in Algorithm 1, where i is the distance between each detected marker and the nearest wrist marker.

2. In piano performance it is unlikely that the x coordinate of the metacar-

pophalangeal points of the left hand's first finger will be larger than the second finger and so on for the third and fourth fingers. The opposite can be considered true for the right hand. This rule for the left hand is set out in pseudo code in Algorithm 2. For the group of detected markers, the x coordinates are evaluated to order the group in increasing value. The marker with the highest value of x is removed from the original vector and put into another vector, $orderedgroup$. This is performed for the next highest value of x and so on until all the markers have been put into the $orderedgroup$ vector. The ordered group is then assigned to first, second, third and fourth fingers respectively.

Algorithm 1 Pseudo code for Heuristic 1

```
if i ≤ maxDistance_meta then
  metagroup ← i
else if i ≤ maxDistance_prox then
  proxgroup ← i
else if i ≤ maxDistance_distal then
  distalgroup ← i
end if
```

Algorithm 2 Pseudo code for Heuristic 2

```
maxval = 0
iterator = 0
for j = 0 to vectorsize do
  for k = 0 to vectorsize_altered do
    if marker k_xval ≥ maxval then
      maxval ← marker k_xval
      iterator ← k
    end if
  end for
  orderedgroup ← marker at iterator k
  remove marker at iterator k from vector
  vectorsize_altered ← vectorsize_altered − 1
end for
```

More advanced constraints can be considered by calculating the angles between joints. An example of an advanced heuristic is as follows:

3. The angle between the proximal interphalangeal and the metacarpophalangeal joints will unlikely be highly different to the angle between the wrist point and the metacarpophalangeal joint. The same rule is applied to the angle between the distal interphalangeal and proximal interphalangeal joint. This algorithm is set out in pseudo code in Algorithm 3. When the angle detected is larger than the maximum angle, it is assumed that two adjacent markers have been wrongly labelled and so their labels are swapped.

Algorithm 3 Pseudo code for Heuristic 3

```
maxangle = 15
if m1 ≥ maxangle AND m2 ≥ maxangle then
  swap marker assignments
end if
```

The benefits of these constraints on the tracking system were calculated by assessing the percentage of markers correctly identified in a series of three frames at a few different points within the test video. The test video was taken from one of the performance videos recorded in the experiment in Chapter 9. These benefits were assessed for three different levels. The first was based on a basic system using only blob detection; the second was an improved system which incorporated basic heuristics to improve the rate of tracking; the third was a more advanced system using the full set of heuristics and blob tracking. Results show the basic system has a tracking accuracy of 63%. The improved system has a 23% increase in accuracy whilst the final system has a 40% increase, bringing the total accuracy in tracking to approximately 88%.

This accuracy was judged for when all points were available to track and not occluded from view as can sometimes happen in piano performance. The estimation of occluded points is dealt with in the next section.

4.4 Occlusion

A significant difficulty in hand tracking arises in occlusion. This happens regularly in piano performance, where the pitch range of notes for both hands overlap or in passages that require fingering patterns which place the thumb underneath the other fingers. The software can estimate the position of any "lost" markers by calculating the average transformation between each frame of the other points in the point set. The affine transforms of each marker are determined using the scaling, rotation and translation matrices below:

$$\begin{bmatrix} X_{\text{scaled}} \\ Y_{\text{scaled}} \\ 1 \end{bmatrix} = \begin{bmatrix} \text{Scale}_{\text{x}} & 0 & 0 \\ 0 & \text{Scale}_{\text{y}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} X_{\text{rotated}} \\ Y_{\text{rotated}} \\ 1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} X_{\text{translated}} \\ Y_{\text{translated}} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & D_{\text{x}} \\ 0 & 1 & D_{\text{y}} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \text{X} \\ \text{Y} \\ 1 \end{bmatrix}$$

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. Future work will calculate the motion vectors of each point, so that the software can predict occlusion and estimate the lost marker's position using the transformation matrices above.

A unique advantage of this software is that it allows a high degree of user intervention, so that any wrongly assigned markers can be corrected, and estimation points can be approved or changed. The software also has functions to allow the re-opening of existing files, allowing users to go back and change stored values.

Having tracked and estimated the positions of all the markers, we can now consider estimating the depth of each marker.

4.5 3D estimation

By using monocular images to track movement, the z position of the markers have to be calculated from reference points in the 2D image. 3D images could be captured by a stereo camera array, however, the resolution for two cameras at 83cm above the keyboard does not improve significantly to justify the extra expense of another camera or the computational processing load to allow raw image capture from another camera in synchrony. In an effort to produce a stable system that is cheap, portable and accurate, only one camera is used. However, the disadvantages of such a system arise when wanting to measure the exact angles of the fingers for any purpose that cannot settle for an estimation of the z axis.

The hand model for the pianists has been designed with several reference markers on the base of the hand to allow 3D estimation. Calculating a range of distances between the markers of the models, the z axis can be estimated by examining the difference of these distances between frames. The distances calculated are seen in Figure 4.8. Although these distances will be different for each person anatomically, as long as an initial frame is recorded that contains both hands laid flat on the keyboard, the z axis can be accurately estimated through the use of trigonometry.

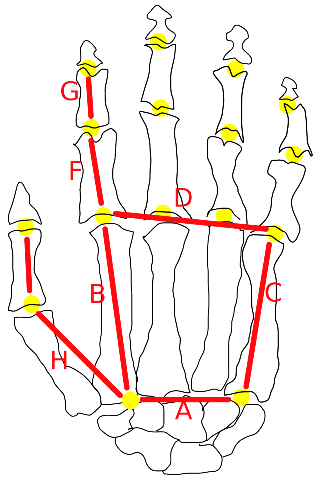


Figure 4.8: Hand Markers with Calculated Distances for 3D Estimation

Distance A is calculated between the two base wrist points, distance B is calculated between the left base wrist point and the first finger metacarpophalangeal point. Distance C is calculated between the right base wrist point and the fourth finger metacarpophalangeal point and distance D is calculated between the first and fourth metacarpophalangeal points. The distance from the thumb metacarpophalangeal point and the left base wrist point is distance H. Distances F_{thumb} and F_1 to F_4 are calculated for each finger as the distance between its metacarpophalangeal and proximal points. Distances G_1 to G_4 are calculated for each finger as the distance between its distal and proximal points.

Considering the view of the camera, we can consider how these distances change with changes in depth, as seen in Figure 4.9. The first image shows the four distances A, B, C and D at a flat level. As the hand is level and approaches the camera, i.e. rises away from the keyboard, distances A, B, C and D will all increase. This is viewed in image(b). Equally, as the hand is level and moves away from the camera, i.e. towards the keyboard, these distances will decrease. As the hand tilts forward and the wrist rises towards the camera, distance A increases whilst all other distances decrease. This is seen in image(c) with the opposite seen in im-

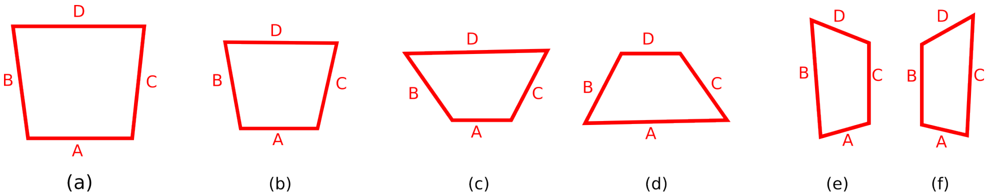


Figure 4.9: Changes in Hand Distances for Different Orientations. Image(a) shows the hand distances as a reference frame, (b) shows the hand moving away from the camera, (c) shows the hand tilting away from the keyboard, (d) shows the hand tilting towards the keyboard, (e) shows the hand tilting to the right and (f) shows the hand tilting to the left.

age(d). As the hand tilts to the right, distances A, B and D will decrease, however, distance C will either increase or stay the same. This is presented in image(e). As the hand tilts to the left, distances A, C and D will decrease, however, distance B will either increase or stay the same. This is presented in image(f). For each of the fingers, distances F and G will decrease as the finger is curved and increase as it is flattened. Considering the thumb separately, distance H will increase as the thumb moves towards the camera, and decrease as it moves towards the keyboard. Using these observations, estimations of depth for each joint can be devised as follows. As the hand has several degrees of freedom, depth for the wrist and metacarpophalangeal joints is calculated by using the average of the nearest two applicable distances in the x and y direction. This then accounts for tilt in the x and y directions. For all depth estimations, the new measurements of distance are compared to the initial frame zero:

Metacarpophalangeal joints

1st and 2nd finger:

$$z = f^2 \times \frac{\tan\left(\frac{B_{t=0} + D_{t=0}}{2} \times \frac{\theta}{p}\right)}{\left(\frac{B_t + D_t}{2}\right)} \tag{4.2}$$

3rd and pinkie finger:

$$z = f^2 \times \frac{\tan\left(\frac{C_{t=0} + D_{t=0}}{2} \times \frac{\theta}{p}\right)}{\left(\frac{C_t + D_t}{2}\right)} \tag{4.3}$$

Thumb:

$$z = f^2 \times \frac{\tan\left(H_{t=0} \times \frac{\theta}{p}\right)}{H_t} \tag{4.4}$$

where f is the objective distance of the camera i.e 830mm
p is the pixel width i.e. 1.1mm
θ is the camera's angular resolution i.e. 0.09°
t is time

Proximal interphalangeal joints

$$z = f^2 \times \frac{\tan\left(F_{n,t=0} \times \frac{\theta}{p}\right)}{F_{n,t}}$$

$$\text{where } n = \text{ finger number} \tag{4.5}$$

Distal interphalangeal joints

$$z = f^2 \times \frac{\tan\left(G_{n,t=0} \times \frac{\theta}{p}\right)}{G_{n,t}}$$

$$\text{where } n = \text{ finger number} \tag{4.6}$$

Wrist positions

Left wrist:

$$z = f^2 \times \frac{\tan\left(\frac{A_{t=0} + B_{t=0}}{2} \times \frac{\theta}{p}\right)}{\left(\frac{A_t + B_t}{2}\right)} \tag{4.7}$$

Right wrist:

$$z = f^2 \times \frac{\tan\left(\frac{A_{t=0} + C_{t=0}}{2} \times \frac{\theta}{p}\right)}{\left(\frac{A_t + C_t}{2}\right)} \tag{4.8}$$

These estimations will not be an altogether accurate calculation of the depth of each marker, as the calculations are based on distances between groups of markers and not the markers themselves. As the hand has so many degrees of freedom, it is highly complicated to calculate the depth for each marker, however, a z index calculation will be sufficient for the purposes of the experiment in Chapter 9.

4.6 Storage

Once 3D estimation is completed, the tracked information is stored in GMS files (see Section 3.4.2) which are structured in scenes, units, channels and tracks. For purposes of the FingerDance software, each scene consists of two units corresponding to each separate hand. Each unit then consists of 16 channels which represent the 16 markers on each hand. Each channel consists of three tracks to store the (x,y,z) coordinate of the marker, as required by the GMS file. This means that the retrieved geometrical data from the image processing software needs to be arranged in the same format to be read in to the GMS file. The data for each frame is stored as a list of numbers with the offset element number for each track recorded. When reading the GMS files, the offsets are used for each frame to locate the correct marker position.

4.7 System Improvements

Various improvements can be made to the system in terms of tracking accuracy, occlusion and 3D estimation. A larger set of advanced heuristics could be based on more restrictive constraints like those of Guan et. al [58]. These constraints are also based on Rijpkema's model but define a set of relationships between the angles of each finger. Occlusion could be improved by also calculating the velocity and direction of each point as it reaches occlusion to better estimate the correct position. Finally, 3D estimation can be improved by deriving a stronger algorithm that incorporates the angular relationships between each finger much like the improvements that can be made to the heuristics.

4.8 Applications

In conclusion, a motion capture system has been described that is cheap, portable, accurate and un-intrusive to performance. It is specifically designed to track finger motion in piano performance and also allows a great deal of user control in its estimation algorithms.

In its current version, this software can be used for a variety of purposes. Being able to track accurate positional information of the hands in piano performance can help to answer pedagogical questions on hand movement, identifying expressive movements and note accents. Investigating how finger curvature affects the acoustic sound in amplitude and in timbre is also possible by analysing the distances between the joint markers.

Future extensions for the software include incorporating models for other types of musical performance e.g. guitar playing and also being able to track fingering patterns by storing the position of the keys.

Chapter 5

Multi-Modal Capture System Design

This chapter describes the design of two full multi-modal capture systems using some of the commercially available capture technologies described in Chapter 3, as well as the specially designed finger motion capture system described in Chapter 4. These two different systems are required due to differing needs in motion capture. The Vicon incorporated system captures full upper body movement whilst the FingerDance incorporated system captures intricate measures of finger movement.

The two systems also demonstrate a number of advantages of using each type of motion capture technology. The Vicon system is entirely stationary and has been used solely within the University of Glasgow Psychology Department. The FingerDance system, however, is entirely portable, fitting on top of any 88-key piano, and has been used at the University of Glasgow, Napier University, the Royal Northern College of Music, Manchester and the Royal College of Music, London.

Self-reporting is included as part of the methodology for both systems, taking place immediately after the recordings. This enables the capture of each pianist's thoughts on their performance, to be used as extra information to inform future data analysis.

5.1 Architecture of Vicon Incorporated System

This multi-modal system is based on using the proprietary Vicon motion capture system. The architecture for this system can be seen in Figure 5.1. The Vicon setup presently in the Psychology Department at Glasgow University consists of twelve infra-red cameras placed around a capture volume of approximately $4 \times 3.8 \times 2$

metres. Retro-reflective markers are attached to the subject either directly onto the skin or applied with velcro to a specialised jacket and cap. Using triangulation, the system records accurate 3D positions of each marker at a rate of 120 frames per second. One of the limitations of using the Vicon system is that it is completely stationary and therefore, only keyboard instruments that are portable into the capture volume can be used. When recording performances, the pianists will play on a Roland RD-150 weighted keyboard.

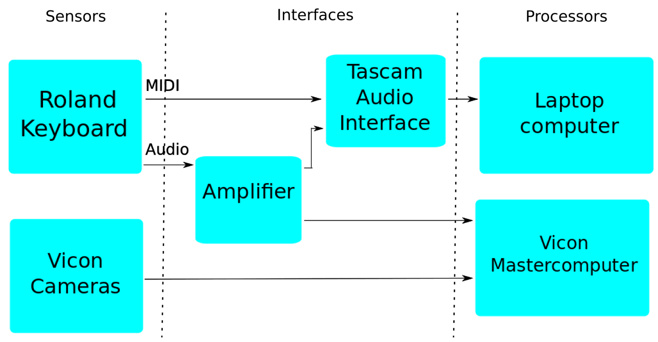


Figure 5.1: System Architecture for Vicon incorporated System

Audio is amplified from the keyboard via a Peavy KA/6 Keyboard Amplifier, and is recorded into a laptop computer via the Tascam US122 Audio Interface. This same audio is sent to the analogue card of the Vicon mastercomputer in synchrony with the motion capture recordings. These two audio recordings are used to synchronise the MIDI recordings with the motion capture data.

The MIDI out jack on the keyboard allows us to capture MIDI directly. This is transported to the computer via the Tascam audio interface. The Jack Audio Client is used to transport audio from the driver to the application Ardour [1] and also to transport the MIDI data to the MIDI sequencer Rosegarden [8]. Jack also allows synchronisation between the audio recording workstation and the MIDI recording software.

To retain a record of the images of the performance, a separate video is recorded by a Sony Handycam video camera placed in an 'audience perspective'. Figure 5.2 shows the setup for this system through the view of the 'audience perspective' video.

This system will be used to record audio and MIDI as well as capturing full body motion of the pianists to answer particular questions on the relationships

Figure 5.2: Audience Perspective of Vicon System Recordings

between body movement and musical structure, as outlined in Part III.

5.2 Architecture of FingerDance Incorporated System

This multi-modal system is based on using the FingerDance system and is designed to be entirely portable. This is vital particularly when working with professional pianists, as it is not always possible for them to travel to a stationary lab to record for an experiment. This system can be fitted around any standard 88-key piano in any venue. This means professional pianists can perform in any venue of their choice and more importantly, with a piano they like and with which they are familiar. Although this may result in recording performances on different makes of piano, the differences in hammer action between pianos are negligible [49]. The architecture for this system can be seen in Figure 5.3.

Video for the image tracking software is recorded through a high frame rate camera, the AVT Guppy F046. This camera is attached onto a microphone stand by using a specially designed thread adapter. The camera is connected to laptop computer B via the Belkin Firewire Interface P81800 and is configured through the open source application Coriander. Coriander records raw video at the fastest possible frame rate using the Region of Interest facility. The raw videos are then stored on an external hard drive. The computational load of recording raw images is to such an extent that an extra laptop is required so that accurate audio and MIDI capture is not sacrificed.

The UV blacklight required to make the passive paint markers fluoresce (as described in Section 4) is suspended above the keyboard of the piano by a specially

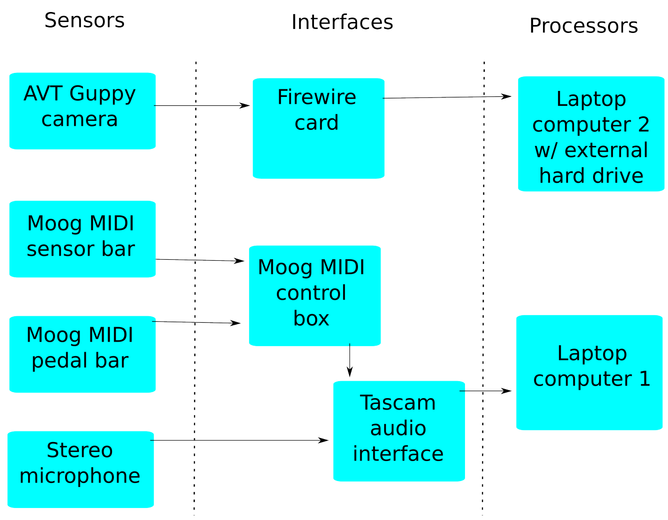


Figure 5.3: System Architecture for FingerDance Incorporated System

designed apparatus with stands at either end of the keyboard. This apparatus is seen in Figure 5.4. Figure 5.4(a) presents a side view of the apparatus, showing how the light is suspended over the keyboard whilst Figure 5.4(b) shows the construction of the adjustable poles at either side of the apparatus, allowing the light to suspend at heights from 122.5cm up to 189cm. Figure 5.4(c) shows how this apparatus is then placed in front of a concert grand piano. When in use, the apparatus is moved so that either side of the stand sits just in front of the keyboard.

The full configuration of the system along with two photographic lights and diffuser umbrellas providing normal lighting is shown in Figure 5.5.

Audio is recorded through a Beyerdynamic MCE82N(C) stereo condenser microphone placed a few feet from the open lid of the grand piano. This is connected via a balanced XLR lead through the Tascam USB audio interface to laptop computer A. Audio is transported from the driver to the audio application Ardour via the Jack Audio Connection Kit. This also provides synchronisation with the MIDI sequencing software and the audio recording software.

MIDI is recorded through the Moog Piano Bar, via the Tascam USB audio interface also to laptop computer A. The two sensors that make up the piano bar are connected to the control module which converts the signals into MIDI protocol. The Moog bar must be calibrated against the piano on which it is placed, with lights above each of the keys indicating whether the bar is sitting too high above or too close to the keys.

A Sony Handycam video camera is set up on a tripod with full view of the performer and the piano to record an 'audience view' of the performance.

(a) Side view of the whole apparatus

(b) Enlarged view of the adjustable poles

(c) Front view of UV apparatus in place in front of the piano

Figure 5.4: Pictures of the Specially Designed UV Light Apparatus

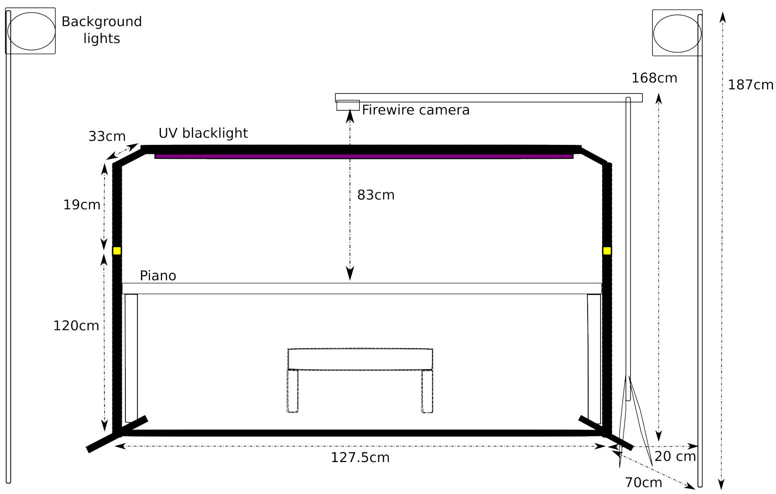


Figure 5.5: Lighting Configuration

Figure 5.6: Audience Perspective of FingerDance system recordings

This system will be used to record audio, MIDI and motion of the pianists' fingers to enable us to answer questions on the relationships between finger movement and acoustic sound, as well as their relationship with musical structure.

5.3 Post-recording Interviews

As part of multi-modal capture, it is necessary to record the performer's own thoughts on their performance, particularly with the research questions of the communication of structure in mind. Self-reporting technique is used as a way of finding out how the performers themselves interpret their performance. In each case, the performer watches the 'audience perspective' video recorded with the Sony Handycam, presented in both audio and video. They are asked to indicate to the experimenter where the sectional and phrase boundaries are. Any other points of interest e.g.harmonic tension are expected to be indicated as well. The experimenter then marks this down on the relevant place in the score. The performers are then asked to perform the same sort of task by looking directly at the score. These markings are compared against the markings made by the experimenter and also to traditional analyses of the piece.

The self-report is conducted in this way to avoid performers becoming traditional 'analysts' when they perform segmentation directly from the score. Gingras and colleagues found that when asking organists to mark directly on a score their manual segmentation of a Bach fugue, their phrase analysis was vastly different to the results taken from what they had performed. He suggests that this was a result of presenting performers with a score as they were potentially using a different rule set to perform traditional segmentation instead of marking down the segmentation they had performed [45].

A general open interview on the performer's views of motion in performance takes place after the audiovisual segmentation exercise. The basic questions that are asked to each performer are:

- How do you express structural features like the ones you have marked on the score?
- What are your views on movement in a performance? Is movement necessary or does it hinder a performance?

Results from these interviews can help in interpreting the numerical performance analysis both in the motion differences between performers as well as the segmentation of the pieces of music.

Chapter 6

Storing and Visualising Data

Acquiring such a plethora of multi-modal data requires effective storage containers and visualisation formats to make musical sense of the data. The following technologies developed at the University of Glasgow work in conjunction to store music and performance data effectively and produce readable output of musical queries valuable to musicologists and other performance scientists.

6.1 Data Storage

One of the difficulties that face performance analysts today is the proper representation and storage of musical data. This section describes a specially developed storage container to allow performance metadata to be stored in alignment with information about the musical score.

6.1.1 PML

Performance Markup Language (PML) developed by Douglas McGilvray at the Centre for Music Technology in the University of Glasgow [82], was particularly designed to accommodate the mark-up of performance information alongside the score. PML is a specification which can be used to extend XML-based score representations such as Music-XML. Analytical, performance and score information are separated into different hierarchies. Since MEI represents these domains in a single hierarchy, which is based on the requirements of the features of the musical score, it makes it a less elegant solution for the representation of other data which may be non-isomorphic with the score. For example, one would not expect the repeated portion of a da capo aria to be performed the same way the second time.

The performance data in a PML file is stored at the end of the MusicXML note list and IDs link aligned performed notes to score notes. This allows more than one performance note to be aligned to one score note.

Conversion into PML begins with MusicXML versions of the musical score and MIDI performance recordings. Several steps are taken to store the separate files of information and create links between the score and performance data. This includes a matching algorithm which uses Dynamic Time Warping to find the optimal mapping between score and performance.

- musicxml2pml - The MusicXML file of the score is converted into the structure of a PML file.

- midi2pml - The MIDI file of the performance is added as a performance structure into the PML file.

The PML file at this point shows the two separate hierarchies for score notes and performance notes. This can be seen in the file fragments in Figure 6.1.

- winmatch - The polyphonic matcher is run to align the performance notes with the score notes. This is done using a running window.

- intermatch - The interpolation algorithm is run to minimise errors in matching.

The pml file at this point now contains links to note ids in the performance part which identify which score note they are associated with. This can be seen in the code fragment in Figure 6.2.

Other formats of performance data can be added such as audio files. Functionality for adding different gesture formats is currently in development.

6.1.2 The future of multi-modal storage

A music and gesture container format is in development that will allow the synchronous storage of different types of music and performance data such as MIDI, audio and gesture data. Music and Gesture format (MGF) [96] files allow the storage of audio and video attached to musical scores. This format extends MusicXML to include data from these other sources inside a compressed archive.

```
<pml>

  <score-partwise>
<work>
</work>
<identification>
</identification>

 <part-list>
        .
        .
        .
</part-list>

  <part id="P1">
    <measure number="1">
        <print>
          <staff-layout>
            <number>2</number>
            <staff-distance>70</staff-distance>
        </staff-layout>
      </print>
<attributes>
       <divisions>8</divisions>
       <key>
      <fifths>3</fifths>
</key>
      <time>
      <beats>3</beats>
      <beat-type>4</beat-type>
         </time>
      <clef number="1">
      <sign>G</sign>
      <line>2</line>
      </clef>
      <staves>2</staves>
                              <clef number="2">
<sign>F</sign>
      <line>4</line>
</clef>  </attributes>
    <note id="note1">
      <rest/>
      <duration>8</duration>
      <voice>1</voice>
      <type>quarter</type>
      <staff>1</staff>
          <starttime>0</starttime>
        </note>
        .
        .
        .
        <barline location="right">
<bar-style>light-heavy</bar-style>
</barline>  </measure>
  </part>
  </score-partwise>

  <performance>
    <perfpart part="P1">
      <event id="pnote1">
        <onset>4.90729</onset>
        <end>5.79063</end>
        <midi>64</midi>
        <velocity>36</velocity>
      </event>
      .
      .
      .
```

Figure 6.1: PML file fragment before performance-score matching

```
<performance>
   <perfpart part="P1">
     <event id="pnote1">
       <onset>4.90729</onset>
       <end>5.79063</end>
       <midi>64</midi>
       <velocity>36</velocity>
       <align note="note3">correct</align></event>
     <event id="pnote2">
       <onset>5.65937</onset>
       <end>5.91979</end>
       <midi>40</midi>
       <velocity>26</velocity>
       <align note="note13">correct</align></event>
```


Figure 6.2: PML file fragment of matched file

6.2 Visualising Data

After storing musical data in a coherent and effective way, the data needs to be visualised in an effective manner. Lists of numbers and graphs are not useful to musicologists trying to establish answers to questions e.g. of dissonant harmony related to expressive tempo. The Pullinger Database provides a tool for visualising the performance metadata alongside the musical score.

The one feature missing from the visualisation tools described in Section 3.5 is a representation of the music being performed. The word 'representation' is used as there is not always a score for music that is not within the Western classical music genre. The Pullinger database [95, 94]provides a tool that can display results of musicological queries alongside a representation of the notated score. It does this by populating a postgreSQL database with information about the score and information about the performance in separate tables linked by IDs.

This database can be used in conjunction with PML information (see section 6.1.1). Once the PML files have been uploaded to the database, they are available for querying. The database uses a particular pitch representation based on the spiral of fifths. This allows operations on the pitch information to be performed in order to analyse chords in terms of consonance and dissonance and also on groups of sequential notes to determine the intervals. The database also uses a representation of time which instead of using a method of implied time like MusicXML, explicitly defines the score time for each note. This allows more specific operations to be performed on timing within a piece of music without having to calculate the projected onset time of each note by parsing the file from the beginning.

After the matched PML file is uploaded to the database, each performance can be queried with musical functions such as highlighting dissonant intervals in the score and showing the inter-onset interval information for the performance of these. After the query is sent to the database, a document is created and then populated with the results of the query using Lilypond typesetter [3]. The results of this query is shown in Figure 6.3.

Since this technology allows easy comparison of different performance values with the notated musical notes intra performance and inter performance, it will be used in the experiments in Part III.

Figure 6.3: Example of database produced result to query on dissonant notes and IOIs

Part III

Experiments and Results

This part describes experiments that have been conducted using the methodologies and tools explained in Part II. These experiments have been designed in order to answer the musicological questions posed in the introductory chapter of how to elucidate musical structure from multi-modal performance data. This also explores how physical gestures ranging from large scale body movements to intricate finger movements align with the performer's interpretive choices and whether these can be used as indicators of structural features.

The first set of experiments explore the relationship between general body movement and phrasing structure, and use this in tandem with audible parameters to examine the multi-modal changes taking place at these structural boundaries. The second experiment uses these relationships to discover structural features where there is a certain ambiguity in traditional score-based analyses. The musical compositions performed by the pianists in each experiment are chosen specifically to expose these relationships between performance and score in a Western classical music context.

Chapter 7

Musical Stimuli

Three Chopin pieces are used as the musical stimuli for these experiments: Prelude in A major Op.28 No.7, Prelude in B minor Op.28 No.6 and the finale movement of the Sonata in B flat minor Op.35. The two preludes come from the same Op.28 set which is a standard set of repertoire for pianists. There also exists a number of analyses on the preludes and they tend to produce coinciding views on their structure. These are ideal pieces to explore the roles of aural and visual parameters in conveying structure. The finale of the sonata however, is a piece that can encourage completely divergent views on its structure. For this reason, it is used as a test piece for being able to use performance parameters to discover musical structure. In both sets of experiments detailed in Section III, the Prelude in A major No.7 is used as a control piece. This chapter provides traditional analyses of each piece from which the investigations into 'performed' structure can proceed.

7.1 Chopin's Prelude in A major op.28 No.7

Prelude No.7 in A major has a strict, rigid structure, with a rhythmically identical two bar phrase occurring eight times in total. As can be seen from Figure 7.1, this binary form 16-bar piece has the main boundary between the two sections occurring exactly halfway through at bar 8. The harmonic climax of the piece occurs with the F sharp minor chord at the end of bar twelve. The two sections of the piece are thought to each contain a set of antecedent-consequent phrases.

This explicit and rigid structure is what makes this particular Prelude a good control piece. Composed rhythm between phrases remains at a constant whereas pitch, harmony and structural importance changes between phrases. This importance is highly dependent on the underlying harmony and melodic contour. This piece is used as a control piece in each of the experiments in Section III.

As mentioned in Chapter 2, Bisesi and Parncutt's accent analysis of this Prelude is included for reference as Figure 7.2.

Figure 7.1: Phrasing Analysis of Chopin's A major Prelude op.28 No.7, with blue marks for sectional boundaries and red marks for phrase groupings

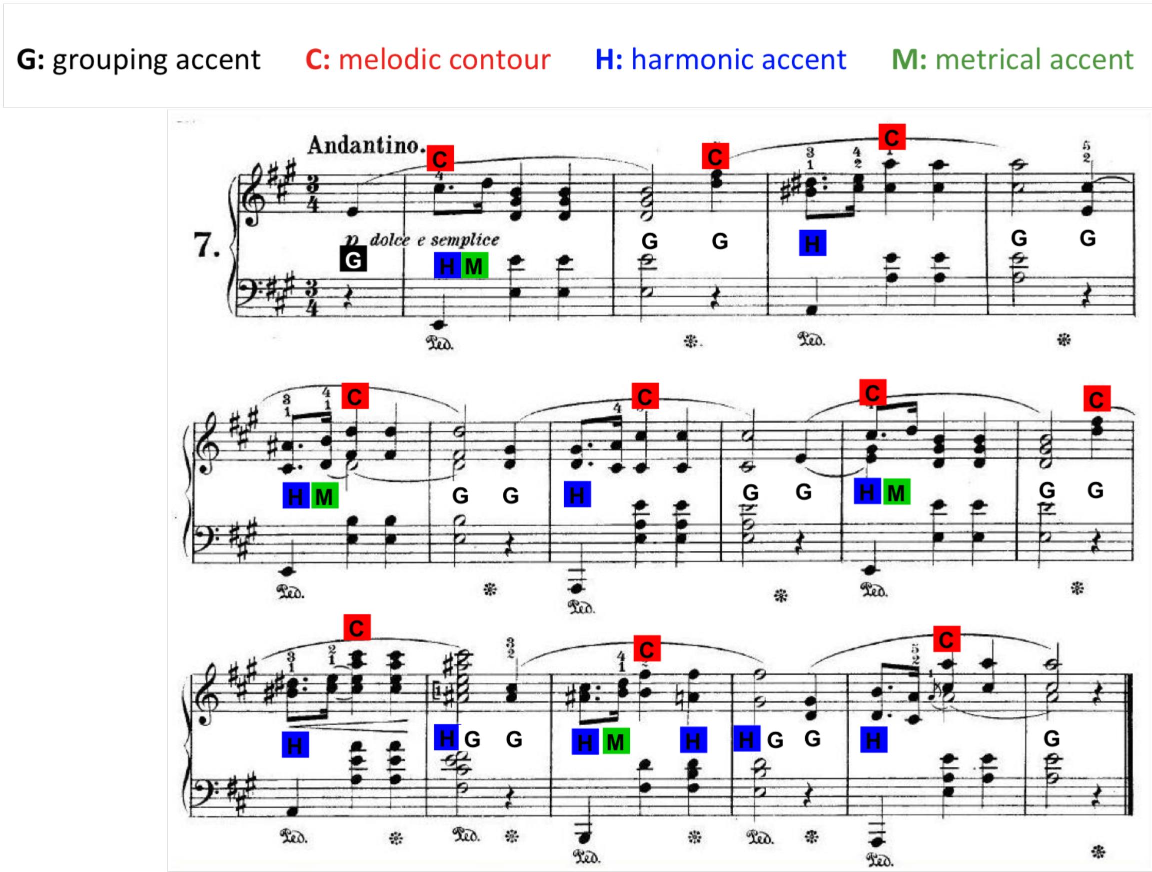


Figure 7.2: Bisesi and Parncutt's Accent Analysis of Chopin's A major Prelude op.28 No.7. Taken from Erica Bisesi and Richard Parncutt, Private Communication. This figure represents a preliminary stage of the analysis by the authors and has been presented by Erica Bisesi at the Opening Ceremony of the Centre for Systematic Musicology - University of Graz, held on 15th October 2009, and is part of her Lise Meitner Research Project M 1186-N23 sponsored by FWF, Austria. Permission to reproduce this figure has been granted by the authors.

7.2 Chopin's Prelude in B minor op.28 No.6

Prelude No.6 in B minor can be segmented into three sections from bars 1-8, bars 9-22 and a coda section from bars 23-26. In the first section we see the representation of an 'extended idea'. As seen in Figure 7.3, Chopin begins with a two-bar motif in B minor. This motif is repeated with a slightly higher pitch range in the next two bars. The first part of the motif is repeated again for a third time and then expands into a four bar phrase ending at bar 8, the first sectional boundary. The second section represents an expansion of this idea. At bar 9, the original two-bar motif is repeated with the next expansion moving into C major. A new four bar phrase is introduced at bar 15, answered by the consequent four bar phrase arriving at the tonic at bar 22, producing the second sectional boundary. The piece concludes with a short coda in B minor in its final phrase [1].

Again, as mentioned in Chapter 2, Bisesi and Parncutt's accent analysis of this Prelude is included for reference as Figure 7.4.

In the experiments in Chapter 8, this piece is used in combination with the control piece to examine how visual gestures relate to phrasing boundaries, and also how tempo, dynamics and motion patterns can be used to detect musical structure. The first three phrases of this Prelude show an extension of the original two-bar phrase. This can be compared structurally against the rhythmically repeating two-bar phrases of Prelude 7 in A major.

[1]This analysis of Chopin's Prelude Op.28 No.6 is combined from Kofi Agawu,V. 'Concepts of Closure and Chopin's opus 28' in Music Theory Spectrum 9:1-17, 1987 [65] and comments made by Jennifer MacRitchie, University of Glasgow, and David Lewis and Christophe Rhodes, Goldsmiths, University of London

Figure 7.3: Phrasing Analysis of Chopin's B minor Prelude op.28 No.6 with blue marks for sectional boundaries and red marks for phrase groupings

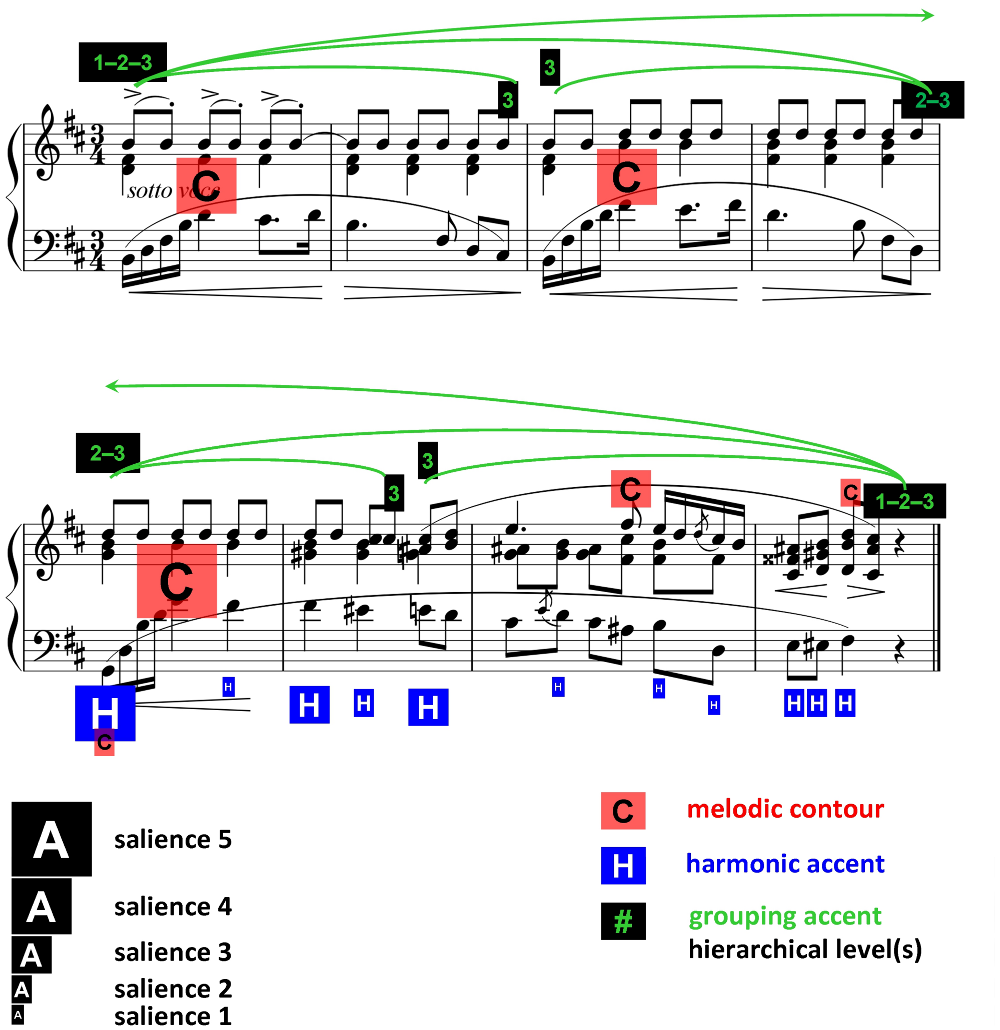


Figure 7.4: Bisesi and Parncutt's Accent Analysis of Chopin's B minor Prelude op.28 No.6. This figure is taken from Bisesi and Parncutt (2010), An Accent-Based Approach to Automatic Rendering of Piano Performance [15]. This figure is reproduced with the authors' permission.

7.3 Chopin's B Flat Minor Sonata op.35 Finale Movement

The finale of Chopin's B Flat Minor Piano Sonata Op.35, the first 8 bars of which can be seen in Figure 7.5, has been referred to as "a wild child, unique and well-nigh indescribable"[116]. A short piece written for the most part in octaves, this rhythmically unrelenting and binary sonata form composition has confounded traditional approaches to its analysis.

The existing written literature on this particular piece is very sparse with comments being both anecdotal and impressionistic, probably due to the problematic nature of the composition. Only Charles Rosen [106] has written an extensive essay and most of his statements are very non-committal, even though his authority as a pianist prompts us to take them seriously. For our purposes, this problematic nature of the work makes the data more suitable for objective, quantitative methods.

Rosen's analysis of the piece sets the first four bars as the introduction in the dominant key of B flat minor, with the chromatic main theme entering in bar 5. After bar 8, there is a transition section where the harmony of the chromaticism gradually defines the dominant of the relative major key. A new theme set in D flat major enters at bar 23 and is repeated an octave higher at bar 27. The recapitulation begins at bar 39 by literally repeating the first eight bars of the composition and then expanding the recapitulation of the following bars with parts of the transition and the second theme, moving towards a cadence.

Another viewpoint on the segmentation of this piece comes from Michael Talbot [115]. His segmentation of the finale is seen in Figure 7.1. Contrary to Rosen's view that the first four bars are set as an introduction, Talbot determines the first eight bars as the first theme.

Further different analyses are summarised by Lindstedt's work on segmenting the finale using computer analysis [74]. One of the first arguable points is the entry of the first theme and establishing whether the first four bars are an introduction. These traditional analyses are taken as a starting point in the following investigation in Chapter 9. From examining patterns of tempo, dynamics and motion at phrasing boundaries in the control piece, the performer's interpretation of structure in the finale can be highlighted and points of agreement and departure across performers can be examined.

Features of this piece which make it ideal for computational analysis are its

Figure 7.5: Chopin's B flat minor sonata op.35 finale movement measures 1-8

constant rhythms, as every single bar except the final few consist of twelve quavers. Any differences in rhythm therefore will be entirely due to the performer's manipulation of inter onset intervals and keypress durations etc. The right hand melody is also perfectly replicated an octave below in the left hand and so chord separation and melody lead are not an issue.

As previously stated, all pieces of music analysed in this Chapter will be used in combinations for experiments in Chapters 8 and 9.

Table 7.1: Talbot's Analysis of Chopin's B Flat minor Sonata Finale Movement Op.35

Bars	*Key*	*Description*	*Comment*
1-8	b flat	first theme	establishing tonic
9-22	modulating	transition	chromatically unstable
23-30	d flat	second theme	diatonic
31-38	modulating	retransition	sequential progressions
39-46	b flat	first theme	reprise of bars 1-8
47-56	modulating	transition/second theme	based on bars 9-30
57-75	b flat	coda	largely diatonic

Chapter 8

Detecting musical structure

This experiment was designed to answer the research questions:

- Can gesture across performers be seen to communicate musical structure?
- Looking at multi-modal parameters, can structure be detected from this?

To enable this exploration of communication, this particular experiment was designed and executed in collaboration with psychologist Bryony Buck, a colleague at the Centre for Music Technology. The full detail of the performance analysis and following audience perception experiments can be seen in [17, 78, 77]. These extend the work of Wanderley and Vines [84, 123] in analysing clarinet performances and the communication of phrasing and tension from performances of the opening of a Brahms sonata. The main aims of our research were to discover how structural information was being conveyed in a performance through aural and visual parameters by using the recorded performances as stimuli to be presented to audiences in audio-only, visual-only and audiovisual modes. The participants were asked to 'shape' the phrasing structure through the use of a slider and the relative contributions of aural and visual information in carrying out these judgements were assessed. As we cannot tell how the audience participants will be making their judgements of visual movements, general body motion is a factor and so for this reason, the Vicon incorporated system (as detailed in Chapter 5.1 was used to capture the recordings. This chapter describes the performance analysis part of these experiments.

To properly analyse how aural and visual gestures are performed within the context of phrasing structure, two pieces are chosen which have a similar structural make-up. The first is an example of a group of phrases with a strictly identical rhythm and the second contains phrases which begin as rhythmically identical but expand the duration of notes in later phrases to form a different rhythm. These are Chopin's Prelude in A major Op.28 No.7 and Chopin's Prelude in B minor Op.28 No.6 respectively whose analyses can be seen in Chapter 7. The first piece is used as a scientific control. These pieces satisfy a number of criteria:

- Short pieces are preferred as the Vicon motion capture system works optimally with short recordings.
- The genre of the music may have an effect on the musical gestures used to express the performer's interpretation and so Western romantic style pieces are used.
- The Chopin Preludes Op.28 set are a widely known and performed set of repertoire with many analyses and recordings available. The structure of these preludes are quite clear with the existing analyses widely agreeing. Differences in interpretation therefore arise from the hierarchical importance of the boundaries and not the position of the phrasing boundaries themselves.

Nine highly trained pianists from different universities and conservatoires in Scotland performed two Chopin Preludes (The A major Prelude No.7 and the B minor Prelude No.6) and were recorded through audio, MIDI and video by use of the Vicon incorporated multi-modal capture system. These nine performers consisted of five music performance undergraduate students, four at the University of Glasgow and one at the University of Edinburgh, two postgraduate students from the Royal Scottish Academy of Music and Drama and two amateur pianists with more than ten years of performance experience. Each pianist was paid a one-off sum of £25 for their participation in the experiment. Given preparation time of just over one calendar month, each pianist had been asked to memorise the pieces in order to guarantee a certain amount of practice and consideration of the structure of the piece. The pianists were asked to perform as in a normal concert setting. No other performance directions were given.

Using the multi-modal capture system described in Section 5.1, performances of the two selected Chopin preludes by nine highly trained pianists were recorded. Retro-reflective markers were placed onto a velcro jacket and hat worn by the performers in the configuration shown in the head and upper body model in Figure 8.1.

This particular model combined the upper body model from Cutti *et al.* [30] with four reference markers for the head positions. Each camera tracks the coordinates of the 28 markers and triangulates their position in order to build a 3-D

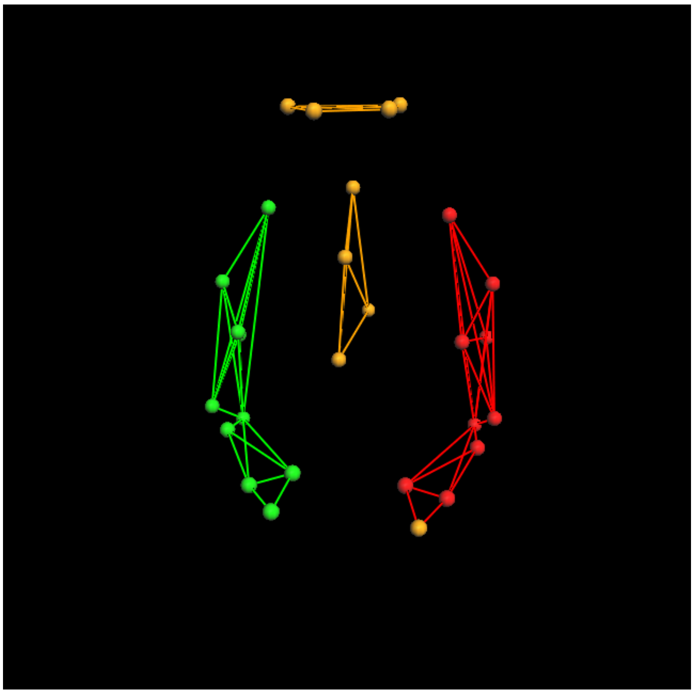


Figure 8.1: Upper body model markers

model of each performer. The models were then reconstructed by post-processing and any points where the cameras had failed to track a certain marker were filled with the estimation models available from the Vicon Nexus software. Problems were encountered particularly with the markers placed on the elbows of the performers. As the markers were placed not directly onto the skin but onto a velcro jacket, there were several points in the recordings where the marker was lost by the camera as the jacket had moved round the elbow and displaced the marker. Although the Vicon interpolation algorithms filled most of these elbow gaps, the system is proprietary, so these algorithms are unavailable for inspection. The accuracy of reconstruction for these elbow gaps must therefore be considered suspect. The abbreviated names for each marker are shown in Figure 8.2. The left and right sides are labelled with respect to the pianist. Each marker is recorded for the x, y and z axes.

MIDI and audio information was recorded for each performance through the setup described in Section 5.1 and aligned and stored as PML files (see Section 6.1.1).

Even though the interpretations will not be wildly diverse, there can be some differences between performers and so each pianist's own interpretation of phrasing is noted in the self-report taken as part of the experiment, in which their views on movement in performance were also noted. Analysis of the aural and visual parameters of each pianist's performance of the rigidly structured Prelude in A

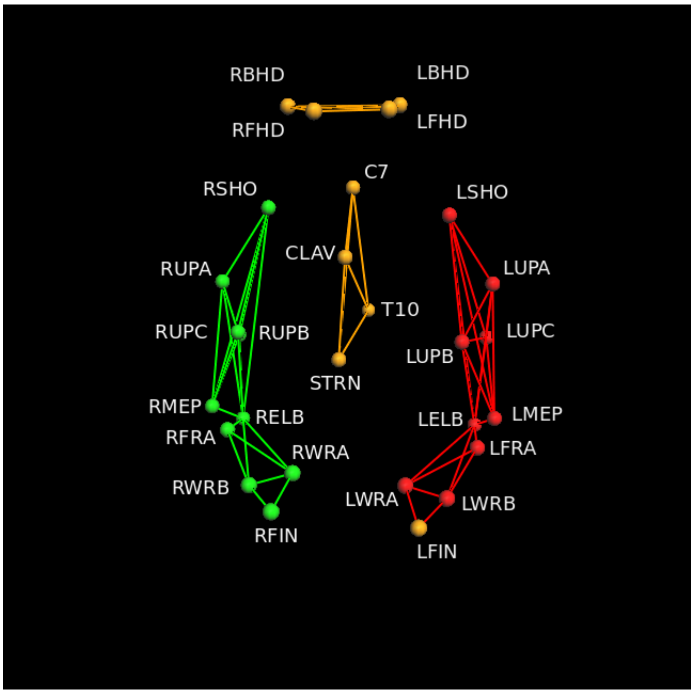


Figure 8.2: Upper body model Marker Definitions

major will provide an impression of performance style, particularly at phrasing boundaries. This piece provides the opportunity to observe movements for each phrase in isolation before moving on to examine slightly more complicated structures in Prelude 6. The hypotheses from these experiments are as follows:

Hypothesis 8.1 *Regardless of the subjective and personal nature of physical gesture in relation to musical structure, there will exist an underlying pattern that is related to phrasing and is common across all performers.*

Hypothesis 8.2 *The underlying motion profile of the performer related to phrasing will be the same across pieces.*

It is expected that performers will show components of motion relating to phrasing in different parts of the body to each other, meaning there is no standard across performers. However, general motion is expected to conform to phrasing patterns of the composed music. This will be compared across Preludes to see how similar the underlying motion norm is of each performer when performing different pieces of music. Gesture is then considered alongside audible parameters for phrasing identification:

Hypothesis 8.3 *When investigating the role of gesture in multi-modal detection of phrasing, a combination of aural and visual parameters will provide the most accurate indicator of phrasing.*

Examination of these parameters will then consider the actual values, in particular the maxima and minima:

Hypothesis 8.4 *Where combinations of global maxima and minima occur in both aural and visual streams of data, these will be related to the most important structural features of the composition.*

First I will examine how gestures relate to the phrasing structure of each piece to establish that there is a relationship between movement and structure. I will then consider the multi-modal parameters to examine how visual gestures and aural gestures interact all within the context of the phrasing. This analysis is an extension of the analysis performed in [78, 77].

8.1 Motion Analysis Techniques

As motion capture always produces such an overwhelming plethora of data, the traditional phrase analysis of each prelude provides us with points from which to start investigation of gestural cues at phrasing boundaries. Each performer's audio recording was annotated in Audacity [2] by a separate professional pianist with the timings of the phrase boundaries noted in Chapter 7. Each performer's own view of the phrase segmentation was also noted in case of any differences to traditional analysis.

As there are 28 markers $\times$ 3 axes of data simply for the motion stream, reductional techniques are applied to reduce the number of variables. Two types of reductional algorithms are principal components analysis (PCA) [92] and the newer functional data analysis (FDA) [97]. However, as FDA warps the timing slightly to fit in factorial curves to the data, it was decided to use PCA to view the overall general motion characteristics of each performer. This was calculated through designated pca modules using non-linear iterative partial least squares (NIPALS) algorithms [6] on the complete set of motion data for each pianist. This improves upon the original analysis performed in [78] in which singular value decomposition was used to devise the principal components. The NIPALS method developed by H.Wold [128] is the most commonly used method for performing PCA and gives more accurate results compared to singular value decomposition. The NIPALS method also avoids calculating the covariance matrix and so greatly reduces the computational processing. This is a particular problem for high dimensional matrices such as the ones created from this motion capture data. The NIPALS PCA motion data is used in the analysis in [77].

Principal Components Analysis allows us to retrieve a comparable motion norm for each performer. It can also calculate details of relationships between the several markers. If all marker trajectories are similar to each other, only one significant principal component will emerge. The variance of the first PC shows if there is commonality between the patterns of motion in each marker. A high variance will show high commonality. e.g. 64% will show considerable commonality between markers but still leaves some room for alternative patterns. We can then see how each marker correlates with these principal components by looking at the loadings scores. These are exactly that - a measure of correlation between each marker and the PCs. If any markers appear to be leading the motion of the rest of the body, we can expect high loadings for a few markers and low loadings for the rest. Each principal component may be considered a 'motion profile' and so by calculating a weighted sum of the components, this gives us a better estimate of overall motion. Reduced dimension curves such as these are good at expressing a general overview but inevitably lose some semantics of the actual movement being performed and so after considering PCA results for each performer, each individual marker is then also examined for reference to phrases, measures and beats.

Each performer's principal component score was mapped against the timings of each phrasing boundary to determine if there was a pattern of movement for each phrase. Three pianists have been chosen to demonstrate the spread of results concisely. These pianists were chosen according to their ability, their standard deviation and variance of movement calculated for intra-performance data on a few selected markers, and also their views on movement during a performance. The pianists' self-reports also conveyed a wide view on the role of movement in performance, with some branding movements extra to sound productive ones as completely unnecessary and something they tried to limit, whilst others felt it vital to move in order to 'feel' the music they were performing. Although physical gestures in performance can be classified either as movements necessary to the actual sound-production or movements that are related to the music but not necessary for the actual sound (i.e. ancillary) [21], it is acknowledged that gestures may still be multi-functional. The performers chosen to display a range of results also reflect these varying opinions on the role of movement in performance. Performer 1 is a highly trained amateur pianist and had a small standard deviation of movement. Performer 2 is a conservatory trained postgraduate student and had a large standard deviation of movement, and Performer 3 is a music undergraduate student and had a mid-range standard deviation. The results from the other performers can be seen in the appendices. Normalization of results allows the movements to be correlated with phrase structure independent of differences in amplitude. The arrows in each graph indicate the point in time where the last note of each phrase ends in the audio stream.

8.2 Gesture Results

Following reconstruction of the markers, principal components analysis (PCA) was performed on the complete set of 84 marker variables (28 markers each containing information for three axes x, y and z) for each pianist. Axis x related to movement towards and away from the keyboard of the piano, axis y related to movement along the keyboard and axis z was related to height. The scores of each principal component reflect the general motion for a reduced dimensionality. The first two principal components are plotted for each performer. As each of these two components have relatively low variance weighting, the weighted combination of the first six principal component scores accounting for more than 90% of the variance was then calculated for each performer and again plotted against the phrasing boundaries. Each performance was time-warped with respect to the audio recordings to allow comparison between performers. This time-warping algorithm resampled each set of data to 10,000 points using the occurrences of phrasing boundaries in the audio stream as references, essentially downsampling for the motion data. When used later for the tempo data, the data is upsampled linearly. This is because the tempo estimations can only be made between a set of two notes, and so the sampling rate will be far smaller than that for the dynamics measure of the audio or the motion data. Ensuring each performer had an identical number of samples allows us to perform statistical tests across the group of performers.

8.2.1 Prelude in A major No.7

Beginning with Prelude 7 in A major, in which the pianists' self-reporting analysis agreed with the traditional phrase segmentation marked in Chapter 7, Figures 8.3, 8.6 and 8.9 show the first two principal components accounting for around 70% of the overall variance in motion for pianists 1, 2 and 3. The patterns of each principal component appear to relate to the phrasing boundaries described by traditional analysis methods. Looking at the loadings results of the PCA, or in other words the correlation between each marker and the resultant PCA scores, there did not appear to be any single prevalent markers causing the most variance in motion. The PCA curves are a result of the variances in a combination of several markers and these differ slightly for each pianist. The top ten correlations between the first two principal components and the body markers are seen in the two tables following each graph with the expanded full list of loadings for Performers 1, 2 and 3 seen in Appendix A. The full list of loadings for every pianist highlighting the top correlations between the first two principal component scores and each marker are also included in the appendices.

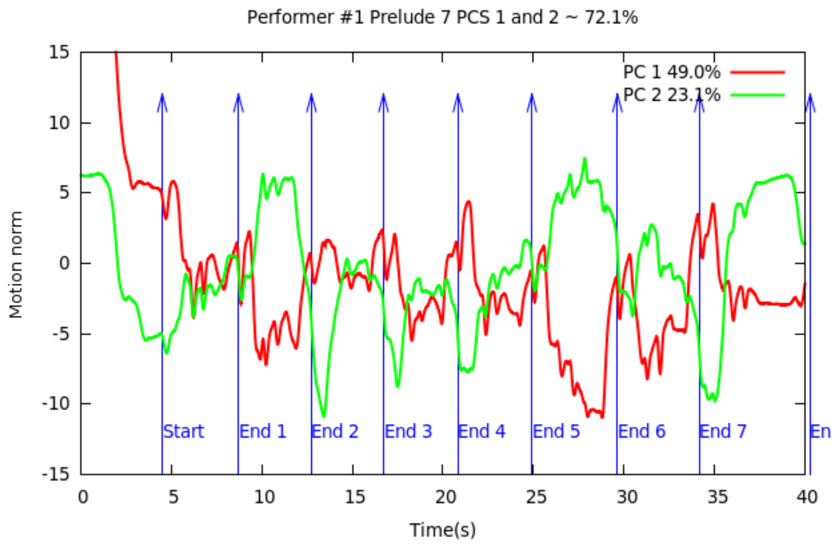


Figure 8.3: First Two Principal Components of Movement for Performer 1 , Prelude 7, the first component accounting for 49% variance and the second component accounting for 23.1% variance, with blue vertical lines representing phrasing boundaries as in the audio recording

Interestingly, Performer 1's recorded opinion on the role of movement in performance leaned towards the view movement in performance did not convey any information on phrasing and that during performances, he/she attempted to minimize movements and facial expressions. However, the graph of the first two principal component scores against phrasing boundaries (seen in Figure 8.3) indicates a clear relationship between overall movement and phrasing. It is noted, however, that the peaks of each component score occur in different points of time within each phrase. Although the peaks of the first component clearly relate to phrasing, the inter-phrase movement appears to move on a lower beat level. The global maximum and minimum of these scores occur at the position of the harmonic arrival

Marker	PC1	PC2	PC3	PC4	PC5	PC6
T10:X	0.12	0.02	0.2	-0.02	0.08	-0.05
LUPA:X	0.12	-0.05	0.15	-0.11	0.09	0.02
LUPB:X	0.14	-0.01	0.05	-0.09	0.06	0.02
LUPC:X	0.14	0.04	0.1	0.01	-0.01	-0.01
LELB:X	0.14	0.08	-0.02	0.04	-0.05	0
LMEP:X	0.15	0.03	-0.01	-0.03	0.01	0.03
LFRA:X	0.14	0.08	-0.01	0.04	-0.06	0.02
RUPC:X	0.11	0.13	0.08	-0.1	-0.06	0.03
RFHD:X	0.12	0.06	0.14	0.06	-0.1	-0.08
LFHD:X	0.12	-0.01	0.2	-0.01	0	-0.03

Figure 8.4: Top Ten Loadings for the First Principal Component, Performer 1, Prelude 7

Marker	PC1	PC2	PC3	PC4	PC5	PC6
CLAV:Y	-0.1	0.17	0.03	0	0.06	0.06
STRN:Y	-0.1	0.17	0.05	-0.02	0.03	0.02
LSHO:Y	-0.1	0.16	0.03	0.02	0.07	0.06
LWRA:X	0.09	0.15	0	0.15	-0.14	-0.03
RSHO:Y	-0.1	0.17	0.01	0	0.11	0.08
RUPA:Y	-0.1	0.16	0	-0.1	0.05	0
RUPC:Y	-0.09	0.16	-0.01	-0.19	0	-0.06
RFHD:Y	-0.09	0.17	0.05	0.06	0.13	0.05
LFHD:Y	0	0.21	0.01	0.1	-0.03	0.02
LBHD:Y	-0.06	0.19	0	0.04	0.1	0.11

Figure 8.5: Top Ten Loadings for the Second Principal Component, Performer 1, Prelude 7

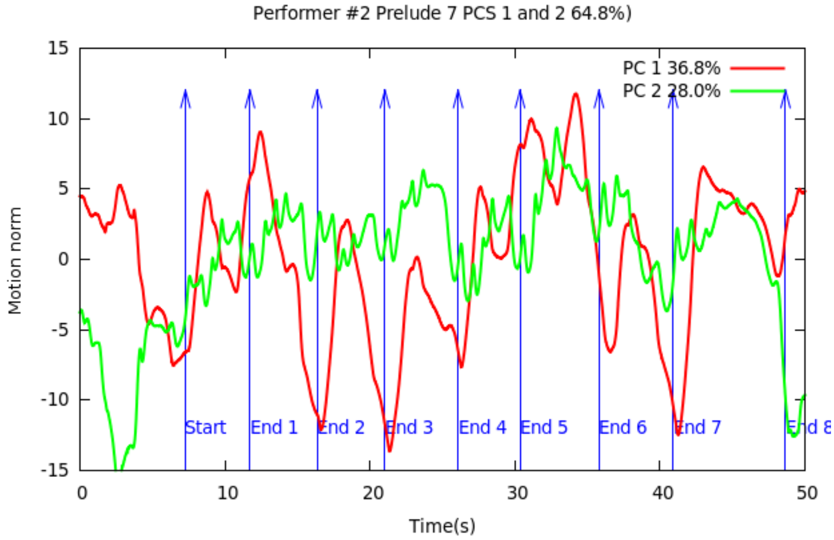


Figure 8.6: First Two Principal Components of Movement for Performer 2, Prelude 7, the first component accounting for 36.8% variance and the second component accounting for 28% variance, with blue vertical lines representing phrasing boundaries as in the audio recording

Marker	PC1	PC2	PC3	PC4	PC5	PC6
T10:Y	0.16	0.09	0.01	0.01	-0.04	0.06
CLAV:Y	0.16	0.09	-0.02	0	-0.04	0.03
STRN:Y	0.16	0.08	0	0.03	-0.03	0.03
LSHO:Y	0.16	0.1	-0.03	0	-0.04	0.03
LUPA:Y	0.16	0.09	-0.02	0.05	-0.03	-0.01
LUPB:Y	0.16	0.09	0	0.05	-0.03	0.02
LUPC:Y	0.16	0.07	-0.01	0.11	-0.03	-0.05
RSHO:Y	0.16	0.09	-0.03	-0.01	-0.06	0.05
RUPA:Y	0.16	0.1	0.01	0.01	-0.01	0.07
RUPC:Y	0.15	0.1	0.01	0.02	0.03	0.08

Figure 8.7: Top Ten Loadings for the First Principal Component, Performer 2, Prelude 7

Marker	PC1	PC2	PC3	PC4	PC5	PC6
T10:Z	0.01	0.18	-0.09	0.03	-0.07	0
LWRA:Z	-0.03	0.15	0.07	-0.19	0.15	-0.03
LWRB:Z	-0.02	0.15	0.07	-0.2	0.14	-0.05
LFRA:Z	0.03	0.13	0.06	-0.25	0.13	-0.09
LFIN:Z	-0.04	0.16	0.08	-0.16	0.1	0.02
RWRA:Z	-0.04	0.13	0.16	0	0.12	-0.03
RWRB:Z	-0.04	0.13	0.17	0.01	0.12	-0.01
RFIN:Y	0.08	0.13	0.07	0.09	0.07	0.2
RFIN:Z	-0.03	0.13	0.14	-0.04	0.13	-0.02
RBHD:Y	0.13	0.12	-0.03	-0.04	-0.09	0

Figure 8.8: Top Ten Loadings for the Second Principal Component, Performer 2, Prelude 7

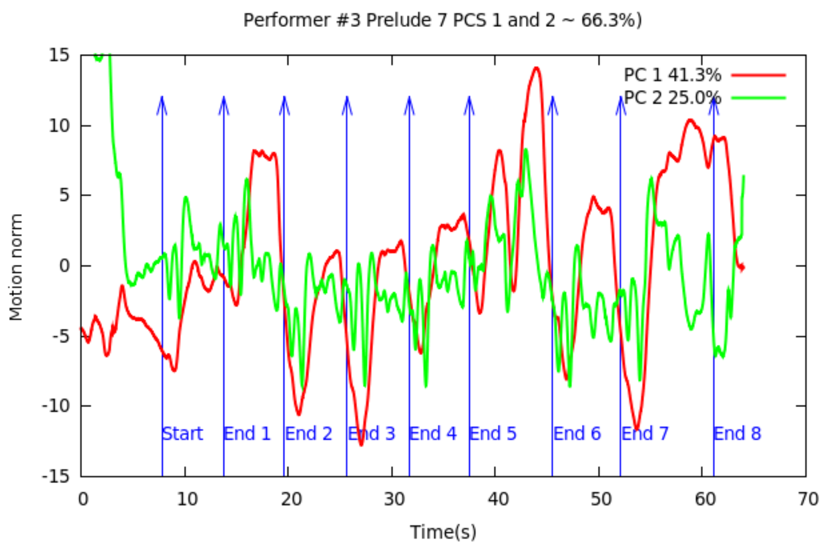


Figure 8.9: First Two Principal Components of Movement for Performer 3, Prelude 7, the first component accounting for 41.3% variance and the second component accounting for 25% variance, with blue vertical lines representing phrasing boundaries as in the audio recording

Marker	PC1	PC2	PC3	PC4	PC5	PC6
C7:Y	0.16	0.01	-0.08	0.02	0.01	0.01
T10:Y	0.16	-0.01	-0.09	-0.02	0.01	0.01
CLAV:Y	0.16	0.03	-0.05	0.03	0	0.01
STRN:Y	0.16	0.03	-0.02	0.02	-0.02	0.02
LSHO:Y	0.16	0.02	-0.05	0.04	-0.01	0
RSHO:Y	0.16	0.01	-0.09	0.02	0.02	-0.01
RUPA:Y	0.17	0	-0.05	0.01	0.03	0.05
RUPB:Y	0.16	-0.01	-0.01	0	0.03	0.12
RUPC:Y	0.16	0	-0.01	0.01	0.05	0.09
RFHD:Y	0.16	0.02	-0.07	0.03	0.01	0.01

Figure 8.10: Top Ten Loadings for the First Principal Component, Performer 3, Prelude 7

Marker	PC1	PC2	PC3	PC4	PC5	PC6
T10:Z	0.01	0.18	-0.09	0.03	-0.07	0
LWRA:Z	-0.03	0.15	0.07	-0.19	0.15	-0.03
LWRB:Z	-0.02	0.15	0.07	-0.2	0.14	-0.05
LFRA:Z	0.03	0.13	0.06	-0.25	0.13	-0.09
LFIN:Z	-0.04	0.16	0.08	-0.16	0.1	0.02
RWRA:Z	-0.04	0.13	0.16	0	0.12	-0.03
RWRB:Z	-0.04	0.13	0.17	0.01	0.12	-0.01
RFIN:Y	0.08	0.13	0.07	0.09	0.07	0.2
RFIN:Z	-0.03	0.13	0.14	-0.04	0.13	-0.02
RBHD:Y	0.13	0.12	-0.03	-0.04	-0.09	0

Figure 8.11: Top Ten Loadings for the Second Principal Component, Performer 3, Prelude 7

within the piece (between end 5 and end 6). The highest loadings for performer 1 as seen in Figures 8.4 and 8.5 relate to movements in the upper arms and the head predominantly in the x axis for the first component, and the chest, upper arms and head predominantly in the y axis.

The second performer's loadings, as seen in Figures 8.7 and 8.8, relate to movements in the upper arms and chest predominantly in the y axis for the first component, and the wrists and fingers predominantly in the z axis for the second component. The first principal component (seen in Figure 8.6) indicates a pattern following the phrasing boundaries, with an exception to this occurring before the end of phrase 6 where the curve is split into two. Suggestions for this occurrence can be found in literature referring to action-chunking [47] where the gesture for a long length of phrase can be split into sections. The second component displays more noise, potentially related to the beats within the phrases. Again the global maximum occurs near the ending of phrase 6 at the harmonic arrival, however the global minimum occurs at the ending of phrase 3.

A similar pattern can be seen for the two principal components of performer 3 (seen in Figure 8.9), where the first component relates highly to phrasing and displays the same split curve in phrase 6, whereas the second component is noisier potentially echoing the inter-phrase beats. The loadings, seen in Figures 8.10 and 8.11, refer to movements in the chest, shoulders and upper arms predominantly in the y axis for the first component, and the head and wrists in both the y and z axes.

In effort to produce a comparable measure of general motion between performers, the addition of the weighted values of the first six principal component scores for each performer produces a motion norm accounting for more than 90% of the variance in movement. The weightings are calculated from the percentage variance of each component over the full dataset. These have been resampled with 10,000 points so that variances in timing between each performance are warped so that results between performers can be directly compared. The distance between each audio phrase boundary is 0.1 and quoted means and standard deviations are calculated for the distances between the peaks of the motion trajectory and its corresponding phrase boundary. These are measured by finding the local maximum for each phrase, using a sliding window. The first three performers' graphs are shown in Figures 8.12, 8.13 and 8.14 whilst the remaining six pianists graphs are included in Appendix B.

Figure 8.12 for Performer 1, at first glance shows no real pattern, however a

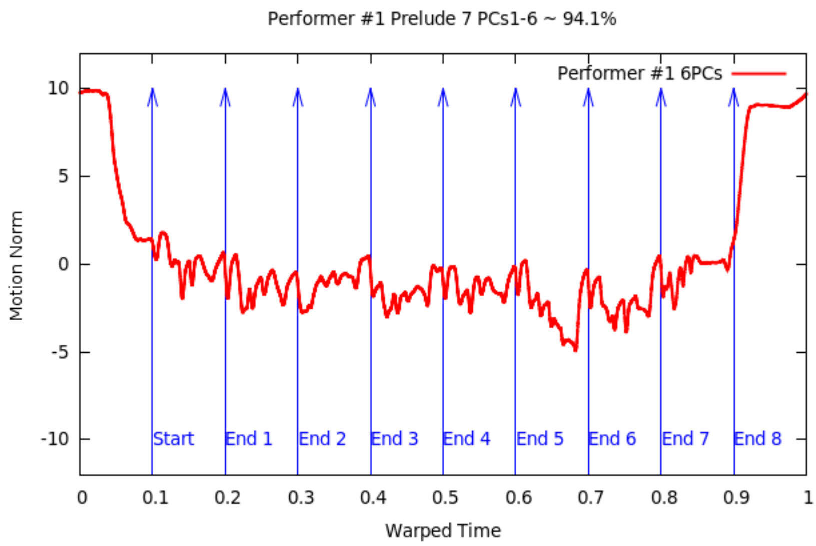


Figure 8.12: Weighted Combination of First Six Principal Components for Performer 1, Prelude 7, accounting for 94.1% variance, plotted in Warped Time

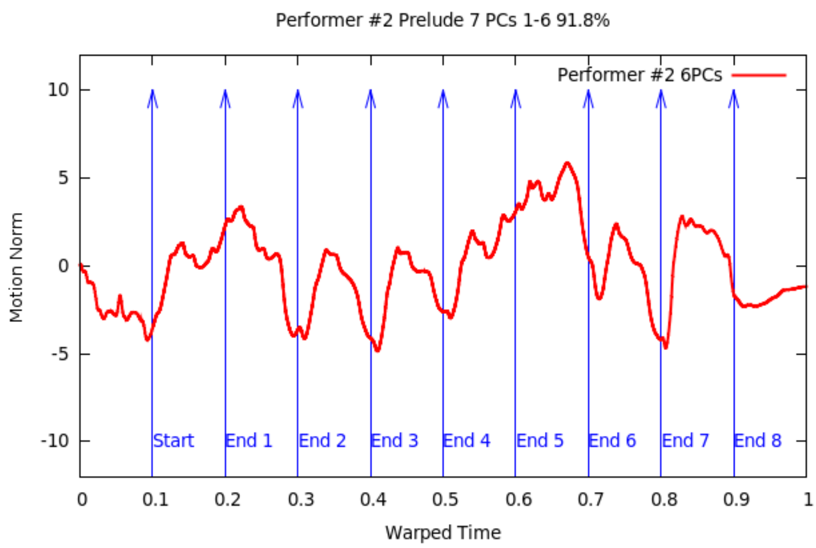


Figure 8.13: Weighted Combination of First Six Principal Components for Performer 2, Prelude 7, accounting for 91.8% variance, plotted in Warped Time

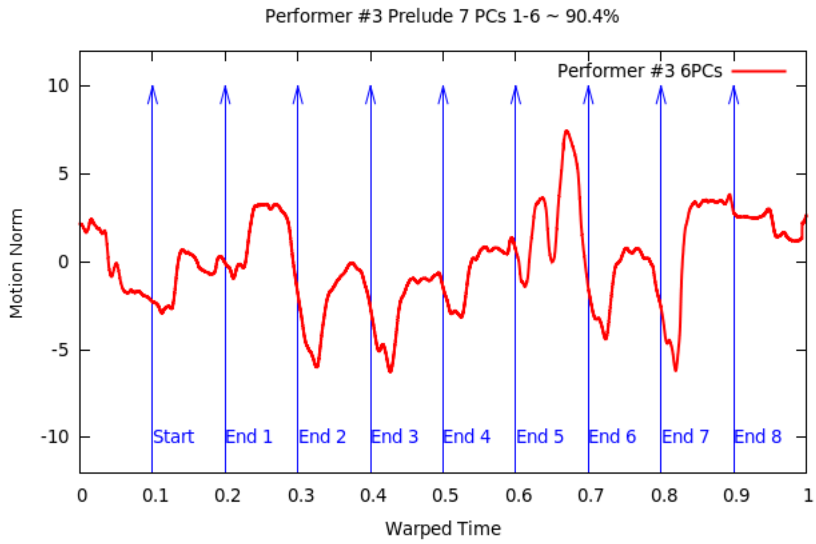


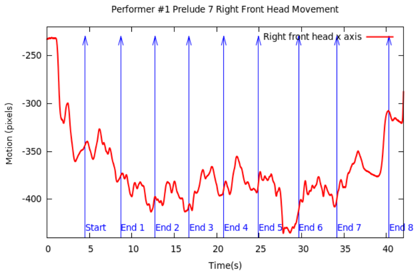
Figure 8.14: Weighted Combination of First Six Principal Components for Performer 3, Prelude 7, accounting for 90.4% variance, plotted in Warped Time

peak always occurs with a phrasing boundary suggesting some underlying pattern (mean = 0.0286, s.d.= 0.026). A large dip occurs at the end of phrase 6, coinciding with the harmonic arrival point. Performer 2's results in Figure 8.13 instead show a very clear pattern of motion with phrasing (mean = -0.046, s.d. = 0.0217). The highest point in the motion norm occurs again at the harmonic arrival point. Finally Figure 8.14 showing results for Performer 3 again shows a clear pattern with the highest point occurring at the end of phrase 6. However, this reflects the split gesture seen in the results of the first two principal components.

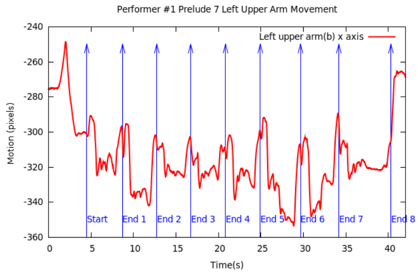
Despite being a good measure of general motion, reductional methods such as PCA can get rid of some of the semantics that singular marker's motion graphs can show. For this reason, the motion of a few particular markers are observed, chosen from those which correlate highest with the first principal components.

The plots for the y axis markers for Performer 1 as seen in Figure 8.15(d), Figure 8.15(e) and Figure 8.15(f) look extremely similar despite being located in different parts of the body. These markers show a trajectory with 8 peaks within the boundaries of the 8 phrases of Prelude 7. The markers plotted for the x axis in Figure 8.15(b) and Figure 8.15(c) show a similar pattern to each other with peaks beginning at each of the phrases. Interestingly the x axis plot for the head marker in Figure 8.15(a) looks entirely different, yet still exhibiting a peak in the motion norm within each of the phrases.

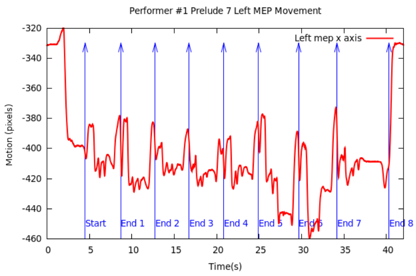
Performer 2's plots of singular markers for the y axis as seen in Figure 8.16(a),



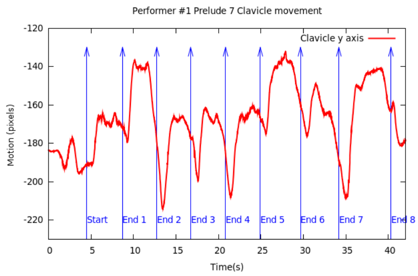
(a) Head (x axis) : RFHD(x)



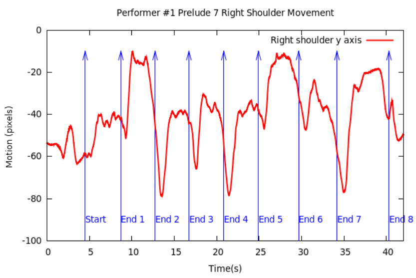
(b) Left upper arm (x axis): LUPB(x)



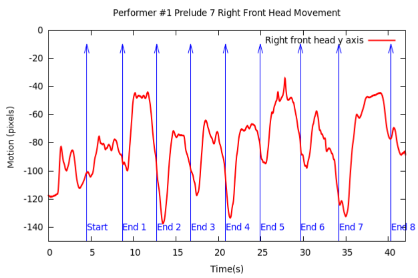
(c) Left mep (x axis): LMEP(x)



(d) Clavicle (y axis): CLAV(y)



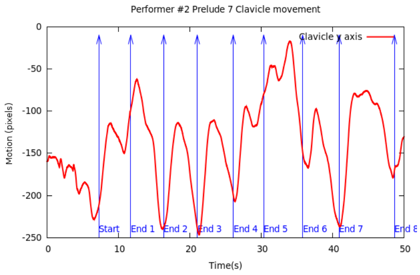
(e) Right shoulder (y axis): RSHO(Y)



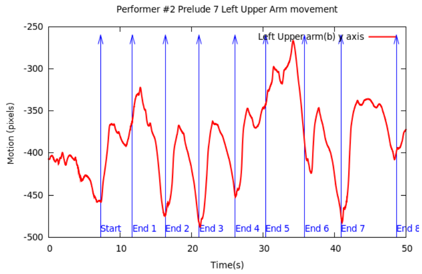
(f) Head (y axis): RFHD(y)

Figure 8.15: Various Raw Marker Data Plotted Against Phrase Boundaries for Performer 1, Prelude 7

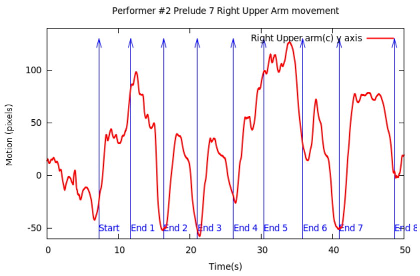
Figure 8.16(b) and Figure 8.16(c) are again remarkably similar to each other, implying a full upper body movement along the *y* axis. One marker from the torso as seen in Figure 8.16(d) plotting the *z* axis movement shows a pattern throughout the 8 phrases albeit not as pronounced as those markers plotted for the *y* axis. The plot for Performer 2's wrist *z* axis as seen in Figure 8.16(e) shows peaks at the start of each phrase, when the left hand plays the first bass note of each phrase and the first chord. The subsequent chords are seen to have not so much of a movement in the *z* axis implying the first two notes are given more stress. The right finger plot in Figure 8.16(f) shows a peak at the end of each phrase, however this is due to the nature of the composition as the performer will need to lift the right hand to prepare for the next phrase.



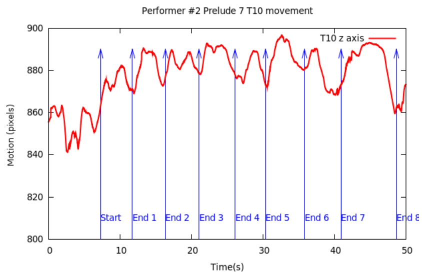
(a) Clavicle (y axis): CLAV(y)



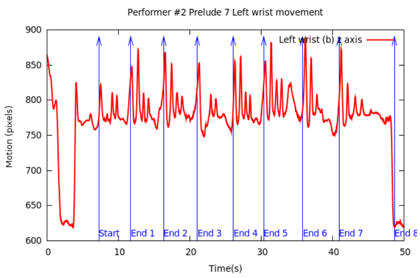
(b) Left upper arm (y axis): LUPB(y)



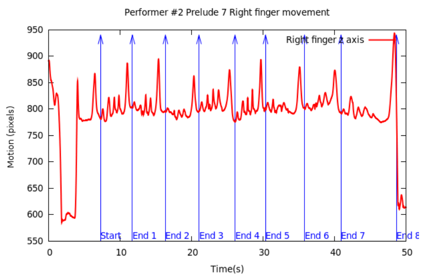
(c) Right upper arm (y axis): RUPC(y)



(d) T10 (z axis): T10(z)



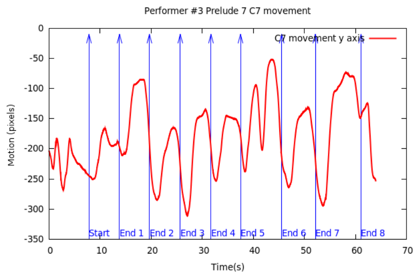
(e) Left wrist (z axis): LWRB(z)



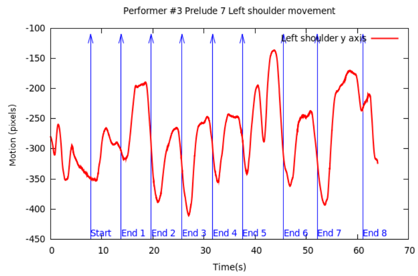
(f) Right finger (z axis): RFIN(z)

Figure 8.16: Various Raw Marker Data Plotted Against Phrase Boundaries for Performer 2, Prelude 7

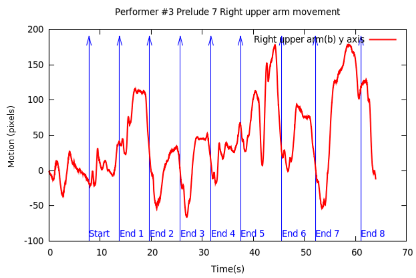
The y axis plots for Figure 8.17(a), Figure 8.17(b) and Figure 8.17(c) for Performer 3 again are remarkably similar in pattern to each other, showing a repeating trajectory for each phrase. The x-axis plots seen in Figure 8.17(d), Figure 8.17(e) and Figure 8.17(f) are not as similar to each other as the y-axis plots but again show patterns for the 8 phrases. Differences at this point lie between the left and right arm markers. This is most likely due to the different rhythms and pitches they are required to play.



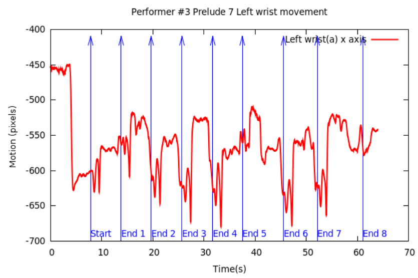
(a) C7 (y axis): C7(y)



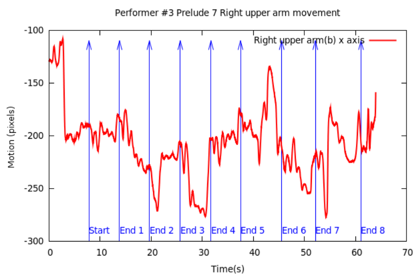
(b) Left shoulder (y axis): LSHO(y)



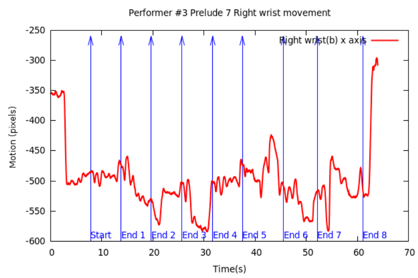
(c) Right upper arm (y axis): RUPB(y)



(d) Left wrist (x axis): LWRA(x)



(e) Right upper arm (x axis): RUPB(x)



(f) Right wrist (x axis): RWRB(x)

Figure 8.17: Various Raw Marker Data Plotted Against Phrase Boundaries for Performer 3, Prelude 7

8.2.2 Prelude in B minor No.6

The initial two-bar motif in prelude 6 is in the left hand melody marked in the score seen in Chapter 7. This motif is varied in the subsequent phrases, first in pitch for the second phrase, then also in rhythm for the third phrase ending at bar 8. Phrase 4 repeats the opening motif and Phrase 5 ends with a modulation into C major. These first five phrases represent an agreement in performers' interpretations and traditional analyses of this prelude. From phrase 6 onwards, performers held diverging views on the structure of the piece. The measured means and standard deviations of distance between motion peak and phrase boundary are therefore taken for the first five phrases only.

Observing Performer 1's results for Prelude 6 (seen in Figure 8.18) and considering the first five phrases, a pattern of phrasing is reflected by the first component. The global maximum occurs at the expansion of the motif in phrase 3 which represents a climax in this particular section. Loadings for performer 1, seen in Figures 8.19 and 8.20 identify correlations in movement of the head, upper arms and chest predominantly in the y axis for the first component and movements of the wrists and fingers predominantly in z axis for the second component.

Performer 2's main loadings seen in Figures 8.22 and 8.23 reflect movements of the upper arms and chest for both the x and y axes for the first component, and the chest, right wrists and fingers for both the y and z axes for the second component. The graph of the two components (seen in Figure 8.21) are highly similar to each other except a slight drag in the second component. An anomaly occurs at the end of phrase 3 where there appears to be an extra peak in the second component. The global maximum can again be seen at the start of the phrase expansion in phrase 3.

As a contrast, the first two principal components for Performer 3, seen in Figure 8.24, appear to be in opposition to each other, yet still in relation with the occurrence of the phrasing boundaries. Again the global maximum is seen at the end of phrase 2, beginning of phrase 3 where the motif is first expanded in rhythm. Loadings can be seen in Figures 8.25 and 8.26 reflecting movements in the head and chest predominantly in the y axis for the first component, and movements in the elbows and wrists predominantly in the x axis for the second component.

When these principal components are combined, into the weighted combination described for in the previous section, we can see clear patterns of phrasing for each of the three pianists examined. These patterns are again repeated for phrases of similar rhythm, although it is interesting to note the differences when compared

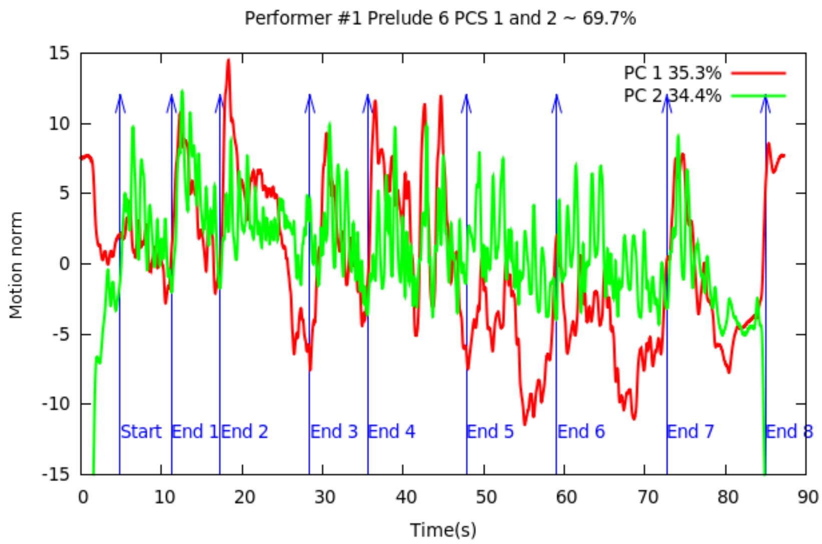


Figure 8.18: First Two Principal Components of Movement for Performer 1, Prelude 6, the first component accounting for 35.3% variance and the second component accounting for 34.4% variance, with blue vertical lines representing the performer's interpretation of phrasing boundaries as in the audio recording

Marker	PC1	PC2	PC3	PC4	PC5	PC6
CLAV:Y	0.17	0.07	-0.02	-0.02	-0.01	-0.04
STRN:Y	0.16	0.07	-0.03	0.05	-0.01	-0.04
LSHO:Y	0.16	0.08	-0.02	-0.04	-0.01	-0.03
LUPA:Y	0.16	0.08	-0.02	0.01	-0.03	0.02
RSHO:Y	0.17	0.07	-0.03	-0.04	-0.02	-0.08
RUPA:Y	0.16	0.08	-0.03	0	-0.01	-0.1
RUPC:Y	0.16	0.07	-0.05	0.02	0.03	-0.09
RFHD:Y	0.16	0.08	0	-0.08	-0.03	-0.06
LFHD:Y	0.17	0.03	0	-0.06	0.06	-0.04
LBHD:Y	0.17	0.06	0	-0.09	0	-0.08

Figure 8.19: Top Ten Loadings for the First Principal Component, Performer 1, Prelude 6

Marker	PC1	PC2	PC3	PC4	PC5	PC6
C7:Z	-0.01	0.15	0.05	0.05	0.2	-0.16
LUPC:Z	0.03	0.15	0.15	-0.09	0.1	0.09
LWRA:Z	-0.01	0.16	0.06	-0.08	-0.13	0.06
LWRB:Z	-0.03	0.16	0.06	-0.11	-0.1	0.07
LFRA:Z	-0.02	0.15	0.12	-0.16	0.06	0.06
LFIN:Z	-0.05	0.16	0.03	-0.07	-0.15	0.07
RWRA:Z	-0.05	0.16	0.05	0.03	-0.13	-0.11
RWRB:Z	-0.06	0.16	0.04	0.04	-0.14	-0.08
RFRA:Z	-0.09	0.14	0.07	0.11	-0.08	-0.06
RFIN:Z	-0.05	0.16	0.02	-0.01	-0.19	0

Figure 8.20: Top Ten Loadings for the Second Principal Component, Performer 1, Prelude 6

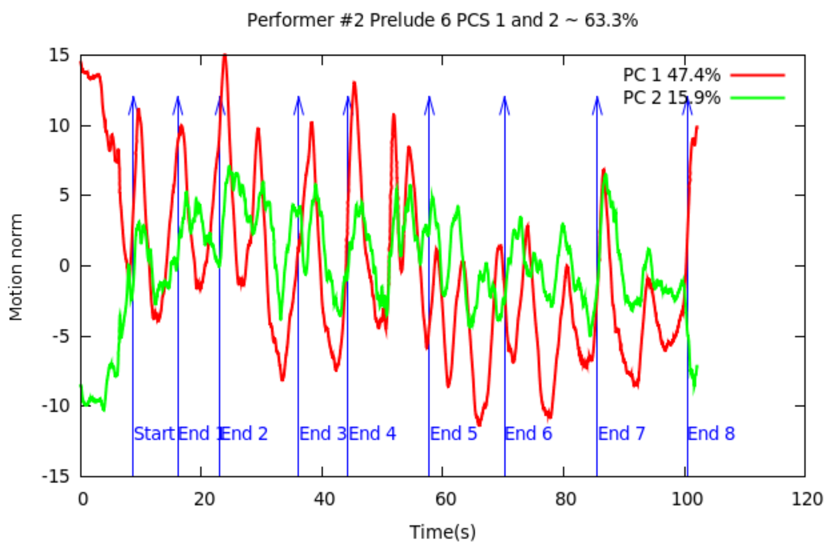


Figure 8.21: First Two Principal Components of Movement for Performer 2, Prelude 6, the first component accounting for 47.4% variance and the second component accounting for 15.9% variance, with blue vertical lines representing the performer's interpretation of phrasing boundaries as in the audio recording

Marker	PC1	PC2	PC3	PC4	PC5	PC6
T10:Y	0.14	0.09	-0.06	-0.02	-0.01	-0.05
CLAV:X	0.14	-0.09	0.09	-0.04	0.04	-0.03
STRN:Y	0.14	0.12	-0.02	0.03	-0.03	0.01
LUPA:Y	0.14	0.11	-0.03	0.06	0	-0.03
LUPB:Y	0.15	0.09	-0.02	0.06	0	-0.04
RSHO:X	0.14	-0.07	0.1	-0.04	0.03	-0.01
RSHO:Y	0.14	0.1	-0.09	-0.03	-0.02	-0.03
RUPA:X	0.14	-0.07	0.08	0.03	-0.03	0.1
RUPA:Y	0.14	0.1	-0.06	-0.02	-0.02	-0.04
RUPC:X	0.14	-0.07	0.06	0.05	-0.05	0.13

Figure 8.22: Top Ten Loadings for the First Principal Component, Performer 2, Prelude 6

Marker	PC1	PC2	PC3	PC4	PC5	PC6
C7:Z	0.08	0.17	0.09	-0.08	0.04	0.16
T10:Z	-0.07	0.2	-0.12	0	-0.04	0.12
RWRA:Y	0.04	0.2	0.07	0.09	0.01	-0.09
RWRA:Z	-0.08	0.19	0.05	-0.08	-0.01	0.05
RWRB:Y	0.04	0.21	0.07	0.08	0	-0.09
RWRB:Z	-0.08	0.19	0.06	-0.08	-0.01	0.04
RFRA:Y	0.09	0.2	0.01	0.04	-0.04	-0.01
RFRA:Z	-0.1	0.17	0.12	-0.06	0.03	0.04
RFIN:Y	0	0.2	0.08	0.06	0.03	-0.1
RFIN:Z	-0.08	0.19	0.05	-0.08	-0.03	0.08

Figure 8.23: Top Ten Loadings for the Second Principal Component, Performer 2, Prelude 6

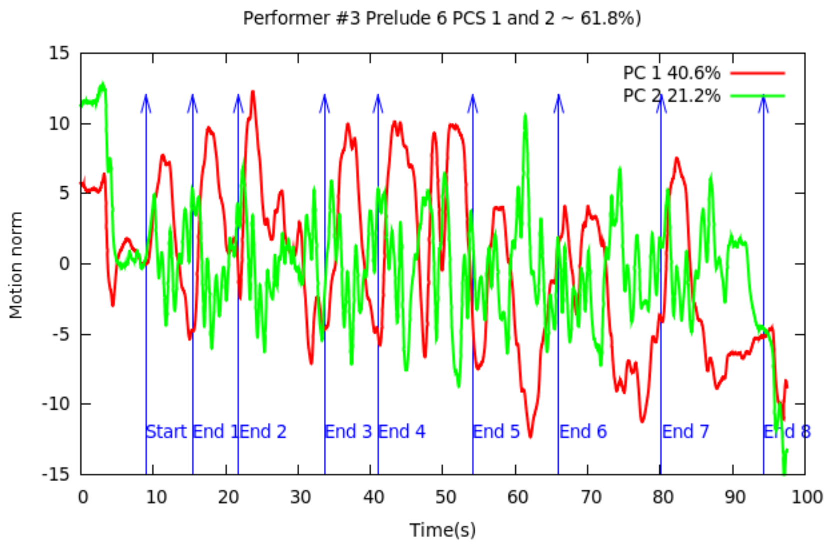


Figure 8.24: First Two Principal Components of Movement for Performer 3, Prelude 6, the first component accounting for 40.6% variance and the second component accounting for 21.2% variance, with blue vertical lines representing the performer's interpretation of phrasing boundaries as in the audio recording

Marker	PC1	PC2	PC3	PC4	PC5	PC6
CLAV:Y	0.16	-0.05	-0.08	0.01	0.01	0.04
STRN:Y	0.16	-0.05	-0.08	0	-0.02	0.03
LSHO:Y	0.16	-0.05	-0.08	0.03	0	0.06
LUPA:Y	0.16	-0.04	-0.08	0	-0.04	0.04
LUPC:Y	0.16	-0.04	-0.09	-0.03	-0.04	0.01
RSHO:Y	0.16	-0.05	-0.07	0.01	0.04	0.06
RFHD:Y	0.16	-0.06	-0.06	0.02	0.05	0.07
LFHD:Y	0.15	-0.06	-0.06	0.02	0.03	0.09
LBHD:Y	0.16	-0.05	-0.07	0.02	0.06	0.06
RBHD:Y	0.16	-0.05	-0.06	0.03	0.08	0.05

Figure 8.25: Top Ten Loadings for the First Principal Component, Performer 3, Prelude 6

Marker	PC1	PC2	PC3	PC4	PC5	PC6
LELB:X	0.09	0.11	0.04	-0.26	0.19	0.01
LWRA:X	0.12	0.11	0.02	-0.2	-0.04	-0.05
LWRB:X	0.11	0.12	0.02	-0.23	-0.01	-0.01
LFRA:X	0.1	0.11	0.04	-0.25	0.15	-0.01
RELB:X	0.07	0.15	0.13	0.13	-0.07	-0.17
RMEP:X	0.07	0.14	0.14	0.13	-0.05	-0.19
RWRA:X	0.09	0.14	0.1	0.13	-0.1	-0.15
RWRB:X	0.08	0.16	0.1	0.11	-0.14	-0.12
RFRA:X	0.07	0.15	0.12	0.12	-0.09	-0.16
RFIN:X	0.09	0.16	0.09	0.12	-0.14	-0.13

Figure 8.26: Top Ten Loadings for the Second Principal Component, Performer 3, Prelude 6

against the expanded rhythm in phrase 3.

Now considering the weighted combinations of principal components, for performer 1 (as seen in Figure 8.27), clear, repeatable, patterns are observed for the first and fourth phrases (mean = -0.0034, s.d = 0.0062), with the fifth phrase split in the middle roughly where the modulation into C major occurs. Performer 2's weighted combination as seen in Figure 8.28, displays a pattern of motion in the first three phrases (mean = -0.0294, s.d = 0.0204). Interestingly, in phrase 3 where the original two-bar motif is expanded, we clearly see two separate movements. As the length of the phrase being performed is just under 12 seconds long, this may relate to the theory of gestures being separated into gesture-units i.e. action-chunking [47]. At which points within a long phrase this action-chunking occurs is most likely related to the smaller rhythmical groupings within the particular phrase. The next two phrases are again split into two sections which corresponds with the performer's own interpretation of the piece. Performer 3's weighted combination as seen in Figure 8.29, shows a repeatable pattern in the first four phrases (mean = -0.0107, s.d = 0.0101). Again we can see a split in the movement occurring during phrase 5, most likely relating to a separate gesture when the modulation into C major occurs.

Examining the differences between each performer's motion profile between their performances of Prelude 7 and Prelude 6, correlations for the first two phrases for each prelude are calculated. The time-adjusted, weighted combinations of principal components for phrase 1 and phrase 2 of Prelude 7 are correlated against the same measurements in phrase 1 and phrase 2 of Prelude 6. This is due to the

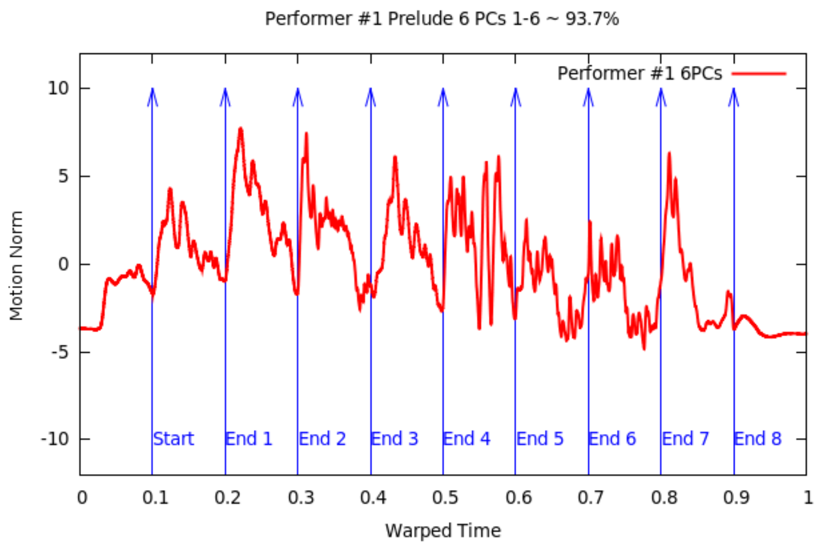


Figure 8.27: Weighted Combination of First Six Principal Components for Performer 1, Prelude 6, accounting for 93.7% variance, plotted in Warped Time

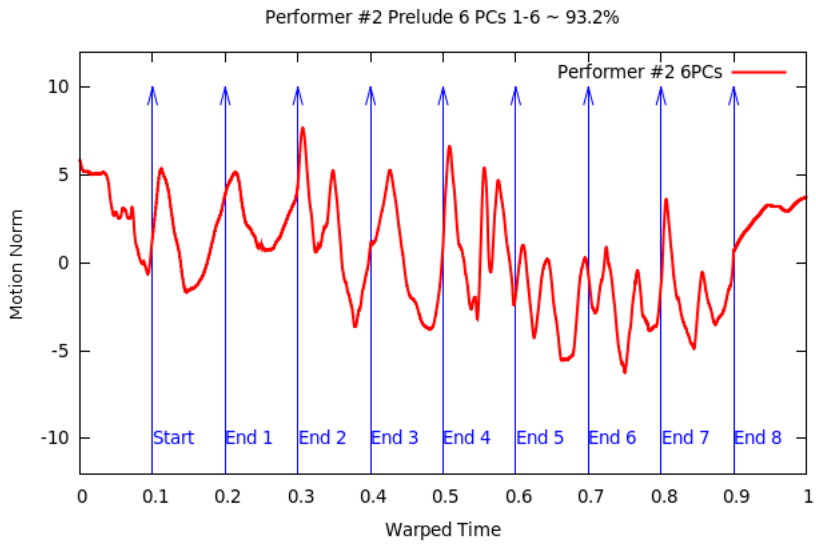


Figure 8.28: Weighted Combination of First Six Principal Components for Performer 2, Prelude 6, accounting for 93.2% variance, plotted in Warped Time

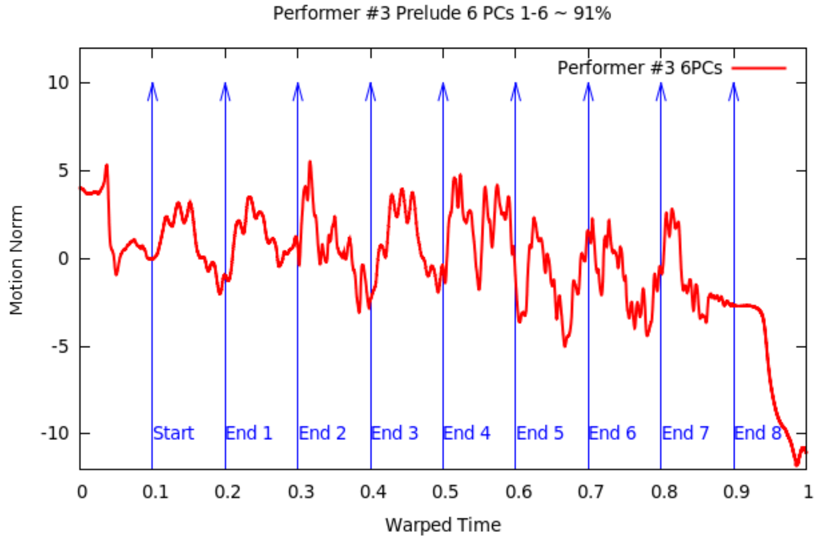


Figure 8.29: Weighted Combination of First Six Principal Components for Performer 3, Prelude 6, accounting for 91% variance, plotted in Warped Time

Table 8.1: Correlations of Performer Motion Profile across Preludes, Results Printed for P<0.01

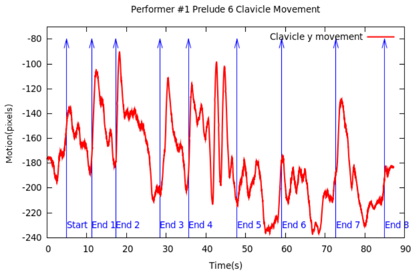
Phrase \ Performer	P1	P2	P3	P4	P5	P6	P7	P8	P9
Phrase One	-0.8154	0.7338	0.8755	0.2246	0.4513	-0.015	0.2978[1]	0.6946	0.3792
Phrase Two	-0.0534[1]	-0.4187	-0.0687[2]	-0.0451[1]	0.3695	0.4210	0.6556	-0.1219	-0.8681

1. Not significant
2. Significant to p<0.05

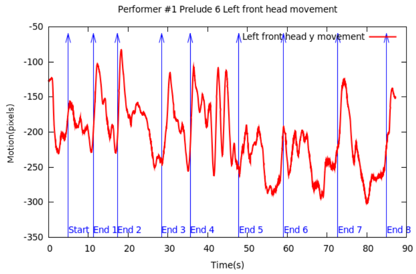
nature of these phrases, both sets being two bars in length, and phrase 2 being a rhythmic replica of phrase 1 in each prelude with changes solely in melody and harmony. The results of this are shown in Table 8.1. Despite some correlations showing results above 0.8 with a significance of $p<0.01$, this is not repeated for the correlation for the same performer in the next phrase of each prelude. Other correlations are either extremely low or not significant. From this we can reject Hypothesis 8.2 as performers' motion profiles appear to differ depending on what piece they are performing.

As principal components analysis is useful in reducing the number of dimensions of data but often loses the semantics of what the condensed data actually represents, it is advantageous to examine the components of each marker trajectory separately so as to better understand their semantics.

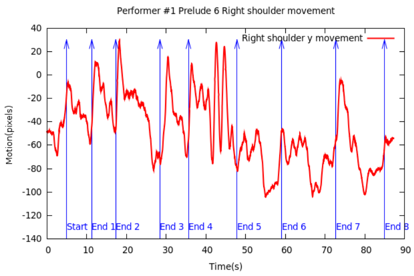
The three plots for y axis markers for Performer 1 seen in Figure 8.30 show a



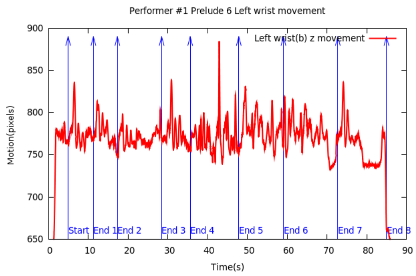
(a) Clavicle (y axis): CLAV(y)



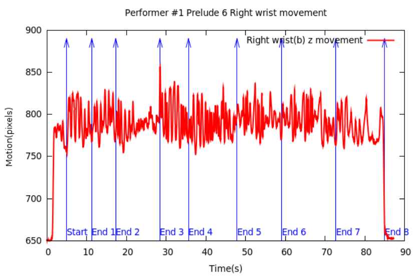
(b) Head (y axis): LFHD(y)



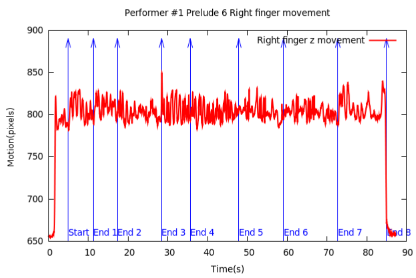
(c) Right shoulder (y axis): RSHO(y)



(d) Left wrist (z axis): LWRB(z)



(e) Right wrist (z axis): RWRB(z)



(f) Right finger (z axis): RFIN(z)

Figure 8.30: Various Raw Marker Data Plotted Against Phrase Boundaries for Performer 1, Prelude 6

similar pattern to each other corresponding to the principal components results discussed previously. The plots for the z axis markers all report the activity of markers on the wrists or fingers and so the data recorded is noisier than for markers further away from the actual keys of the piano. These graphs however, still show a pattern within each phrase.

Performer 2's plots for the y axis as seen in Figure 8.31(b) and Figure 8.31(c) show a similar pattern to each other which repeats over the 8 marked phrases. The graph for x axis movement in the upper body as seen in Figure 8.31(a) shows peaks occurring at phrase boundaries which is again similar for the z axis movement seen in Figure 8.31(d). The remaining two graphs plotting the movement of the right wrist in the y and z axis despite being closer to the movements needed for note production, do show certain patterns that can be attributed to producing a phrasing contour.

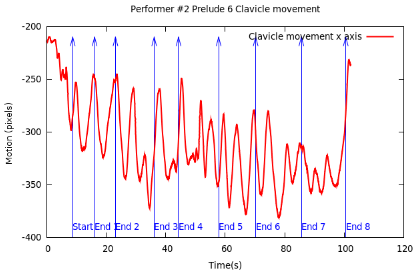
Performer 3's graphs shown in Figure 8.32 again show the markers for the upper body in the y axis moving simultaneously in the same direction which is similar to the principal components motion norm displayed previously. The x axis movement of the right elbow seen in Figure 8.32(e), and Figure 8.32(f) are reflective of each other and show patterns involving peaks in the trajectories at the beginning of phrases. The x axis of the left elbow seen in Figure 8.32(d) does not show entirely a clear pattern of phrasing but places peaks in the trajectory at certain points in the music, notably at the beginning of phrase 3, being the climax of the first section of this piece with the highest pitch repetition of the original two-bar motif and expansion into four bars.

8.2.3 Conclusions

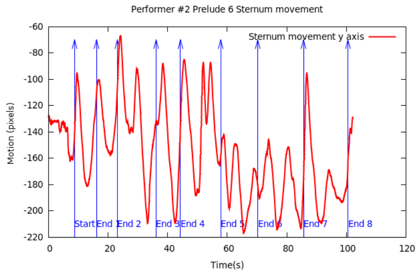
By examining movement of nine performers across two Chopin Preludes, it is demonstrated that each pianist's movement is entirely subjective and personal. No two performers appear to move in exactly the same way for any one piece of music. However, there appears to be an underlying pattern within these gestures that relate to phrasing structure. The results from the principal components analysis for Prelude 7 show clear patterns between the calculated motion norm and the phrasing boundaries indicating that hypothesis 8.1 is correct. Local maxima in the motion norm are consistent across phrases in their distance from the phrasing boundary suggesting that with repeated phrases, performers will reliably produce the same overall motion. This is reflected by the trajectories shown by plotting the raw marker data for the highest correlated markers indicated by the loadings. The loadings for each performer show that the movement cannot be attributed to any singular marker but instead a combination of many from different parts of the body. Marker trajectories particularly for the y axis (along the length of the keyboard) reflect the phrases dictated by traditional analysis. Also we see that markers in the head, upper torso and shoulders tend to reflect the phrasing structure more clearly, whereas markers getting closer to the elbows and wrists will show the beats of each performed note, due to the necessary gestures required for actual note production. Interestingly, for each performer, their loadings do not stay consistent between pieces of music. Their calculated motion norm trajectories as well as the trajectories for the raw marker data are also different between pieces, suggesting that gesture is not used in the same way across pieces, but may have qualities influenced by rhythm and pitch. This rejects hypothesis 8.2. The identical rhythm in the phrases of Prelude 7 helps highlight gestures being produced in a situation where rhythm is controlled. Despite the identical nature of these phrases in rhythm, each performer's gesture for each phrase is not entirely identical suggesting that variables such as pitch and harmony contribute to gesture production.

Examining performers' gestures for Prelude 6 up until the end of phrase 5, we again see patterns developing between phrases with some slight differences, particularly at the expansion of the original motif in phrase 3. Some pianists expand their gesture to cover the entire phrase whereas some are producing almost two peaks within a gesture, so sub-chunking the movement.

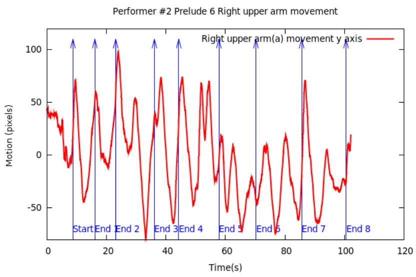
Overall gesture appears to be a good identifier of phrasing structure across these two pieces despite the pattern within each performer not being consistent. It will now be examined how aural parameters contribute to the phrasing contour of the piece and how these interact with gesture at important points in the structure.



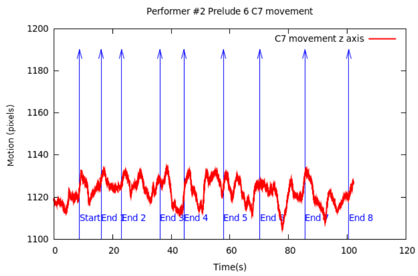
(a) Clavicle (x axis): CLAV(y)



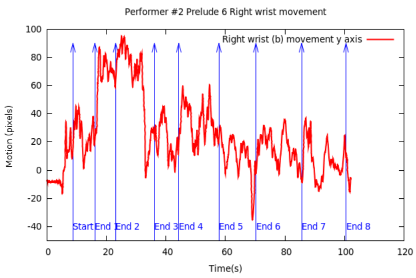
(b) Sternum (y axis): STRN(y)



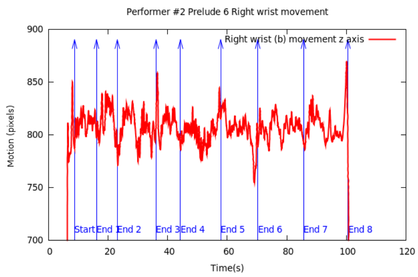
(c) Right upper arm (y axis): RUPA(y)



(d) C7 (z axis): C7(z)

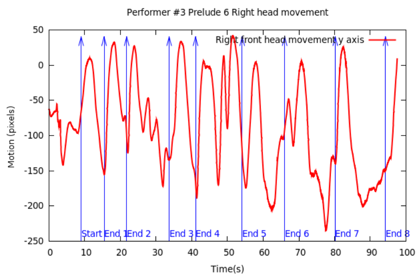


(e) Right wrist (y axis): RWRB(y)

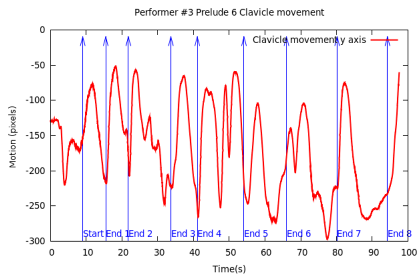


(f) Right wrist (z axis): RWRB(z)

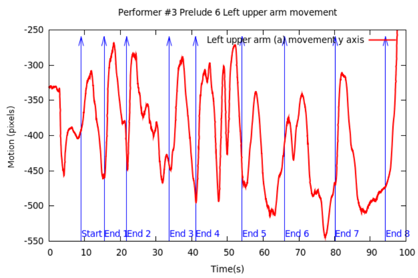
Figure 8.31: Various Raw Marker Data Plotted Against Phrase Boundaries for Performer 2, Prelude 6



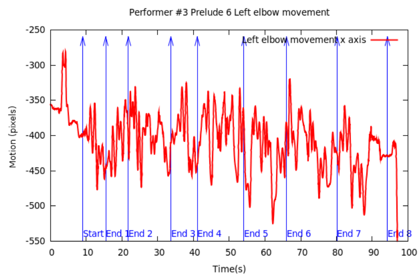
(a) Head (y axis): RFHD(y)



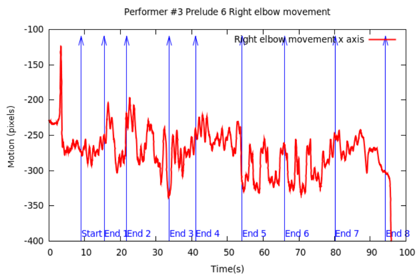
(b) Clavicle (y axis): CLAV(y)



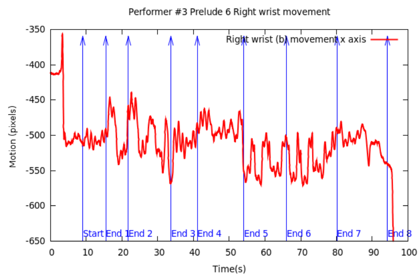
(c) Left upper arm (y axis): LUPA(y)



(d) Left elbow (x axis): LELB(x)



(e) Right elbow (x axis): RELB(x)



(f) Right wrist (x axis): RWRB(x)

Figure 8.32: Various Raw Marker Data Plotted Against Phrase Boundaries for Performer 3, Prelude 6

8.3 Multi-Modal Analysis

Exploring how aural parameters work with visual parameters to convey structure in musical performances, two parameters taken from the audio and MIDI data are examined. The movement parameter is taken from the weighted combinations of principal components describing the overall body movement, explored in Chapter 8.2. To estimate tempo, the MIDI notes were matched to the MusicXML score notes in the processes involved in creating Performance Markup Language files (seen in Section 6.1.1). These files were uploaded to the database (described in Section 6.2) and queried for the calculation of inter-onset intervals(IOIs) for each matched note. Each of these values were normalised to a crotchet beat and divided by 1/60 to give an estimation of beats per minute. Outliers in tempo for specific notes were removed due to habits of performers when performing semiquavers following dotted quavers. The semiquaver part of this pair of notes tended to be highly elongated in comparison to the other notes and was considered to be a stylistic point. For this reason, these particular semiquavers were removed from Prelude 7 and Prelude 6 from the calculations of tempo. Dynamics, or loudness, was estimated by calculating the RMS amplitude of the audio signal using a short Python script.

Again three of the nine pianists are taken as examples to examine the spread of data. Each performer's audio, MIDI and video data is plotted against the phrase boundaries as they occur in the audio stream. Tempo estimations are plotted at the note onset of the first of the pair of notes used for calculating the inter-onset interval. The graphs of the remaining six pianists are seen in Appendix C.

8.3.1 Prelude No.7 in A major

Examining parameters of tempo and loudness against the previously analysed motion norms for each performer we can view interactions between the aural and visual parameters.

Performer 1's multi-modal graph of Prelude 7, as seen in Figure 8.33, shows slight peaks in tempo (representing an acceleration) after each phrasing boundary and dips in rms amplitude (representing a diminuendo) just before the boundary. This loudness measure will be largely affected by the makeup of each phrase which ends with a minim. As the piano plays the chord, the measure of rms amplitude will reduce exponentially. However, the shape of the rms envelope throughout the phrase will be manipulated by the performer. Particular points to note in

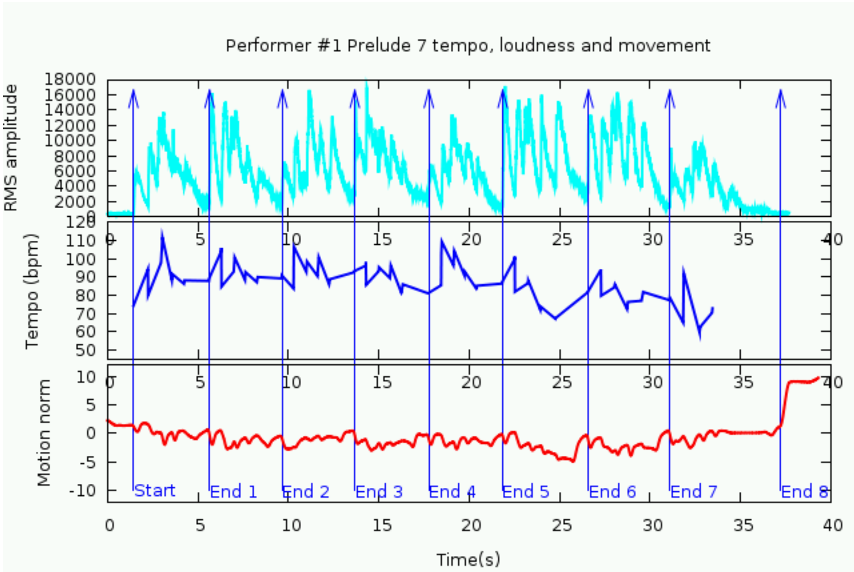


Figure 8.33: Motion, Tempo and Dynamics for Performer 1, Prelude 7 with blue vertical lines representing phrasing boundaries as noted in the recorded audio

this performance are during phrase 6 where the harmonic arrival occurs (between end 5 and end 6 on the graph) where the tempo measured reaches a global minimum and the rms amplitude instead of reducing throughout the phrase stays at a constant level. This also aligns with a global minimum in the motion norm.

Performer 2's multi-modal parameters as seen in Figure 8.34 highlights particular points in the piece such as the halfway point at the end of phrase 4 where we see a global maximum in the tempo calculation. This occurs directly after a large dip in loudness. Another point of interest occurs at the global maximum in the motion norm at the end of phrase 6 at the harmonic arrival, which corresponds with a global minimum in tempo.

Performer 3's graph seen in Figure 8.35 again shows a global minimum in tempo occurring alongside a global maximum in motion at the harmonic arrival in phrase 6. Looking at the whole graph we can also see a reflection of the tempo curve in the motion norm.

Observing these traits across all performers, a direct comparison can be taken by warping each stream of data with respect to the occurrence of phrase boundaries in the audio stream. Distances to the local minima for dynamics and tempo curves were extracted for each phrase boundary. Distances to the local maxima were extracted for the motion curves. Two-way ANOVAs showed significant effects of performer on motion norm ($F=12.07$, $p<0.001$), a significant effect of performer on dynamics ($F=6.26$, $p<0.05$) and of phrase number on tempo ($F=11.43$,

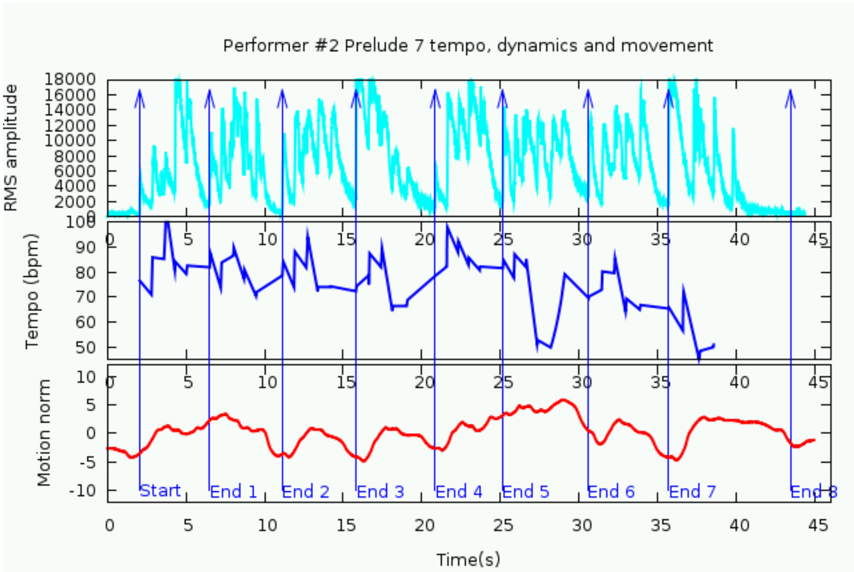


Figure 8.34: Motion, Tempo and Dynamics for Performer 2, Prelude 7 with blue vertical lines representing phrasing boundaries as noted in the recorded audio

p<0.001). Other effects were not significant. This suggests that performers have a distinct style of motion and diminuendo. Although they may not vary in their use of ritardando, this is varied between phrases.

Observing these three graphs we can see patterns of motion, tempo and loudness which occur within each phrase and extremes of these datasets corresponding to points of interest within the piece such as the end of a section or a particular point of notice in harmony. To discern whether the extremes in the measured parameters correspond to important points in musical structure, the measurements are sampled at the point of each note onset for the previously calculated IOI, motion norm and rms amplitude.

Box-plots showing the spread of data for each parameter can be seen in Figure 8.36 for tempo, motion norm and rms amplitude respectively. Measurements for tempo and rms amplitude and motion are normalised between 0 and 1 for each performer. Each box-plot shows a red line for the median of the data, and the surrounding box shows the first and third quartiles. The extremes of the data not considered to be outliers are identified by the whiskers of each box, with the outliers marked as red crosses. From these box-plots we can also view the preferences or style of each performer in their use of tempo, dynamics and motion. A thin box with many outliers suggests that the performer uses a very small range of a certain parameter throughout the majority of the piece, reserving the extremes for a few specific points. A large box covering most of the data range suggests that

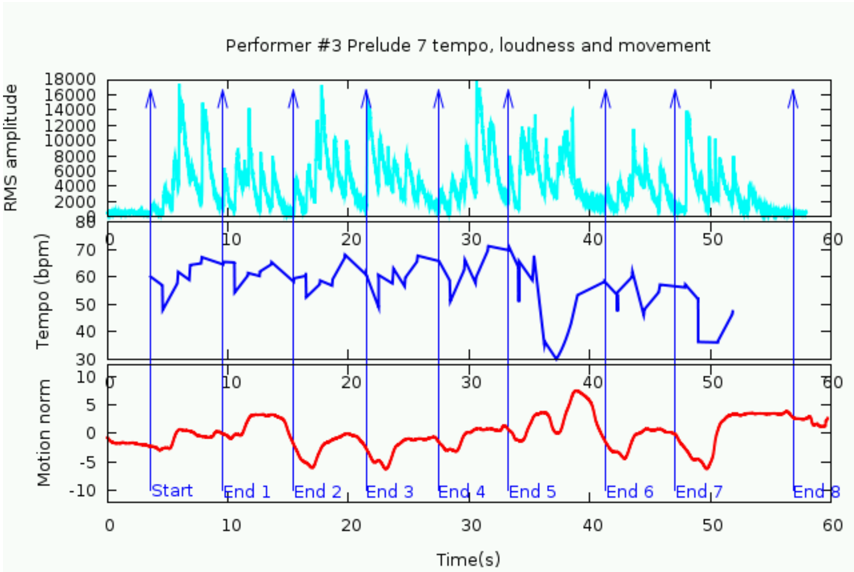


Figure 8.35: Motion, Tempo and Dynamics for Performer 3, Prelude 7 with blue vertical lines representing phrasing boundaries as noted in the recorded audio

the performer uses a larger spread of the parameter throughout the piece.

These box-plots show the spread of each parameter for each performer to be very different to each other demonstrating that each pianist has a particular style of expressing the notes of the piece. Also as not every box suggests a normal distribution, there appear to be underlying patterns skewed to certain values. The first two performers show a fairly normal distribution for motion and dynamics whereas the tempo is slightly skewed. The third performer shows skewed values for all three parameters.

Calculating the 5th and 95th percentiles for each parameter for each performer, the extremes of the filtered data below the 5th and above the 95th percentile are extracted and compared against their occurrence within the score. Hypothesis 8.4 states that the extremes in tempo, dynamics and motion are where the most important notes of the piece occur. For each of the three example performers, in Figures 8.37, 8.39 and 8.41, a scatter plot exhibits the spread of extracted data for each of the three parameters as red crosses.

The extremes of the motion data are plotted over the top of the dataset as blue markers, the extremes of the tempo data as green markers and the extremes of the dynamics data as pink markers. These are translated onto a score of the piece with corresponding colours, as seen in Figures 8.38, 8.40 and 8.42.

From the scatter plot of data for Performer 1 seen in Figure 8.37, we can see that most data points lie in a cluster in the middle of the graph, however, the few

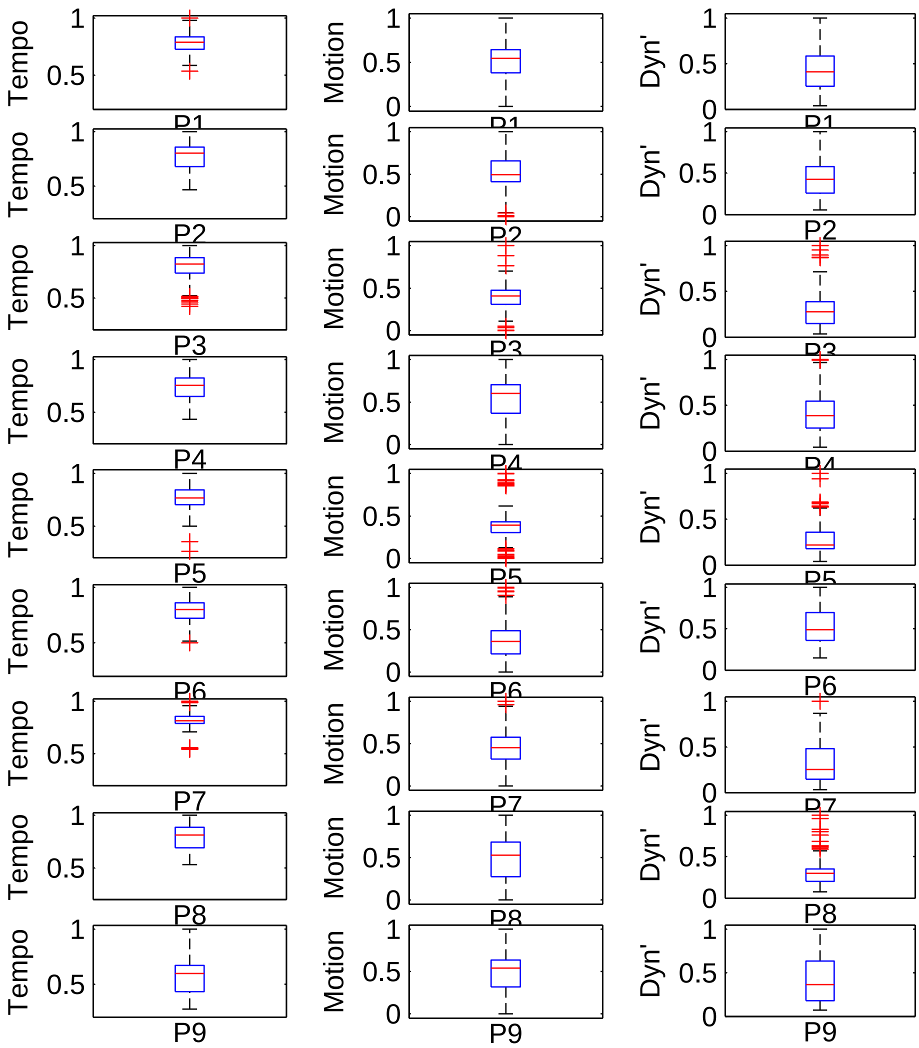


Figure 8.36: Box-plots for all Nine Performers measuring Tempo, Motion norm and Dynamics used in Performances of Prelude 7

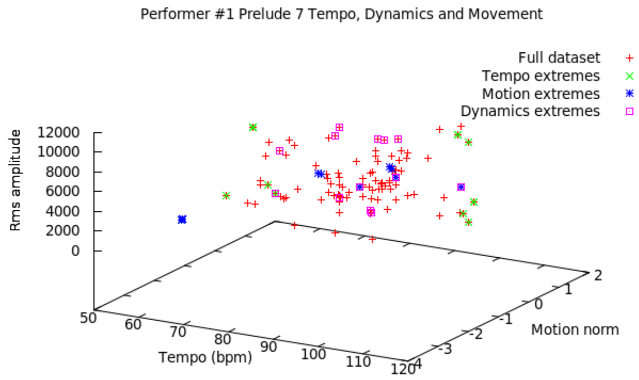


Figure 8.37: Scatter Plot Showing Extremes in Tempo, Dynamics and Motion for Performer 1, Prelude 7

outliers identified by the pink, green and blue markers indicate particular places of interest. Some points show duplicates of extremes, with a blue marker occurring at the same place as a pink box, showing a point where the motion has been varied to a global maximum or minimum at the same point where dynamics have been varied to a global maximum or minimum. Another point to note is clusters such as the maxima in tempo denoted by green markers to the right hand side of the scatter plot, which seem to occur with high values in motion norm, suggesting that the pianist ties in fast tempi with higher values of their motion profile. To see how these extremes lie on top of the structural boundaries of the music, these maxima and minima are plotted on top of the original score. From the translated score image identifying the outliers in each of the parameters for performer 1 in Figure 8.38, we can see points of interest particularly at the beginning of the piece with shows combinations of extremes from motion and dynamics. Also at the harmonic arrival we can see extremes of tempo and dynamics leading up to the end of the phrase in bar 12, which is also characterised by extremes in the motion norm. The end of the piece also sees a combination of parameters in their extremes and bar 9 which marks the beginning of the second half of the piece with a repetition of the original phrase sees extremes in both tempo and dynamics. Particular notes within phrases being accented by extremes of these parameters include the second beat of bars 1, 3, 9, 11, 13 and 15. These correspond to melodic accents within each phrase as marked out in Parncutt's theory of accents in piano performance [91] (see also Chapter 7). Performer 1's particular accents correspond to the first pair of phrases, the first phrase of the second section, the harmonic arrival and the last pair of phrases.

Music engraving by LilyPond 2.12.2—www.lilypond.org

Figure 8.38: Annotated Score for Performer 1, Prelude 7 noting extremes in tempo(T), dynamics(D) and motion(M)

Observing the scatter plot for Performer 2's data during their performance of Prelude 7 in Figure 8.39, we see the extremes in dynamics denoted by pink marks, lie within the mid range of tempo values, something that challenges the general theory that faster tempi more often than not result in higher dynamics. For the other extremes, we see a spread of data, however there are quite a few points where an extreme in motion (blue) coincides with an extreme in dynamics(pink). For Performer 2's score plot in Figure 8.40, we can see these combinations of extremes occurring at the end of section 1 in the fourth phrase at bar 7 and the beginning of section 2 in the fifth phrase at bar 9. Again the harmonic arrival and the end of the piece are characterised by extremes in motion and tempo and extremes in all three parameters respectively. Another point of interest is noted at end of phrase 2 and beginning of phrase 3 at bar 4. This is marked by an extreme in dy-

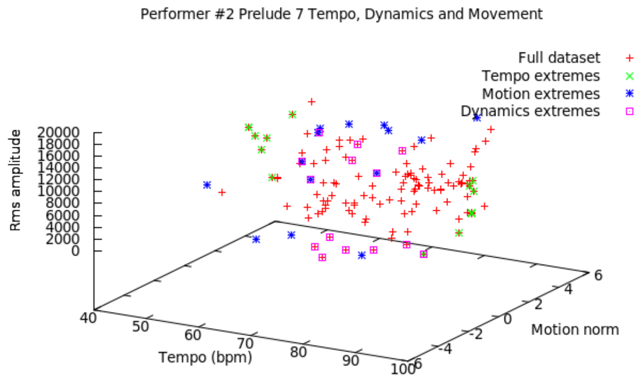


Figure 8.39: Scatter Plot Showing Extremes in Tempo, Dynamics and Motion for Performer 2, Prelude 7

namics followed by an extreme in motion. Where each extreme is located on the singular note-level, Performer 2's accents fall on the grouping and metrical accents suggested by Parncutt and so tend to fall on the first and last notes of each phrase for grouping as well as the first note in the bass for each phrase suggesting a more rhythmical accent.

The scatter plot of data for Performer 3 as seen in Figure 8.41, shows a large cluster of maxima and minima in dynamics, tempo and motion, occurring on the left hand side of the plot, where tempo values are at their lowest. Combinations of extremes are seen in all three data streams occurring with these minima in tempo. What is interesting to consider is whether these occur at several points within the performance, or are located within one particular phrase. Performer 3's score plot in Figure 8.42 shows less use of extremes at several points in the piece and instead shows a large cluster of points around phrase 6 which includes the harmonic arrival at bar 12 marked by extremes in all three parameters. The beginning of phrase 3 at bar 5 is also marked by extremes in motion and dynamics. Singular parameter extremes feature at the beginning of the piece for dynamics and the end of the first section at phrase 4 in bar 7 for motion. Again, Performer 3's extremities tend to fall on the first and last notes of the phrase on grouping accents as well as the first beat of the bass for the metrical accent. The exception for this appears to be at the phrase containing the harmonic arrival at bar 12 where the performer uses combinations of all three parameters to emphasise this feature.

Although these three performers are entirely different in their use of aural and visual parameters, they tend to mark out similar sections with extremes in measurement, particularly the harmonic arrival (considered the climax of the piece),

Figure 8.40: Annotated Score for Performer 2, Prelude 7 noting extremes in tempo(T), dynamics(D) and motion(M)

the beginning and end of the piece. Some performers also pick out the halfway point of the piece at the end of phrase 4. These most important structural features tend to be characterised by a combination of extremes in the aural and visual stream.

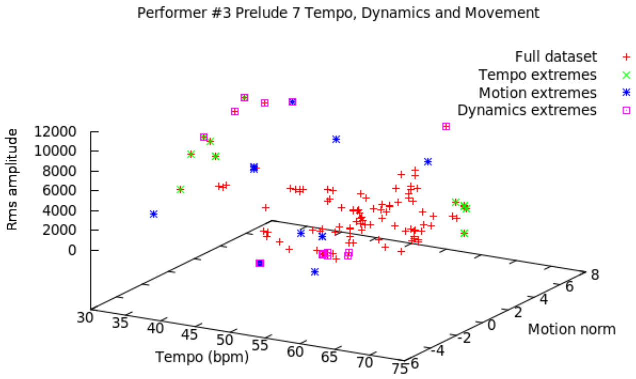


Figure 8.41: Scatter Plot Showing Extremes in Tempo, Dynamics and Motion for Performer 3, Prelude 7

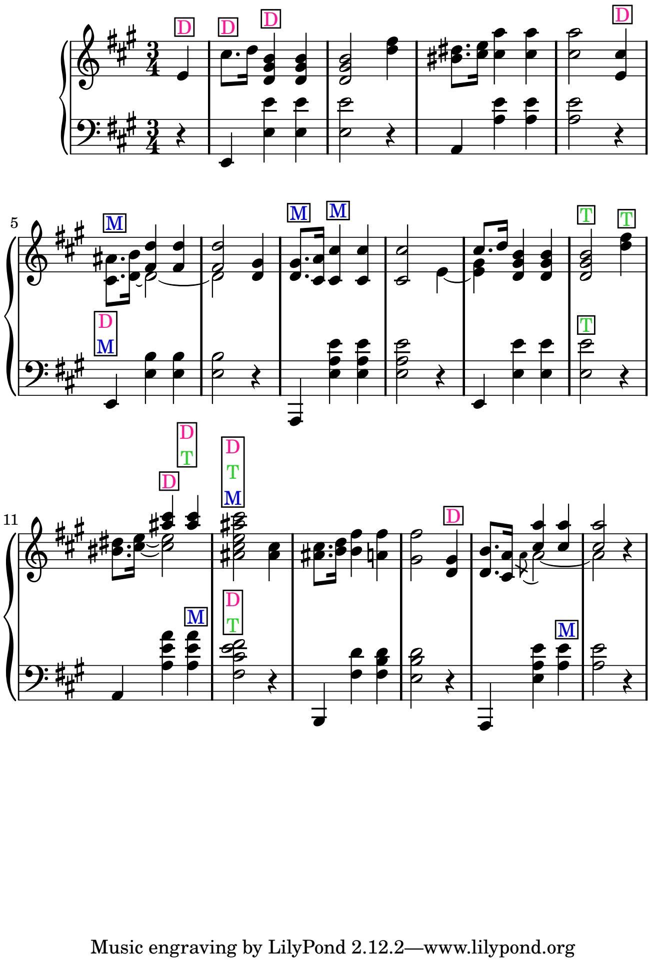


Figure 8.42: Annotated Score for Performer 3, Prelude 7 noting extremes in tempo(T), dynamics(D) and motion(M)

8.3.2 Prelude No.6 in B minor

Again for this Prelude as in Section 8.2.2 the analysis will refer solely to the first five phrases as marked out in traditional analysis in Chapter 7. These five phrases represent an agreement among performers with an added split in the middle of phrase 5 where the piece modulates into C major. The multi-modal graphs for each performer will show their own interpretation of phrasing boundaries marked out by a blue vertical line identifying their occurrence in the audio stream.

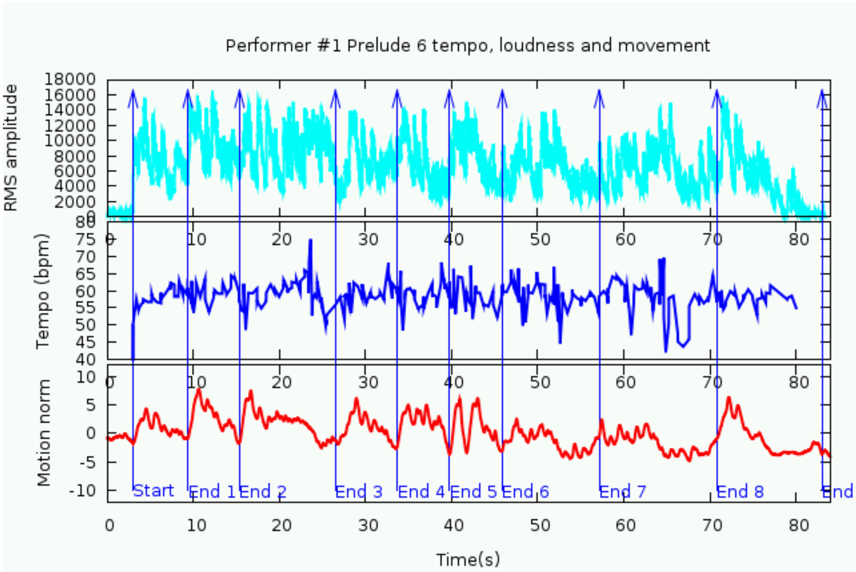


Figure 8.43: Motion, Tempo and Dynamics for Performer 1, Prelude 6 with blue vertical lines representing the performer's interpretation of phrasing boundaries as in the recorded audio

Performer 1's graph for multi-modal parameters seen in Figure 8.43. Particular points of interest include the global maximum in tempo at the end of phrase 3 followed by a local minimum which marks the end of the first section of the piece. The motion trajectory as analysed in the previous section shows distinct patterns between phrases.

Global maxima in the motion norm for Performer 2's multi-modal graph seen in Figure 8.44 appear to correspond to the global maxima in tempo for each phrase. This is also reflected in the rms amplitude measurement.

Performer 3's graph of multi-modal parameters seen in Figure 8.45 shows a peak within phrase 3 in the motion norm which is echoed in the dynamics and tempo measurements.

Again these streams of data are re-sampled with a time-warping algorithm which takes the phrasing boundaries into account. Two-way ANOVAs performed

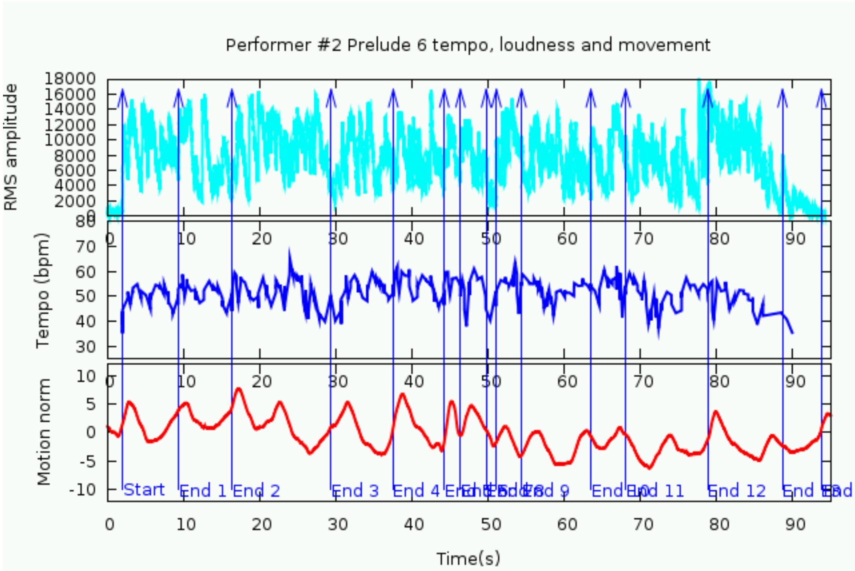


Figure 8.44: Motion, Tempo and Dynamics for Performer 2, Prelude 6 with blue vertical lines representing the performer's interpretation of phrasing boundaries as in the recorded audio

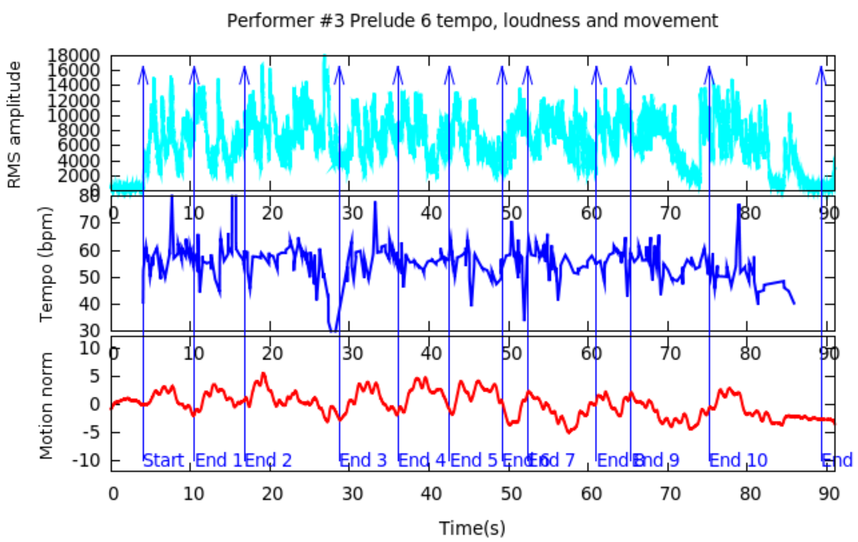


Figure 8.45: Motion, Tempo and Dynamics for Performer 3, Prelude 6 with blue vertical lines representing the performer's interpretation of phrasing boundaries as in the recorded audio

on the data for the first five phrases showed a significant effect of performer on motion norm (F=8.27, $p<0.05$) and of phrase number on dynamics (F=4.81, $p<0.05$). No other significant effects were found. Again we see that each performer uses motion differently on their approach to phrasing boundaries but are consistent across phrases. The effect of phrasing on dynamics could be a product of the structure of the phrases, as phrase 3 is expanded into 4 bars instead of the original 2. Stronger effects may be noted if the performers' interpretations were in agreement allowing extraction of data at all phrase boundaries.

The measurements of extremes of each parameter performed for Prelude 7 are repeated for this prelude. The resulting box plots are shown in Figure 8.46. The spread of data for all nine performers is remarkably similar.

Comparing these to the spread of data seen in the box-plots in Figure 8.36 for performances of Prelude 7, we can see some similarities for each performer between their data sets from each prelude. This suggests that although the use of parameters for highlighting particular features can change across pieces, performers tend to use the same spread of tempi, dynamics and motion, implying that they each have a certain style.

Again, the results of extracting the extremes of data below the 5th percentile and above the 95th percentile in the spread of data are plotted in scatter plots seen in Figures 8.47, 8.49 and 8.51. These extremes are examined to see how they correspond to the structure of the music being performed.

Observing the scatter plot of Performer 1's data (seen in Figure 8.47) from the performance of Prelude 6, we can see extremes in motion occurring with both maxima and minima in tempo and dynamics. This is different to the interaction of parameters noted for the same performer's Prelude 7 (see Figure 8.37). This is another suggestion that performers use these parameters differently for different pieces. Also noted are a number of combinations in extremes, particularly motion(blue) and dynamics(pink). From the translated score image in Figure 8.48 identifying the outliers in each of the parameters, for performer 1 we can see points of interest particularly at second beat of each phrase marked by a crotchet in the left hand melody. This is in line with Parncutt's analysis of the prelude for melodic accents which appear to be marked here by combinations of extremes in dynamics, motion and tempo (see Chapter 7).

The scatter plot for Performer 2, as seen in Figure 8.49, shows a spread of motion extremes throughout values of tempo and dynamics, however the minima in the dynamic range appear to coincide with low values of motion norm whilst the

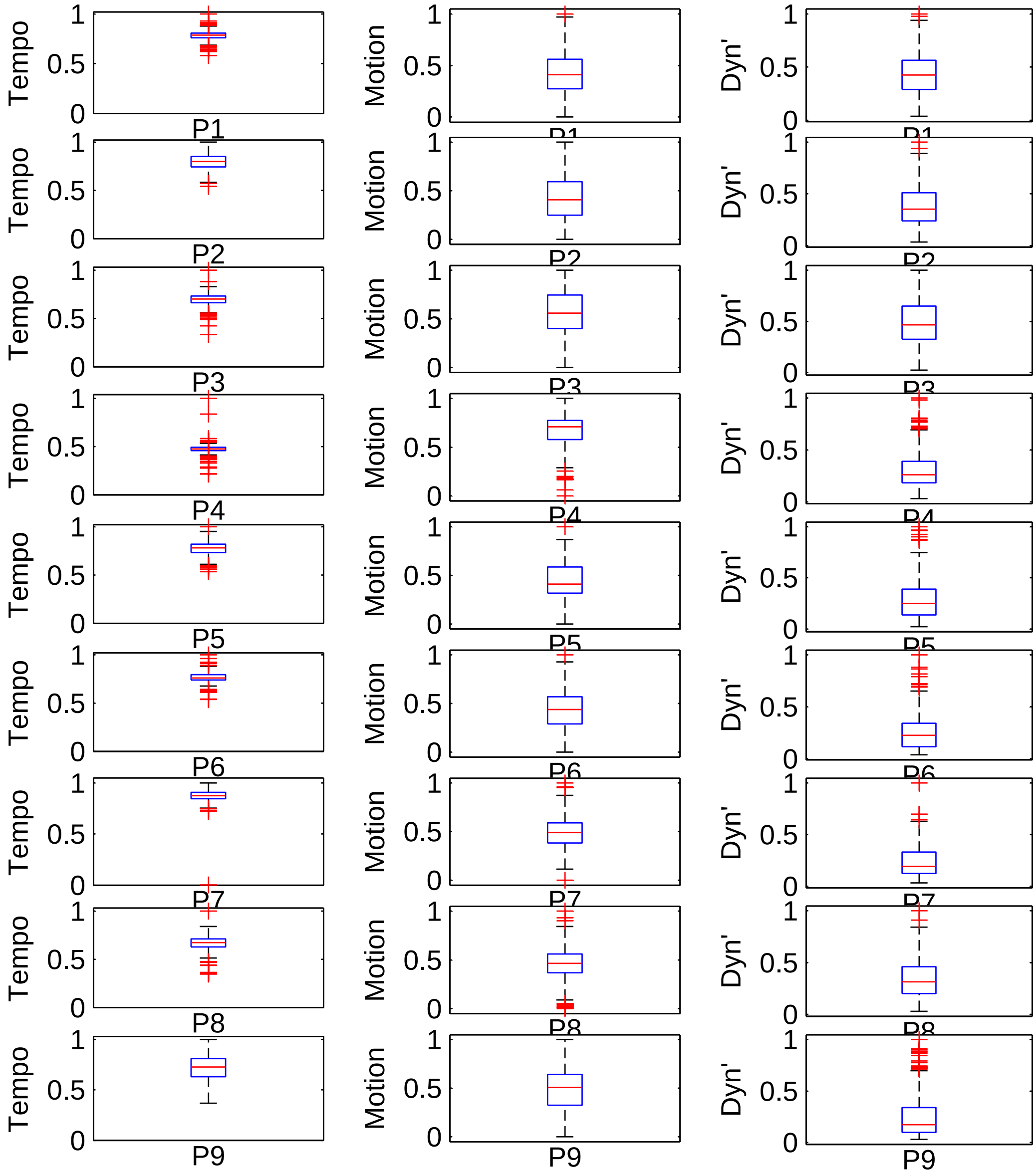


Figure 8.46: Box-plots for all Nine Performers measuring Tempo, Motion norm and Dynamics used in Performances of Prelude 6

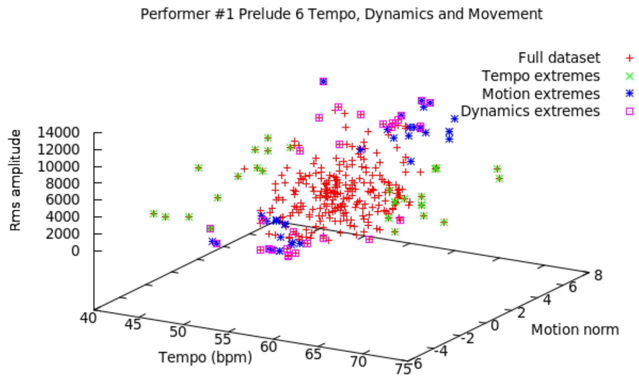


Figure 8.47: Scatter Plot Showing Extremes in Tempo, Dynamics and Motion for Performer 1, Prelude 6

maxima in the dynamic range coincide with high values of motion norm. Again, as for this performer's Prelude 7, the motion and dynamic data tend to occur across a spread of tempi, not limited to low or high values. For Performer 2's score plot in Figure 8.40 shows the most combination of extremes at the beginning of the piece and at the beginning of phrase 3 in bar 5. These align with Parncutt's grouping accents which mark out the beginning and end of phrases. A large cluster of tempo extremes is seen at the end of this first section in bar 8. The beginning of phrase 5 at bar 11 also sees a combination of tempo and motion extremes marking the modulation into C major.

Performer 3's scatter plot of data seen in Figure 8.51, shows distinct groupings of maxima in motion norm occurring at high values of rms amplitude, and minima of motion norm occurring at low values of rms amplitude. This is slightly different to the spread of data found in the same performer's interpretation of Prelude 7 (as seen in Figure 8.41). These values are slightly skewed for tempo as well with the lower extremes in motion and dynamics occurring in the bottom half of the tempo range, and the higher extremes occurring in the top half. Contrary to this performer's use of parameters in Prelude 7, there appear to be more parameter extremes occurring simultaneously with one another. Performer 3's score plot as seen in Figure 8.52, shows a particular cluster of extremes in motion and tempo at bar 7 which in Parncutt's theory contains a cluster of melodic accents. The end of phrase 3 at bar 8 is marked by a cluster of dynamics and tempo extremes. Following this, the melodic accents on beat 2 of each phrase is marked by either dynamics and tempo or dynamics and motion. Bar 13 onwards marks a cluster of motion and tempo extremes as the piece modulates into C major.

Music engraving by LilyPond 2.12.2—www.lilypond.org

Figure 8.48: Annotated Score for Performer 1, Prelude 6 noting extremes in tempo(T), dynamics(D) and motion(M)

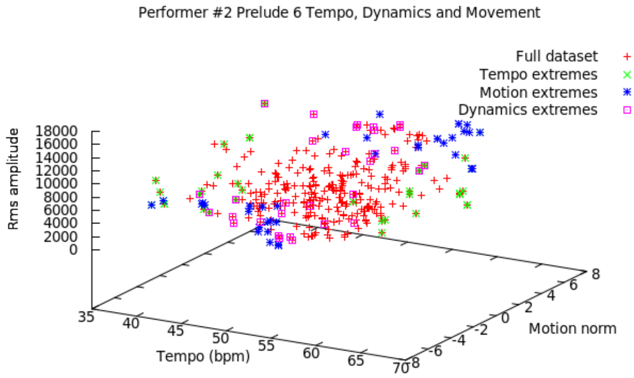


Figure 8.49: Scatter Plot Showing Extremes in Tempo, Dynamics and Motion for Performer 2, Prelude 6

Music engraving by LilyPond 2.12.2—www.lilypond.org

Figure 8.50: Annotated Score for Performer 2, Prelude 6 noting extremes in tempo(T), dynamics(D) and motion(M)

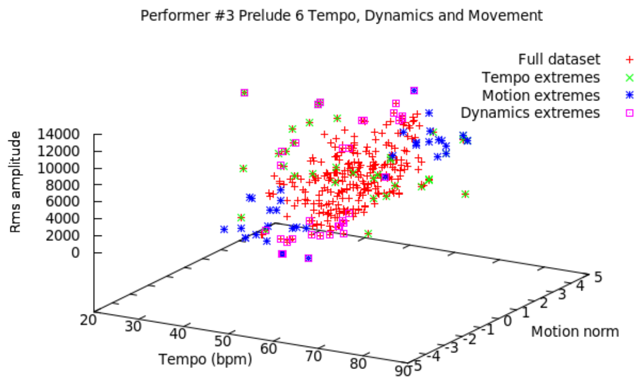


Figure 8.51: Scatter Plot Showing Extremes in Tempo, Dynamics and Motion for Performer 3, Prelude 6

Music engraving by LilyPond 2.12.2—www.lilypond.org

Figure 8.52: Annotated Score for Performer 3, Prelude 6 noting extremes in tempo(T), dynamics(D) and motion(M)

8.4 Conclusions

At the beginning of this chapter, four hypotheses were set out suggesting how performers manipulated parameters such as tempo, dynamics and overall motion in accordance with phrasing structure of the music being performed. Hypothesis 8.1 stated that regardless of the subjective and personal nature of physical gesture in relation to musical structure, there would exist an underlying pattern that was related to phrasing and was common across all performers. Hypothesis 8.2 stated that the underlying motion profile of the performer related to phrasing would be the same across pieces. Hypothesis 8.3 stated that when investigating the role of gesture in multi-modal detection of phrasing, a combination of aural and visual parameters would provide the most accurate indicator of phrasing and following on from this, Hypothesis 8.4 stated that where combinations of global maxima and minima occurred in both aural and visual streams of data, these would be related to the most important structural features of the composition.

From the gestural motion studies conducted in the earlier part of this chapter, it was shown that despite the idiosyncratic nature of the performers' gestures in performances of both preludes, the underlying motion norm suggested the same phrasing structure. This was confirmed by measuring the local maxima of the motion profile between phrases for each pianist. These local maxima occurred reliably at the same point for each phrase for each performer. These patterns were evident across all performers despite their background and ideas on movement within performance. This confirms Hypothesis 8.1. Correlating each performer's patterns of motion profile across their performances of the Preludes, it is shown that few result in a high correlation. Some even result in negative correlations. This suggests that the motion profile for each performer changes depending on which piece they are performing. This rejects Hypothesis 8.2. Factors for this may be due to changes in rhythm, melody or harmony, however, seeing as the rhythmically repeating phrases of Prelude 7 tend to produce similar motion patterns for each performer, it suggests that motion may be highly linked to rhythm. Moving onto the results of the multi-modal analysis, structural information appears to be intrinsic in pianists' use of both aural and visual parameters within their performances. By the box-plots of data for motion norm, dynamics and tempo for each performer for each piece, we can see that the spread of these parameters is not consistent across performers but is similar across pieces. This suggests that each performer has a particular style of playing. This is reinforced by the two-way ANOVAs performed on the distances between the local maxima and minima and the nearest phrase boundary which present a significant effect of performer on motion for both pieces and for dynamics in Prelude 7. Significant effects of phrase number on tempo for Prelude 7 and tempo for Prelude 6 suggest that performers' use of these parameters at the ends of phrases is dependent on the position within the score. As each performer 'style' is different and the use of these parameters can be varied according to the position on the score, it becomes apparent through observation of the multi-modal graphs that a combination of parameters indicates phrasing boundaries. An example of this is clearest at the harmonic arrival between phrases 5 and 6 where global maxima and minima in motion, dynamics and tempo coincide. This suggests that Hypothesis 8.3 is correct.When examining the maxima and minima of the dataset and their occurrence in the musical score, it is clear that performers tend to use combinations of these extremes at important points in structure, suggesting that Hypothesis 8.4 is correct. The location of these extremes in motion dynamics and tempo occur at particular accents of harmony, melody and rhythm set out by Parncutt.

In conclusion, structural information can be elucidated from examining certain performance parameters. The continuous multi-modal streams form patterns for each of the phrases and the extremes of the performer's use of tempo, dynamics and motion identify the most important structural features.

Chapter 9

Elucidating musical structure

In the previous chapter, results suggested that there are underlying patterns of physical gesture across all performers and that these could be used to identify phrasing boundaries. In combination with aural parameters of tempo and dynamics, clues concerning the hierarchical phrasing structure can be detected. This experiment is designed to indicate whether phrasing structure can be predicted purely from patterns of performance parameters, particularly for pieces of music where the structure is not so explicit. Within this experiment I also explore the role of finger gesture in piano performance and whether enhanced movements can be related to specific accents. This follows on from the exploration of accents illustrated in Chapter 8.3.

Six professional pianists were recorded performing Chopin's Prelude in A major (Op.28 No.7) and the finale of Chopin's B flat minor sonata (Op.35). These recordings were captured through audio, MIDI and finger motion analysis. This chapter analyses the recordings taken with the multi-modal system described in Chapter 9.1, building on the methodology described in [76] and extending the preliminary results published in [79, 80]. The pianists are directed to perform both Chopin pieces as they would in a normal concert situation. These six professional pianists are combined from a mixture of lecturers in piano studies and postgraduate students in the following music conservatories: the Royal Scottish Academy of Music and Drama, Glasgow (Fali Pavri and Carlisle Beresford Anderson Frank), Napier University, Edinburgh (Simon Coverdale), Royal College of Music, London (Jessica Chan), Royal Northern College of Music, Manchester (Lauren Hibberd) and Royal Academy of Music, London (Martin Jones). In order for structure to be discovered in cases where no a priori information on phrasing is available, this experiment requires the performer to have concrete ideas on the finale and experience of performing the whole sonata. For this instance, professional pianists who have these pieces as part of their performing repertoire are used.

The pianists are also recorded for their interpretation of Chopin's Prelude in A major to provide a set of control data. For this piece, we can establish how each performer uses aural and visual parameters at the phrasing boundaries. As these phrasing boundaries for the finale are not explicitly known, the analysis will focus on the note level with the first two phrases of the prelude and the first five bars of the finale.

9.1 Method

As in Chapter 8, these multi-modal parameters of tempo, dynamics and motion will initially be measured for an explicitly structured piece, Chopin's Prelude in A major op.28 no.7, which contains a rhythmically repeating two bar phrase (see Chapter 7.1). These techniques will then be used in identifying performers' interpretation of an ambiguously structured piece, Chopin's B flat minor sonata op.35 finale movement, the opening bars of which are shown in Chapter 7.3. Despite the diverse opinions as to its analysis, this finale is still a widely performed piece as part of the B flat minor sonata.

Each performer is recorded using the data capture system described in [79] and in Chapter 5.2. This system captures audio, MIDI and video data whilst ensuring minimum disturbance to the performer. The audio data is acquired through the open source application Ardour whilst the MIDI data is captured through the Moog Piano bar device [5] into the application Rosegarden. As the finale is generally performed at fast tempi and is also technically difficult, it is expected that performers' full body movements will be restricted [84] and so the motion analysis focusses entirely on hand movements. Each knuckle and joint of each hand is detected as an x,y coordinate with the z coordinate estimated from the 3D algorithms in Chapter 4. A post-recording self-report was conducted for each of the performers, providing us with their own interpretations and comments on each piece. The hypotheses for this experiment follow on from the results in Chapter 8. It is expected that by examining patterns of finger motion, tempo and dynamics for performances of Chopin's Prelude in A major, phrasing boundaries will be discovered in performances of the finale. As this is a subjective measurement, the stated hypotheses that follow will be more observations on the analysis of these parameters.

Hypothesis 9.1 *Trajectories of finger motion in the x, y and z axis will reflect expressive accents within the phrase.*

Hypothesis 9.2 *It is expected that wrist motion that reflects movements toward the soundboard of the keyboard, and movements toward the key-bed will produce high values of rms amplitude.*

Tempo and dynamics information are extracted from the MIDI and audio data streams respectively. The MIDI data is processed to create matched PML files (see Chapter 6.1.1 for this process) where each performed note is aligned to a score note through the use of note IDs. The PML file is then submitted to the database designed in [94]. This data is then queried for inter-onset intervals (IOIs) between matched notes and keypress durations and returns a text file with this information for each note. A separate query produces these IOIs and keypress durations as bars plotted above notes in a score produced using the music typesetting program Lilypond [3]. This is particularly useful in fast pieces where a normal time graph may lose the intricacies of measurements for each note. The calculated IOIs are converted into an estimation of tempo by normalising each note to a crotchet or dotted crotched beat depending on the time signature of the piece, and dividing by 1/60 producing a beats per minute (bpm) value.

Despite the availability of onset and offset pedal information from the MIDI bar, the pedal markings are not included in the displays of keypress durations. It was decided that as the pedal information is only present for when the sustain pedal is fully depressed, it may not be of much use as professional pianists use a range of pedal angles to alter the sound. Observations of the spectra of notes with pedal on and off are analysed in [67]. The keypress durations are therefore, not exactly a measure of the length of time that a particular note is audible but rather a reference of how long the key is held down for as an estimation of articulation. This will provide certain clues on accenting of particular notes.

Dynamics are estimated by taking a simple measure of the rms amplitude of the audio signal. This was preferred over velocity values for each note as the rms amplitude would provide a better estimation of the overall loudness. This was decided as the more important value when considering how the performer communicates phrasing structure.

The motion data is extracted as x, y, z coordinates for each marker. As the hand has such a high number of degrees of freedom, it was decided not to condense the data using principal components analysis as in the previous chapter, but instead to examine particular markers of interest individually. As audio and video results are not yet stored within the PML representation (although this is in development), they are linked with the performed music by using the open source program Audacity to manually label the bars of the piece from the audio recording.

Particular issues arising from the methodology occurred in both the audio and video recordings. An unforeseen issue with the Moog Piano bar arose from recording the full range of MIDI notes through one channel. The default function for the device is to split the keyboard into two channels at the D flat below middle C and so for the first three recordings (those of Fali Pavri, Simon Coverdale and Carlisle Frank), only a percentage of MIDI notes have been recorded. This was corrected for the latter recordings of Jessica Chan, Martin Jones and Lauren Hibberd. The motion capture method of placing the camera directly above the keyboard appeared to capture the most information of the hand movement. However in two cases, those of Simon Coverdale and Jessica Chan, the performer moved their head over parts of the hands, obscuring them to the camera. The motion data was estimated for these particular cases and would be improved with better estimation algorithms as detailed in Chapter 4. For the following results, the examples taken are those of Lauren Hibberd, Martin Jones and Jessica Chan. The results for the other three performers can be seen in Appendix D.

Tempo and dynamics information are plotted against the phrasing boundaries in the same form as for the previous experiment. As the fingers of the hand can move largely independent of one another (with some biomechanical constraints), it is decided that PCA analysis will not be useful in examining finger motion during performance. Instead, a few markers on each hand will be examined in isolation, enabling the exploration of how finger movement contributes to the overall phrasing of the piece. Considering the finger motion data, the x axis relates to movement along the length of the keyboard, the y axis is movement towards and away from the keyboard and the z axis relates to the height estimation. These three axes are heavily influenced by the arrangement of notes being played and height needed in preparation to physically play each note. These factors are all closely related to sound production. However, what becomes apparent from these measurements are products such as specific accents and groupings of notes, which contribute to the performer's interpretation. Larger body movements in fast pieces such as these are few and far between, so it is expected that these small measurements will provide the most information.

9.2 Results

Each performer is examined first for their performance of the Prelude and then for the Finale. For a concise spread of results, three performers out of the six are examined with the remaining graphs attached in Appendix D. It can be observed from performances of the Prelude how performers employ expressive techniques to communicate structural information such as phrase endings. Multi-modal parameters are displayed as stacked graphs for overall comparison of aural and visual cues. For these graphs, tempo is plotted as an estimation of beats per minute extracted from the MIDI data, an estimation of dynamics is presented as the rms amplitude of the audio signal and the left wrist marker and thumb's metacarpophalangeal marker movement for each hand is plotted as an example of general hand movement. For these movement graphs, the y axis reflects movement towards and away from the keyboard, with the y value increasing as the marker moves further away from the keyboard. The x axis reflects movement from the left to the right of the keyboard, the x value increasing as the marker moves to the right. The z axis estimate reflects movement in height from the key-bed towards the camera, with the z value increasing as the marker moves away from the camera. This value is an estimate subject to noise, due to the limitations of the system design (see Chapter 4.5) in an effort to construct a lightweight, low-cost image capture application. It should therefore be taken as an indication of height changes instead of a strict measurement. This, however, will provide valuable information as to how finger height changes throughout each phrase and in relation to other audible parameters.

9.2.1 Martin Jones

Martin Jones' performance of the Prelude can be examined in Figure 9.1 where we observe tempo, dynamics and wrist movement parameters plotted for the first two phrases. The two large movements for the left hand wrist marker in the x axis around 16s and again at 17s are most likely products of the physical movement required to play the consecutive chords at the end of each phrase, following the bass note at the beginning. Most of the movement and tempo fluctuations appear to be at the start of this phrase.

The thumb motion for the proximal interphalangeal marker is displayed in Figure 9.2 allowing observation of particular fingers' motion alongside the aural parameters. The x axis can be viewed as a representation of pitch for the two

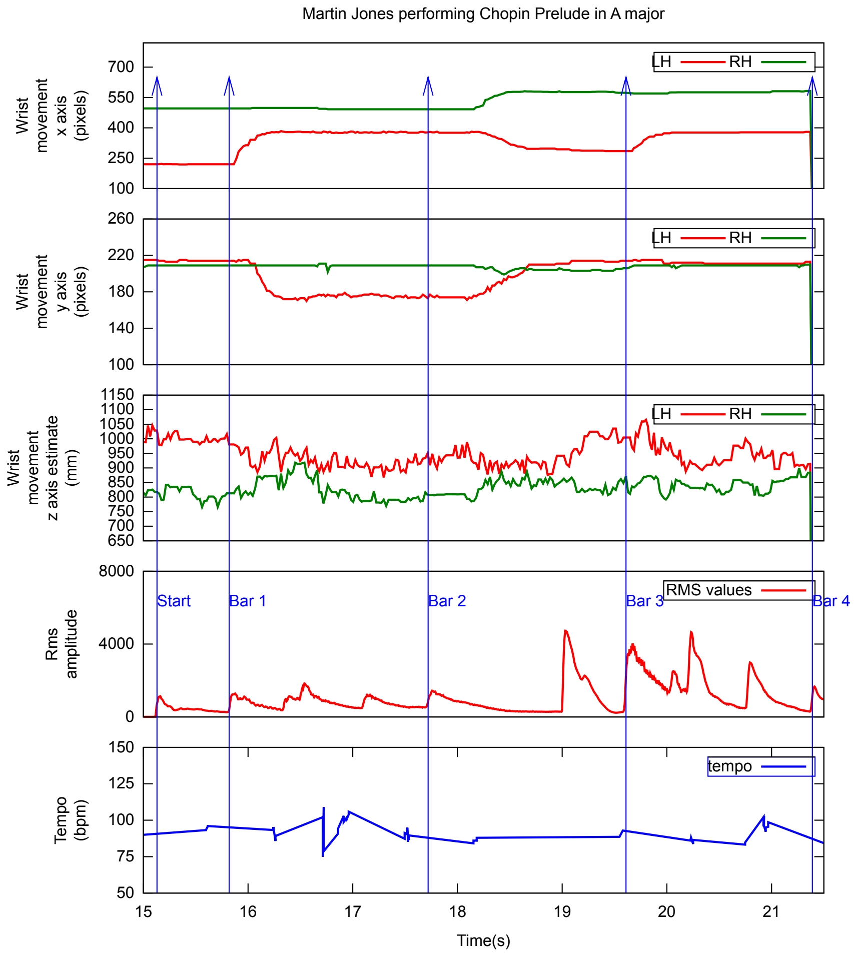


Figure 9.1: Wrist Motion, Tempo and Dynamics for Martin Jones, Prelude in A Major, with the first phrase running from the first blue vertical line until two-thirds through bar 2 and the second phrase running from this point until two-thirds through bar 4.

hands. In the z axis, in both hands, there is an increase in distance away from the camera, towards the key-bed, towards the middle of the phrase. This phrase runs from the start arrow to two-thirds through the second bar and also shows a decrease in distance towards the end. This suggests that the hands are shaping the phrase with the emphasis being on the middle of bar one. The three chords at the end of the phrase seem to be played with decreasing height, which would suggest the chords are being played with a lighter touch, and we would expect smaller measures of rms amplitude for each consequent chord. This is seen in the measurements of dynamics underneath.

The results of the lilypond typeset graphic produced from the database are displayed in Figure 9.3 showing both IOIs and keypress durations in columns underneath each matched note. The first column represents the IOIs data and the second column represents the keypress duration data. In this particular case, as the Prelude is short and relatively simple in harmony and structure, we do not glean much more information from this representation than noted in the previous time graphs, and so this figure is included purely for interest as it is still a better representation considering each note.

This alternative representation although useful for scrutiny of every single note, does not provide far more information than the original graphs in this case. They may however, be useful for exploring comparisons between performers in a note-to-note basis. For the rest of the examples, the database figures will only be shown for performances of the finale. The remaining three pianists' database results for the finale are seen in Appendix E.

For Martin Jones' performance of the finale, the wrist motion is displayed alongside tempo and dynamics in Figure 9.5. Again, the x axis can be considered a representation of pitch and to an extent, the y axis represents the playing of black notes, as the hand is generally moved into the piano to allow the performer to reach the note. This is not exclusive however, as we see the difference between the first two bars in the y axis despite the construction of notes in each bar in terms of black and white notes is similar. Interestingly, the left and right hand do not show the same pattern of movement which may be expected as the y axis movement was entirely dependent on the position of the white and black notes in the score. In this axis, the right hand shows a repeated movement spanning the first four bars which suggests the twelve quavers in each bar are separated into groups of six which is also reflected in the tempo patterns. The z axis of wrist movement shows peaks (which reflect a higher distance away from the camera, and thus a movement towards the key-bed) at the beginning of bar 2 and in the first and second halves of bar 3. Dynamics show a clear separation in the middle of bar 3 and again a dip at the beginning of bar 5.

Observing the thumb motion in Figure 9.6 for which the right hand plays what may be considered by some analyses the most accented notes in each bar, we can see a similar pattern in the y axis to the wrist motion in Figure 9.5 where the dips correspond to the notes being played in the right hand. The left hand thumb plays different notes to the right hand and so exhibits a different pattern of motion suggesting this accenting may be true for this particular performance.

Focussing more on the aural parameters and the particular note accents, the database result of the IOIs and keypress durations is presented in Figures 9.7 and 9.8. This shows a particularly elongated note for the F at the very beginning of the piece, with the B flat in the 3rd bar also held down for considerably longer than the consecutive notes. This elongated note coincides with the emphasised movement seen in Figure 9.5. This note elongation is not imitated at the beginning of bar 5, suggesting the performer is not making an effort to distinguish this bar from the previous notes. The approach to this potential boundary is not characterised by notable fluctuations in tempo, however, there is a slight diminuendo at the end of bar 4.

For Martin Jones' performance of the finale, we can infer that bar 5 is not marked particularly as a phrasing boundary but simply a continuation, particularly as his measurements from the prelude appear to highlight the start of new phrases with all three measured parameters. Attention is drawn to bar 3, where particular accents in movement and tempo are most likely a result of the change in composition where each group of six quavers are now different pitches as opposed to the repetition in pitch of six quaver groups seen at the beginning.

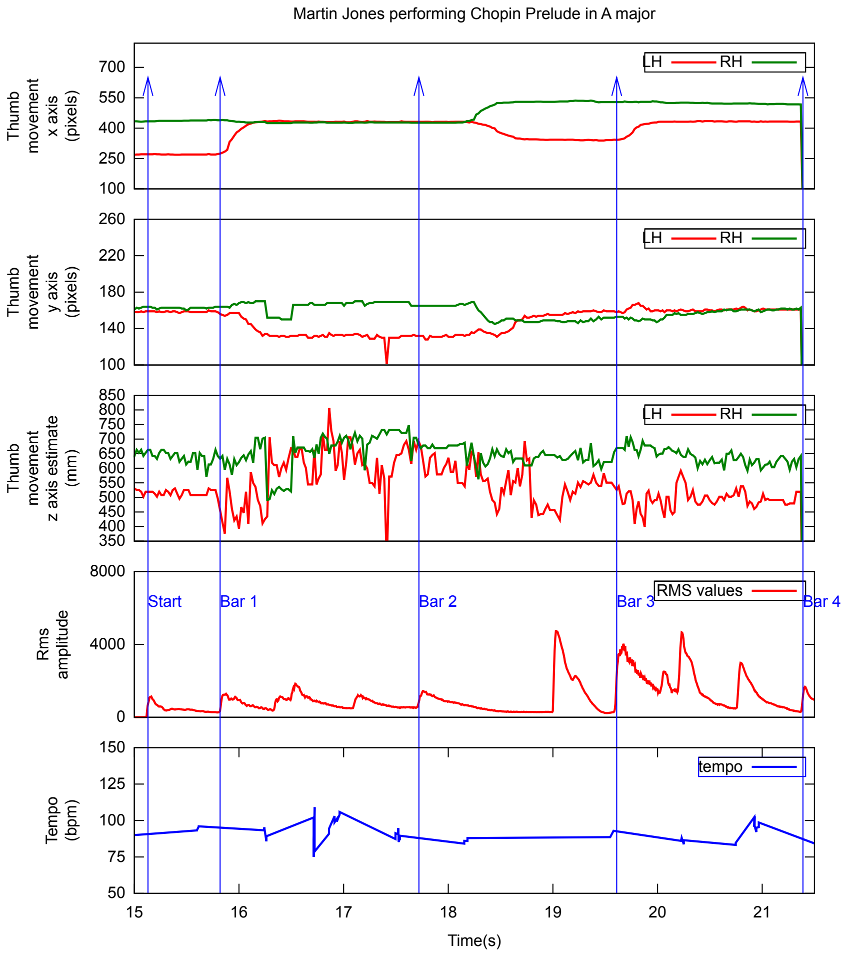


Figure 9.2: Thumb Motion, Tempo and Dynamics for Martin Jones, Prelude in A Major, with the first phrase running from the first blue vertical line until two-thirds through bar 2 and the second phrase running from this point until two-thirds through bar 4.

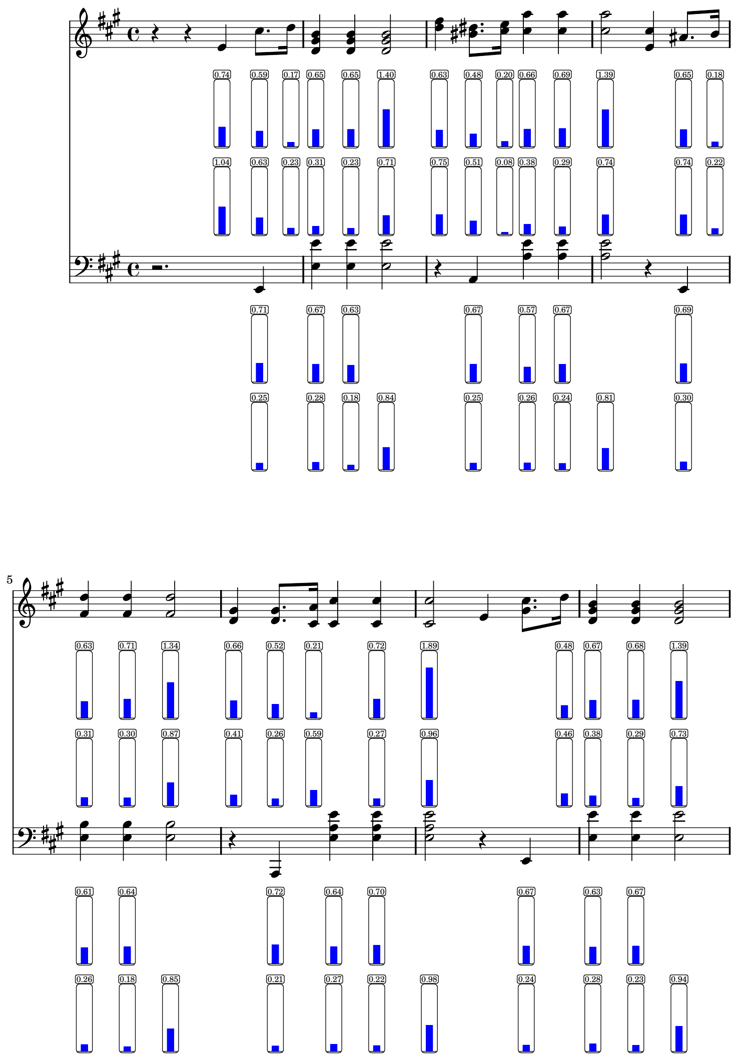


Figure 9.3: Database Results Page 1 for Martin Jones, Prelude in A Major, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

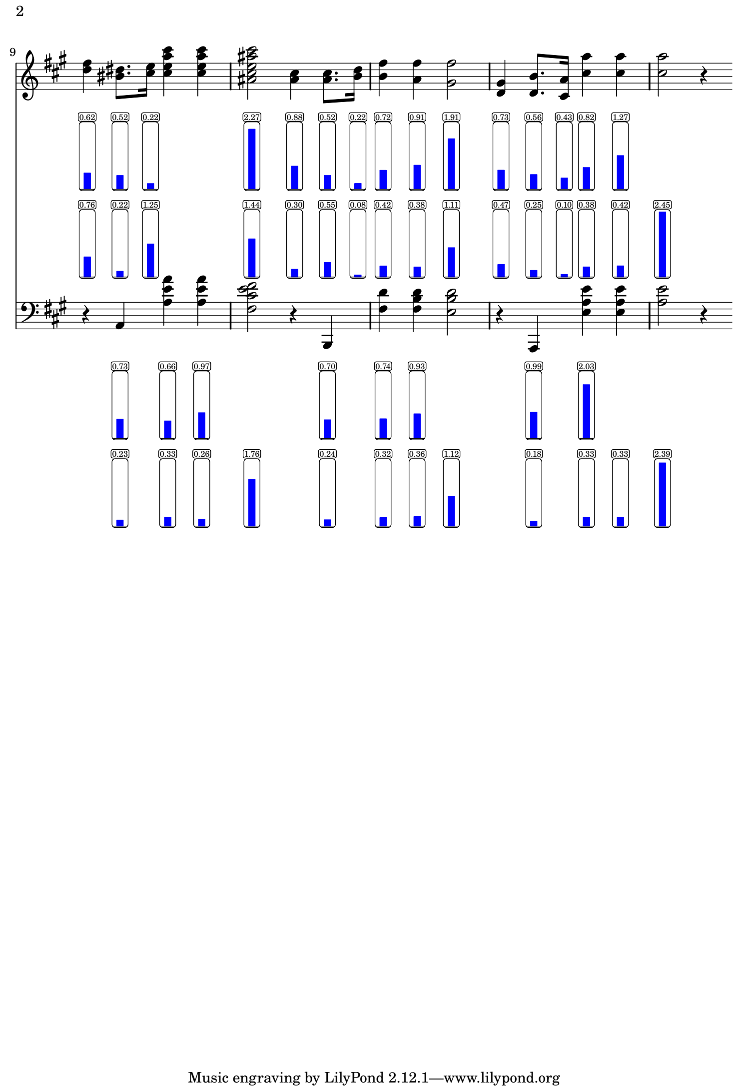


Figure 9.4: Database Results Page 2 for Martin Jones, Prelude in A Major, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

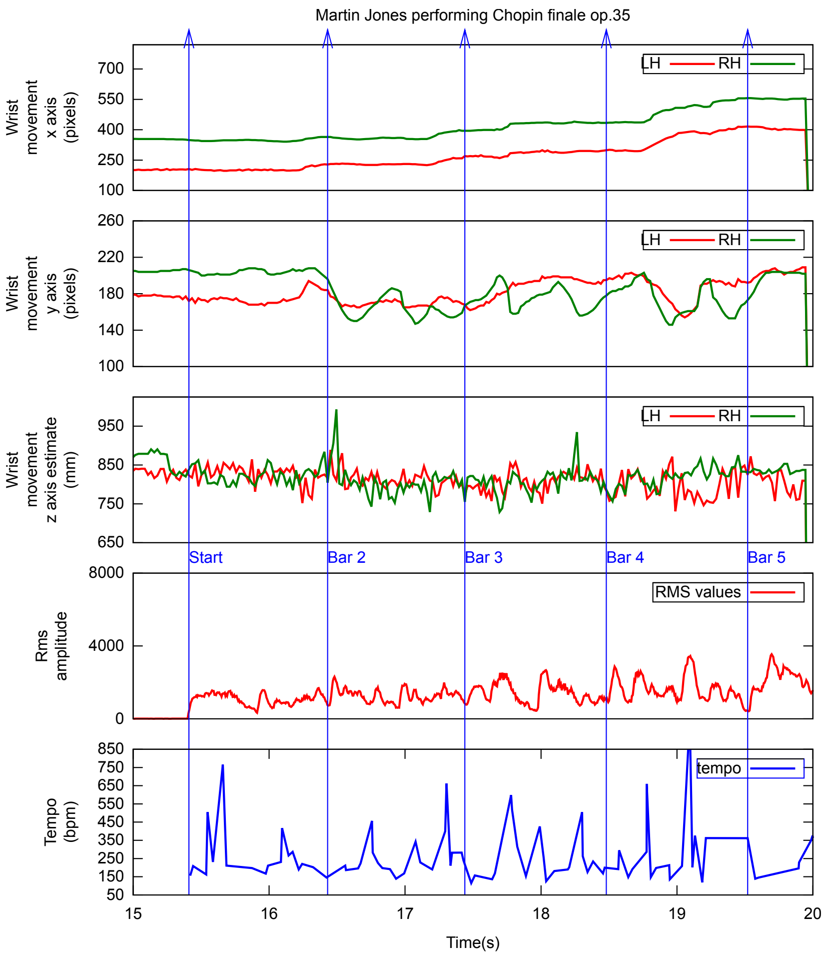


Figure 9.5: Wrist Motion, Tempo and Dynamics for Martin Jones performing the Chopin finale

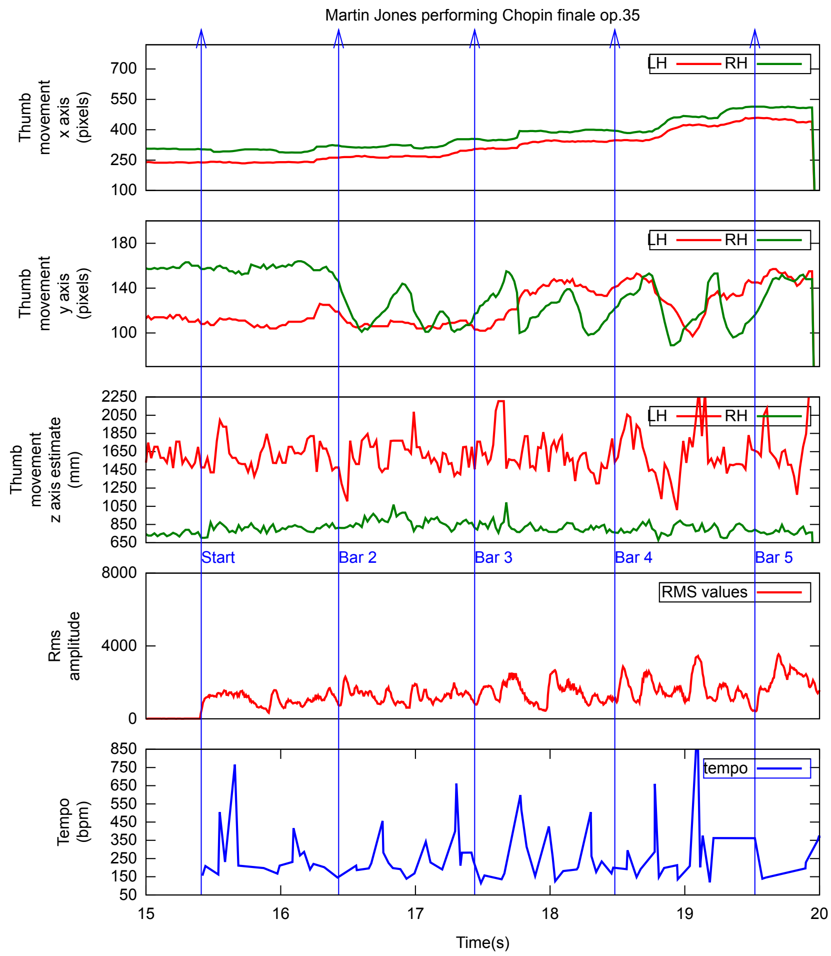


Figure 9.6: Thumb Motion, Tempo and Dynamics for Martin Jones performing the Chopin finale

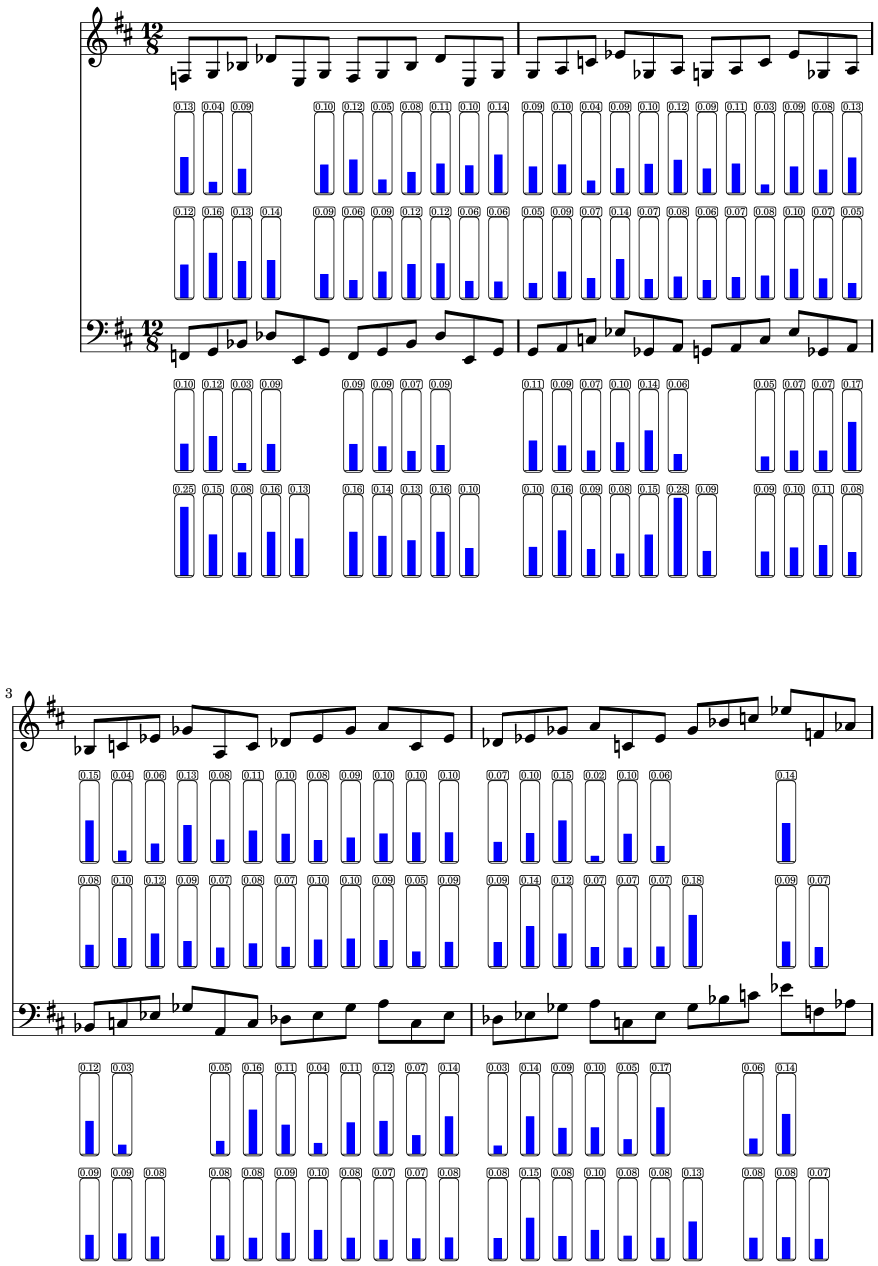


Figure 9.7: Database Results Page 1 for Martin Jones, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

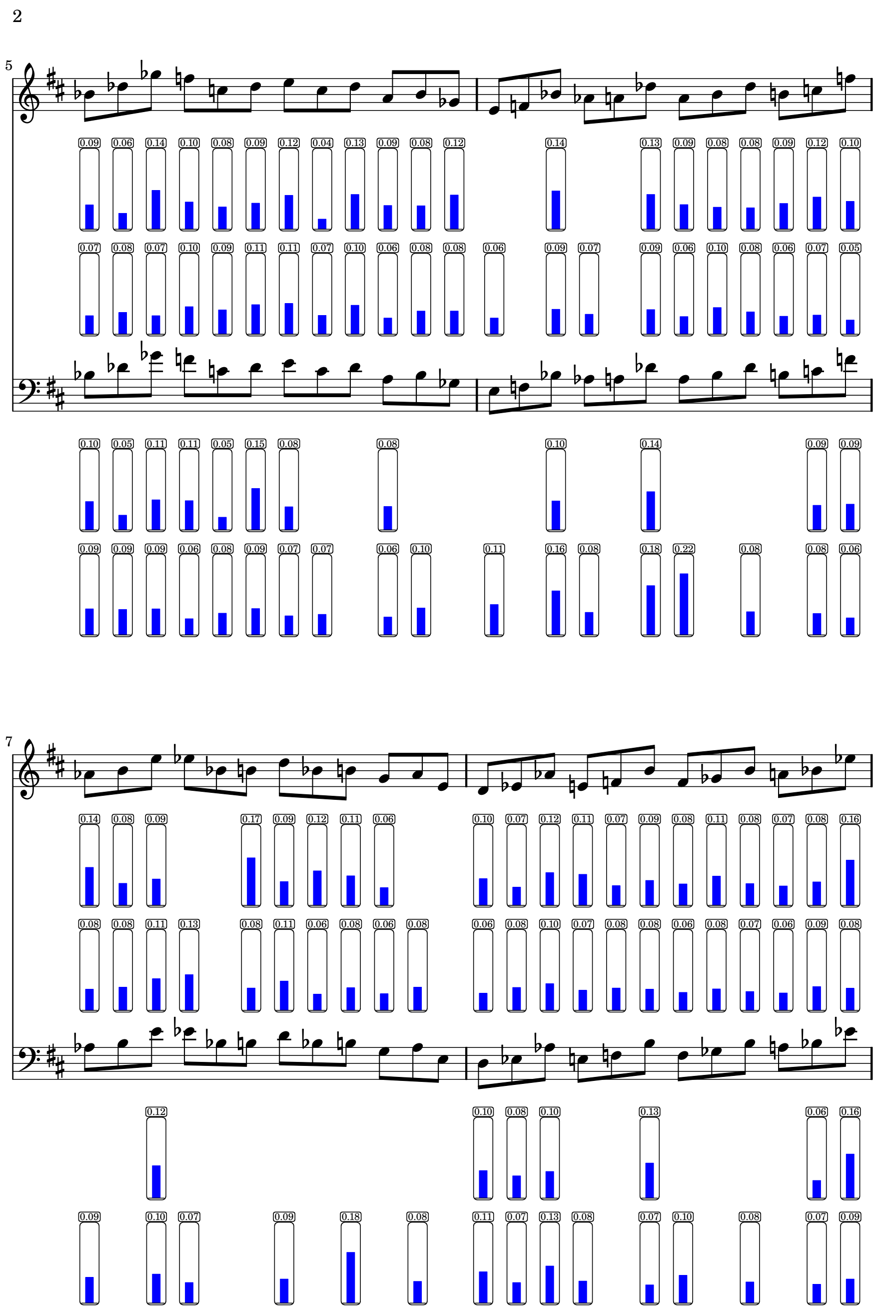


Figure 9.8: Database Results Page 2 for Martin Jones, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

9.2.2 Jessica Chan

Considering another set of performances, Jessica Chan's Prelude performance is plotted in Figure 9.9. A similar pattern in the x axis of the wrist motion reflects the pitch of the phrase. Jessica's wrists move more frequently towards and away from the keyboard, with each chord being shaped by the movement of the hand. From this we can see how the performer 'releases' each chord by movements away from the keyboard and towards the camera in height. A general decrease in dynamics is seen throughout the phrase (from the beginning of the piece until two thirds of the way through bar 2) with a dynamic peak on the metrical accent on the first beat of every two bars.

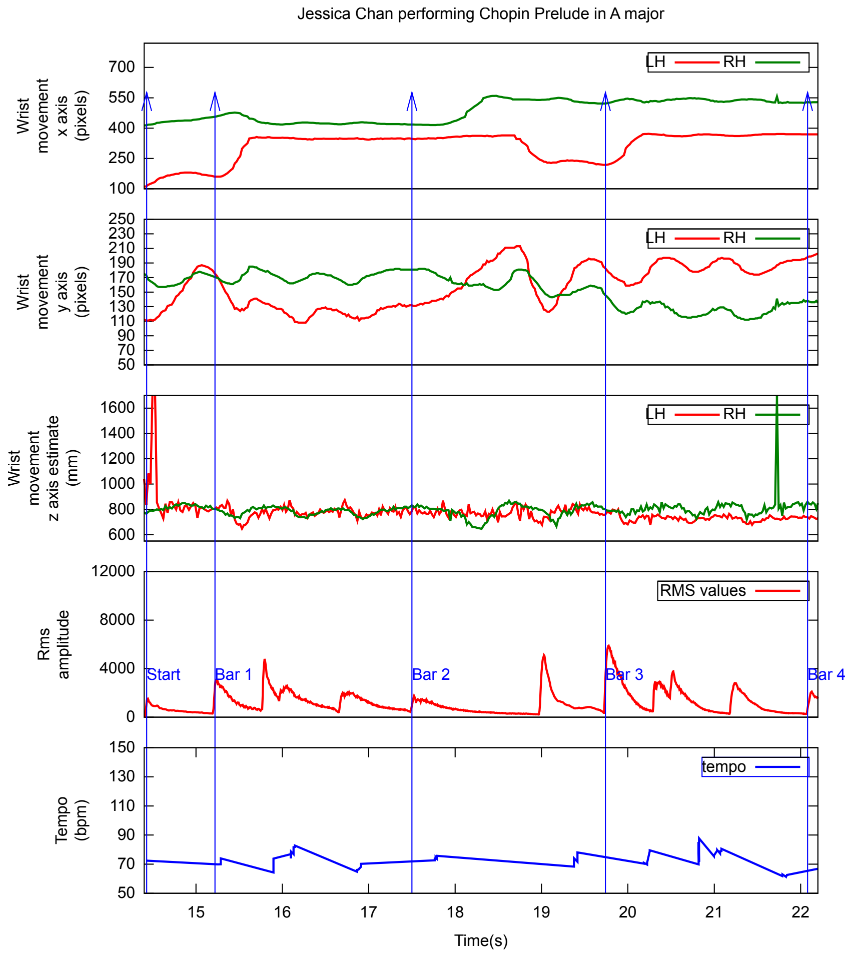


Figure 9.9: Wrist Motion, Tempo and Dynamics for Jessica Chan, Prelude in A Major, with the first phrase running from the first blue vertical line until two-thirds through bar 2 and the second phrase running from this point until two-thirds through bar 4.

The thumb motion suffered during recording by being obscured by the head, but the data displayed in Figure 9.10 still shows a pattern where each chord experiences a movement in the hand which may be the performer 'releasing the chord'. The biggest thumb height fluctuation is seen at the beginning of each phrase (at the start and at the end of bar 2) despite it not being responsible for the production of each beginning note.

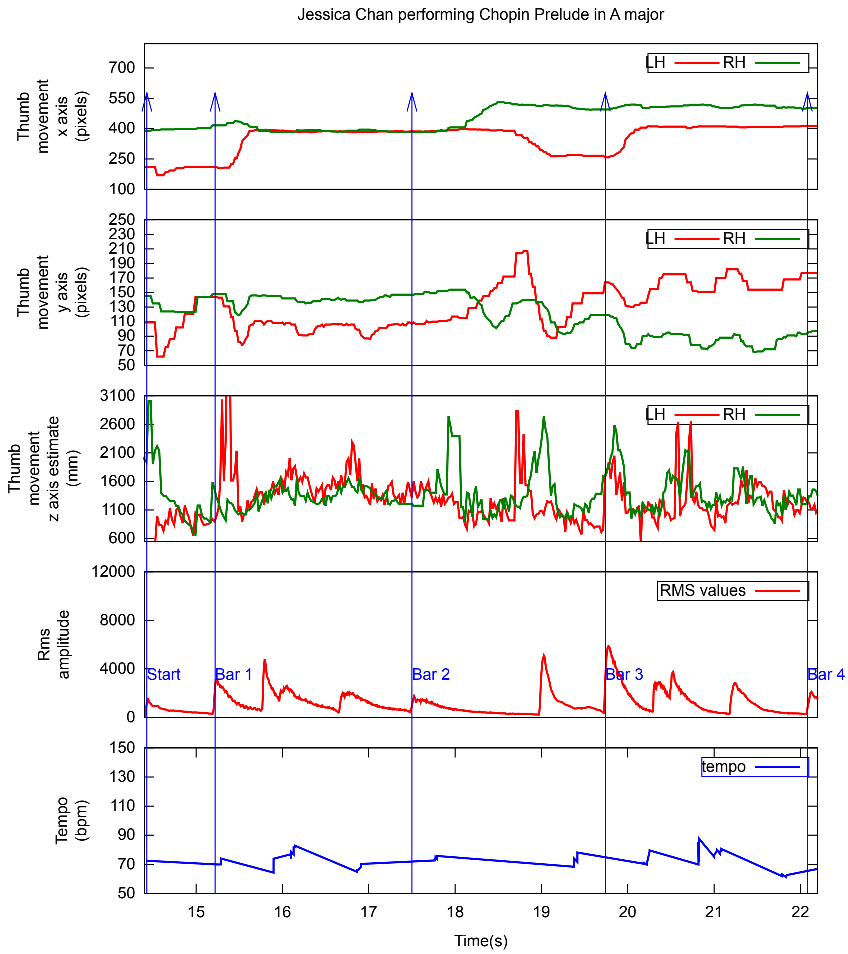


Figure 9.10: Thumb Motion, Tempo and Dynamics for Jessica Chan, Prelude in A Major, with the first phrase running from the first blue vertical line until two-thirds through bar 2 and the second phrase running from this point until two-thirds through bar 4.

Jessica Chan's performance of the finale is seen for these same parameters in Figure 9.11 which much like Martin Jones' tempo estimations suggests grouping the notes into sixes. The dynamics here reflect this grouping to a certain extent with the peaks in bar 2 and bar 3 and large peak just before bar 5. The grouping is demonstrated physically by the y axis movement. An interesting point in the z axis movement occurs simultaneously with the peak in rms amplitude occurring around the E flat of the 4th bar which is also the highest pitch occurring across the piece so far. Another point occurs before this in the z axis where there is an decrease in distance from the camera coinciding with a shift away from the keyboard in the y axis. This could possibly be a product of a fingering change resulting in a quick lift away from the keyboard.

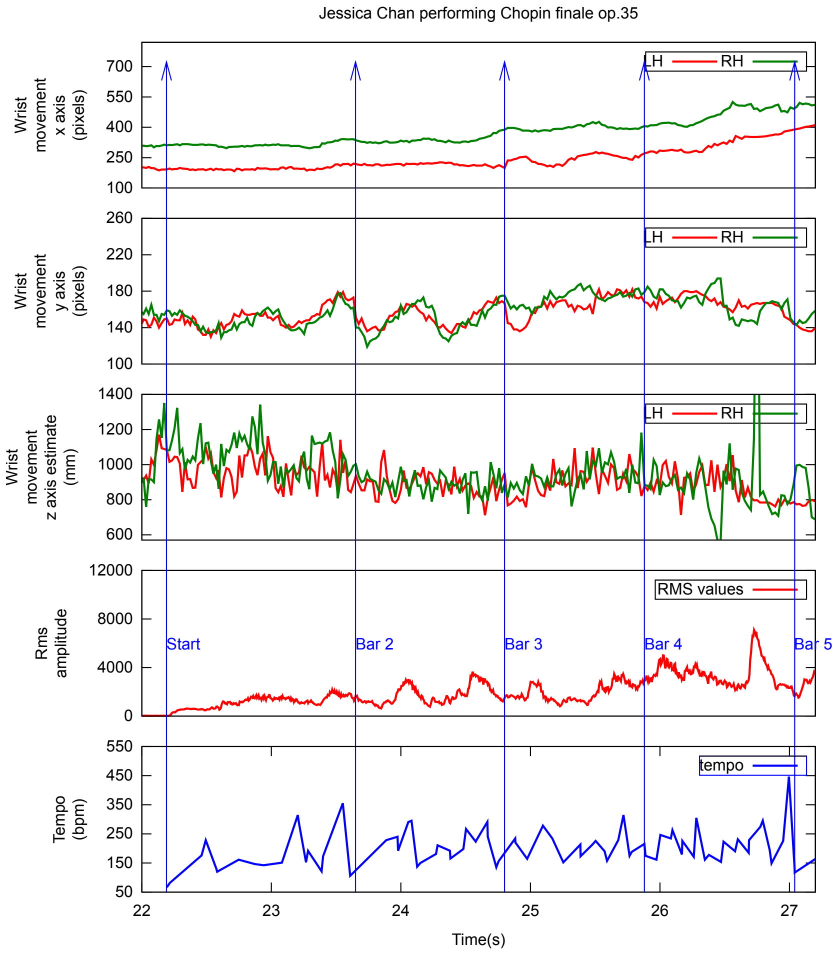


Figure 9.11: Wrist Motion, Tempo and Dynamics for Jessica Chan, performing the Chopin finale

The thumb motion is seen in Figure 9.12 shows large increases in distance from the camera, moving towards the key-bed in the first bar, particularly in the right hand. This pattern does not continue suggesting that this particular emphasis is just for the opening bar of the phrase. Another peak in distance in the z axis is seen in the second half of bar 3, much like Martin Jones' emphasis of this change in composition.

Delving into the note level of the aural parameters, the database result for Jessica Chan is displayed in Figures 9.13 and 9.14. We observe again a slight elongation in IOI and keypress duration for the first note in the piece but not as pronounced as in Martin Jones' performance. This coincides with the emphasised

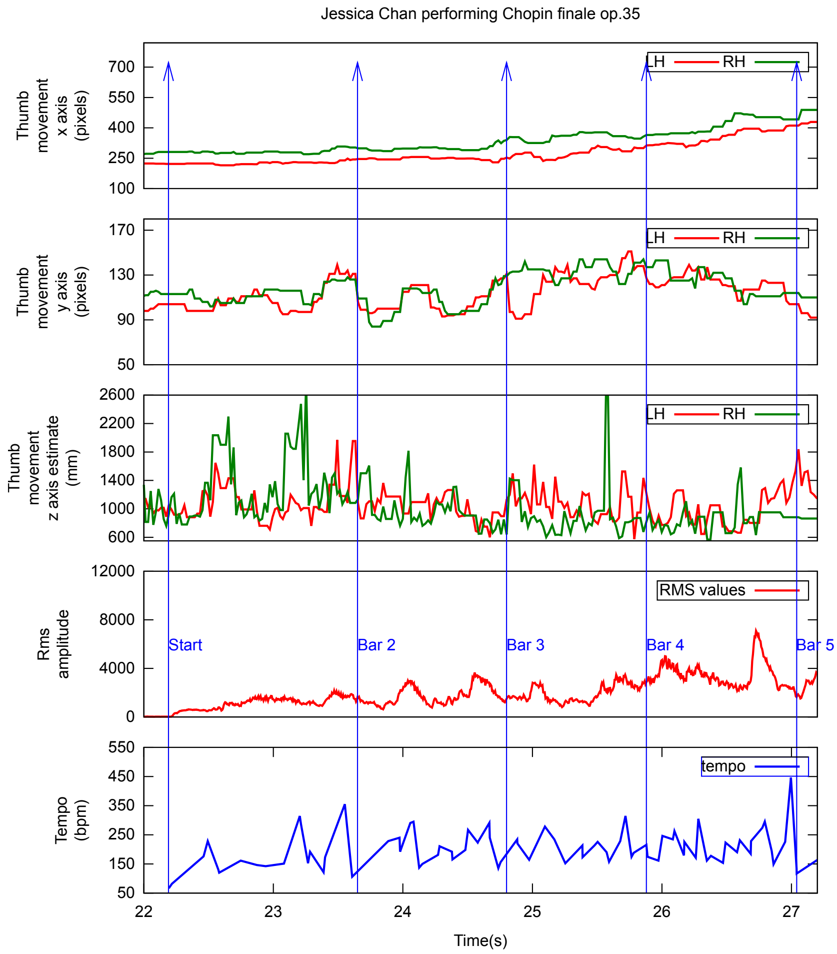


Figure 9.12: Thumb Motion, Tempo and Dynamics for Jessica Chan, performing the Chopin finale

movements seen in the first bar for the thumb movement in Figure 9.12. The E flat in the fourth bar also shows a particularly held on note, reflected in the previous movement and dynamics analysis. However, no specific accents appear to occur at the beginning of bar 5.

Relating this information back to her performance of the Prelude, we see a large fluctuation in dynamics and movement near the end of bar 4 in the finale which would suggest the end of a phrase, however, this does not appear to be characterised by the same movement in tempo. We can infer from this that although bar 5 has not been 'marked' by the performer as a definite phrasing boundary, it is still recognised as a juncture where the notes experience a change in composition and key, much like the change midway through bar 3.

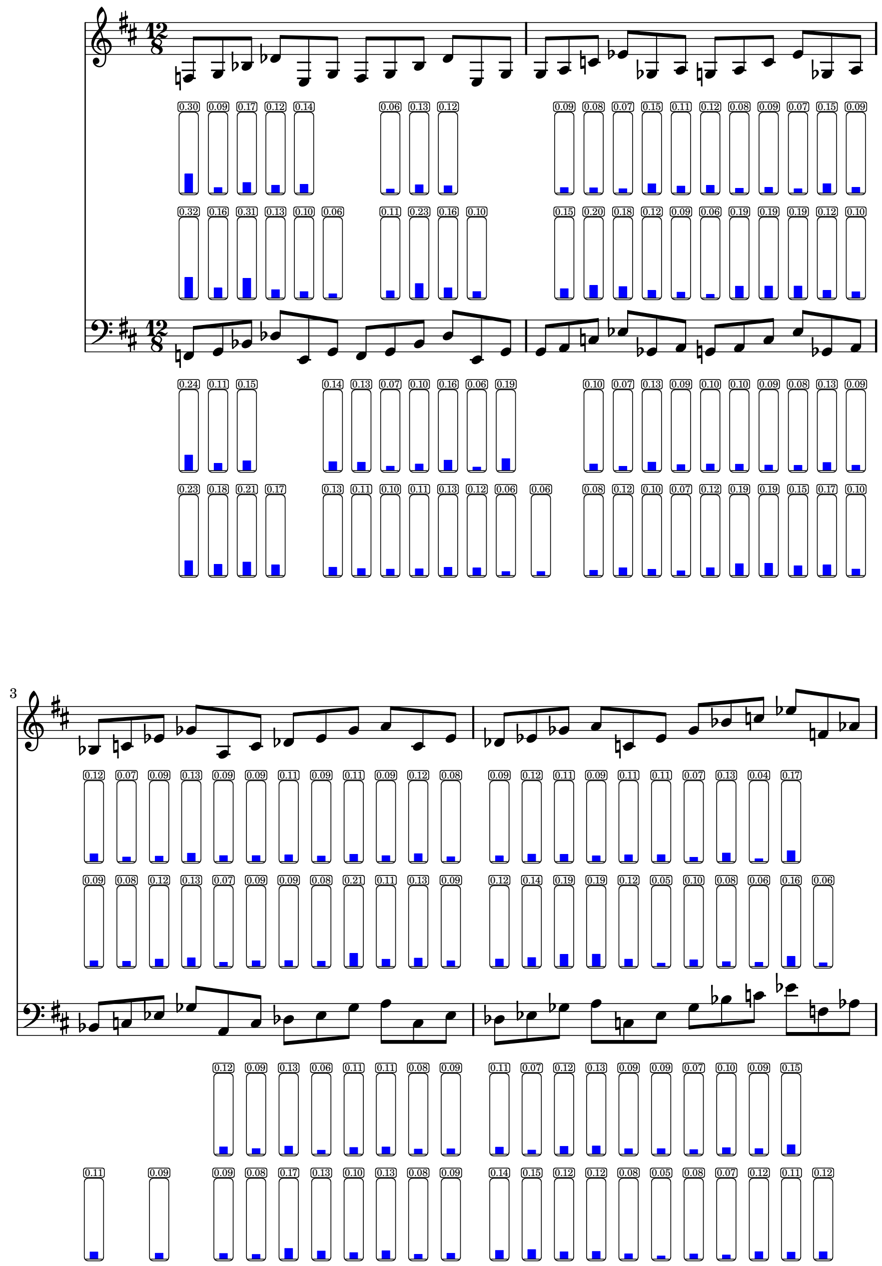


Figure 9.13: Database Results Page 1 for Jessica Chan, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

Figure 9.14: Database Results Page 2 for Jessica Chan, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

9.2.3 Lauren Hibberd

The final example of results is produced from Lauren Hibberd's performances. Figure 9.15 shows only slight fluctuations in tempo for each phrase, increasing slightly in the middle at halfway through bar 1, and decreasing slightly for the end of the phrase at two thirds through bar 2. The dynamics increase from the beginning of each phrase and does not show any overall decreases apart from those characteristic to the piano action. The wrist movement shows the 'releasing' action of each chord much like the performance of Jessica Chan. A increase in distance away from the camera at the beginning of bar 1 in the left hand could reflect the metrical accent of the first beat of that bar, an accent which occurs again at the beginning of bar 3. The general shape of the wrist movement of the left hand in the z axis reflects this phrase shaping, with the accent on the first beat, followed by the three chords played at the same height.

The thumb motion as displayed in Figure 9.16 shows a slightly different pattern in the y axis movement with a movement into the keyboard at the last chord of phrase 1 occurring two thirds through bar 2. This also occurs in the left hand at the beginning of bar 1 and 3 where the metrical accents of the phrase occur. Slight decreases in height above the normal can be seen at these accents as well. Height fluctuations in the left hand show particular emphasis at these metrical accents as well as on the last chord of the phrase.

Observing this pattern of performance parameters for Lauren Hibberd's performance of the finale, as seen in Figure 9.17, we can see a general crescendo in dynamics towards bar 5. The tempo fluctuations again appear to group each bar of quavers into sixes. This grouping can be seen reflected in the y axis movement, however this pronounced shaping ceases at bar 4 where the left hand decreases at one point near the start of the bar, and the right hand decreases at another point near the end of the bar. Both of these occur simultaneously with increases in the rms amplitude. Wrist movements in the z axis in bars 1-2 also reflect this grouping with a particular increase in height towards the key-bed halfway through bar 4 coinciding with a peak in the rms amplitude.

The thumb motion for the finale as seen in Figure 9.18 shows similar results to the wrist motion in the y axis, with three distinct peaks towards and away from the piano in bar 3. The thumb height also reflects the grouping movements in bars 1 and 2.

The database results in Figures 9.19 and 9.20 shed light on the previous graph findings by displaying increased IOIs and keypress durations at the 3rd and 4th

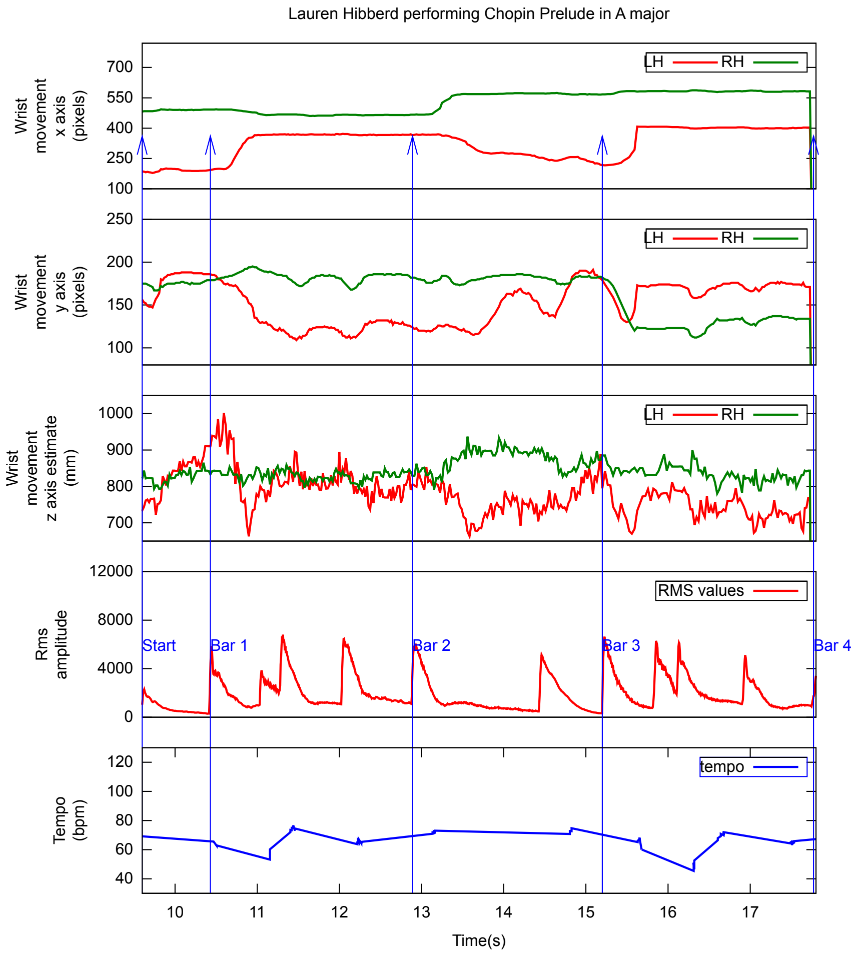


Figure 9.15: Wrist Motion, Tempo and Dynamics for Lauren Hibberd, Prelude in A Major, with the first phrase running from the first blue vertical line until two-thirds through bar 2 and the second phrase running from this point until two-thirds through bar 4.

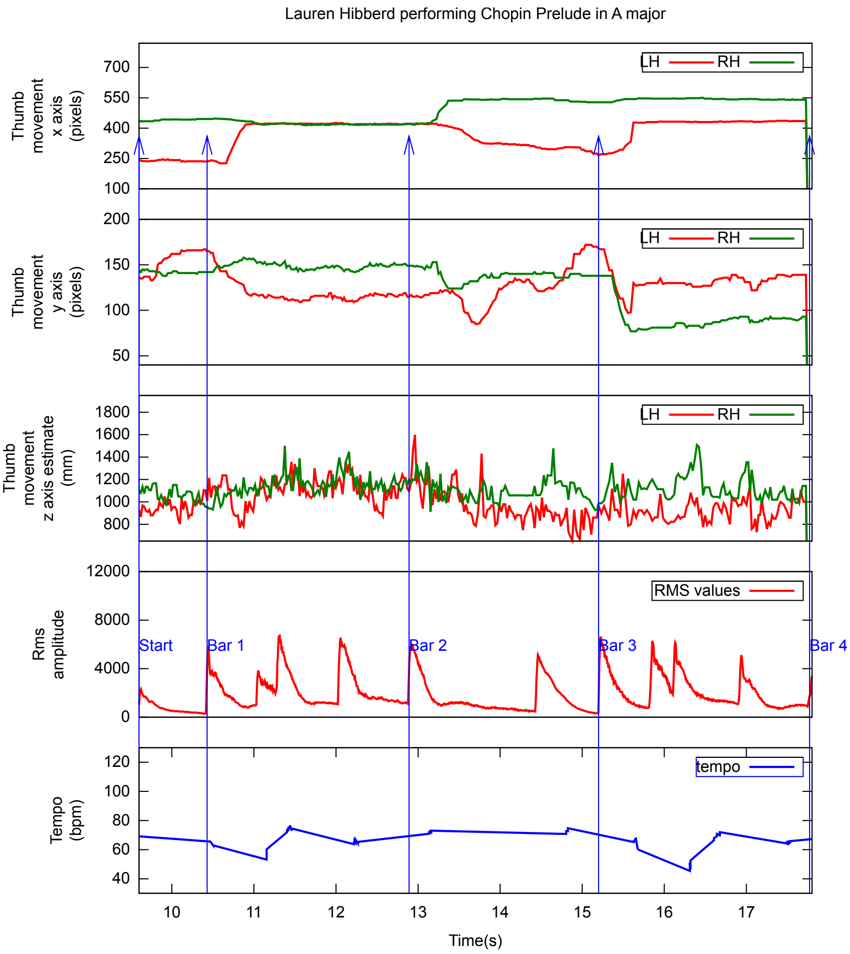


Figure 9.16: Thumb Motion, Tempo and Dynamics for Lauren Hibberd, Prelude in A Major, with the first phrase running from the first blue vertical line until two-thirds through bar 2 and the second phrase running from this point until two-thirds through bar 4.

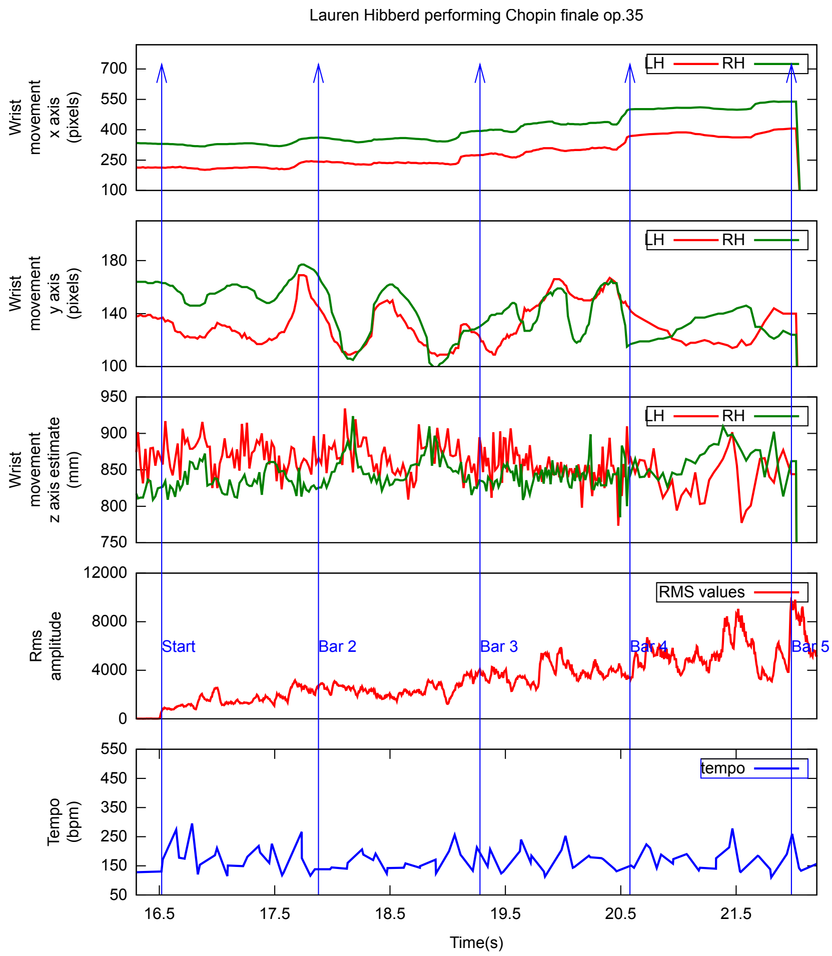


Figure 9.17: Wrist Motion, Tempo and Dynamics for Lauren Hibberd, performing the Chopin finale

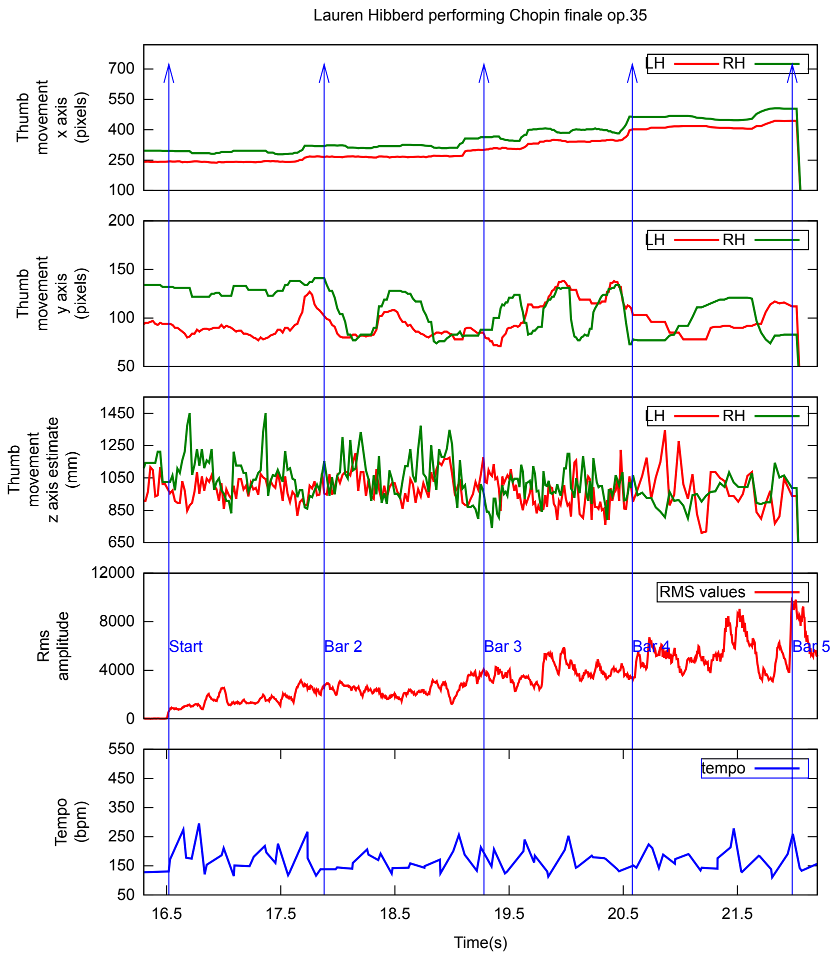


Figure 9.18: Thumb Motion, Tempo and Dynamics for Lauren Hibberd, performing the Chopin finale

quaver of each group of six. This does not fluctuate much on the approach to bar 5. From these parameters we could infer again that this performer does not mark bar 5 as a definitive phrasing boundary.

These results were similar for most performances with one exception being Simon Coverdale whose elongated notes in bar 5 matched with a ritardando and diminuendo on the approach suggest the presence of a phrasing boundary. Fali Pavri and Carlisle Frank showed similar increases in dynamics much like Lauren Hibberd's performance, however, this was not matched by similar fluctuations in tempo. This in-depth note analysis presented by the database alongside graphs of 3D motion in the pianists' fingers has allowed us to observe the particularly accented or 'stressed' notes in an effort to elucidate the structure being performed. Analysis of movement in the wrist and thumb markers have indicated particular accents on notes in both performances of the prelude and finale, confirming hypothesis 9.1. Measurements of height appear to relate to the dynamics of the resultant notes in the performance, however, as the z axis is an estimate, we cannot outrightly confirm hypothesis 9.2. From these continuous results of movement, tempo and dynamics, the maxima and minima of each dataset are examined for their position in the score.

9.2.4 Statistical Results

Other measurements taken over the complete set of data were those such as average tempo. This was calculated from the median of the inter-onset intervals and gave a tempo range for performances of the prelude ranging from 40 bpm (Carlisle Frank) up to 72 bpm (Simon Coverdale). The same calculations from the first five bars of the finale produced a tempo range from 161 bpm (Simon Coverdale) to 220bpm (Fali Pavri). The mode of the tempo measurements were compared against the median to give a measure of expressiveness, as performed in Repp's study on expressive timing [101]. Fali Pavri's performance was indicated as the most expressive by this measure and the least expressive was Simon Coverdale's. However, this calculation fails to address expressiveness at particular points in time and whether fluctuations in tempo occur at important points in structure.

In the same pattern as the previous experimental chapter, the statistics of the spread of parameters are examined. For each note onset value, an rms amplitude value, a tempo value and a motion value are extracted. For the motion value the y axis of the left hand wrist marker is used to show movement towards and away

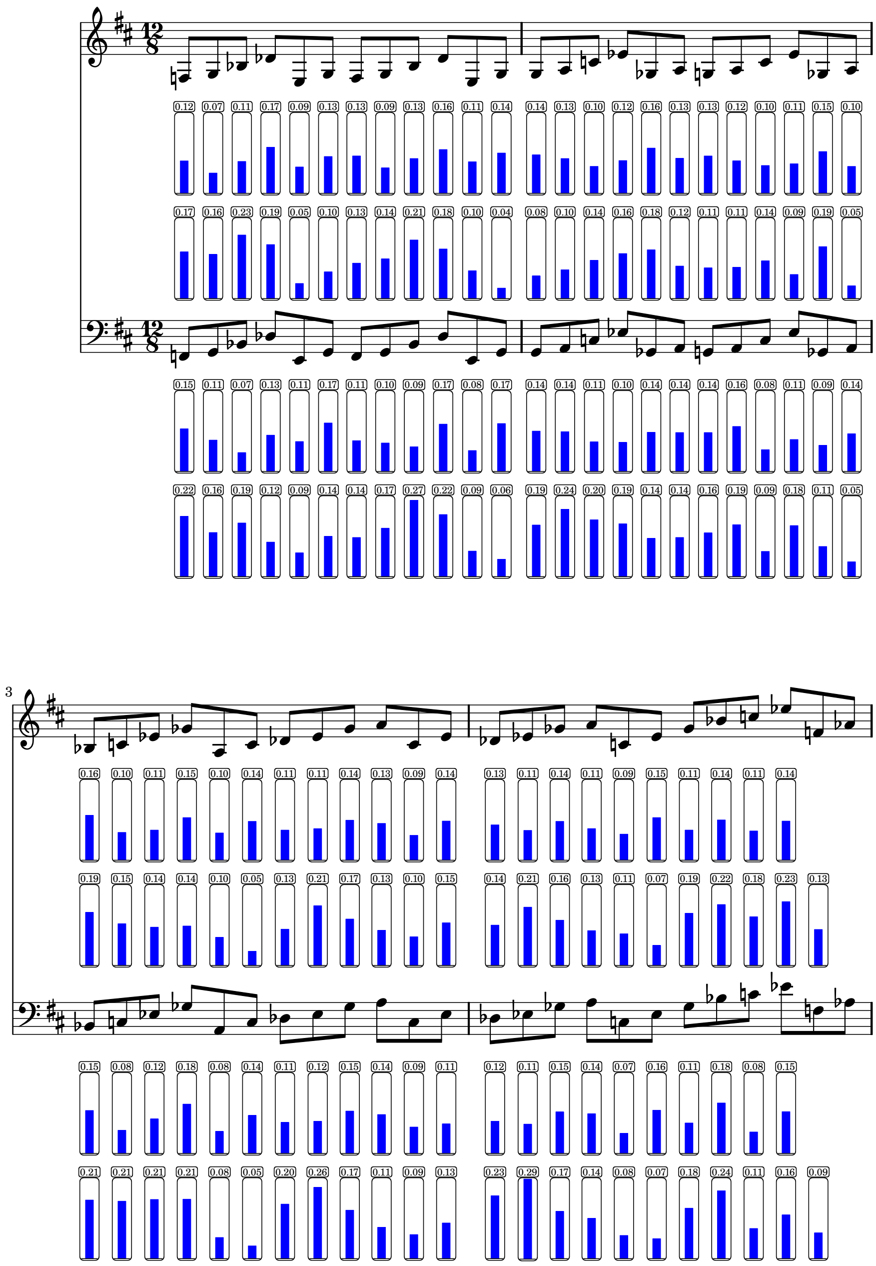


Figure 9.19: Database Results Page 1 for Lauren Hibberd, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

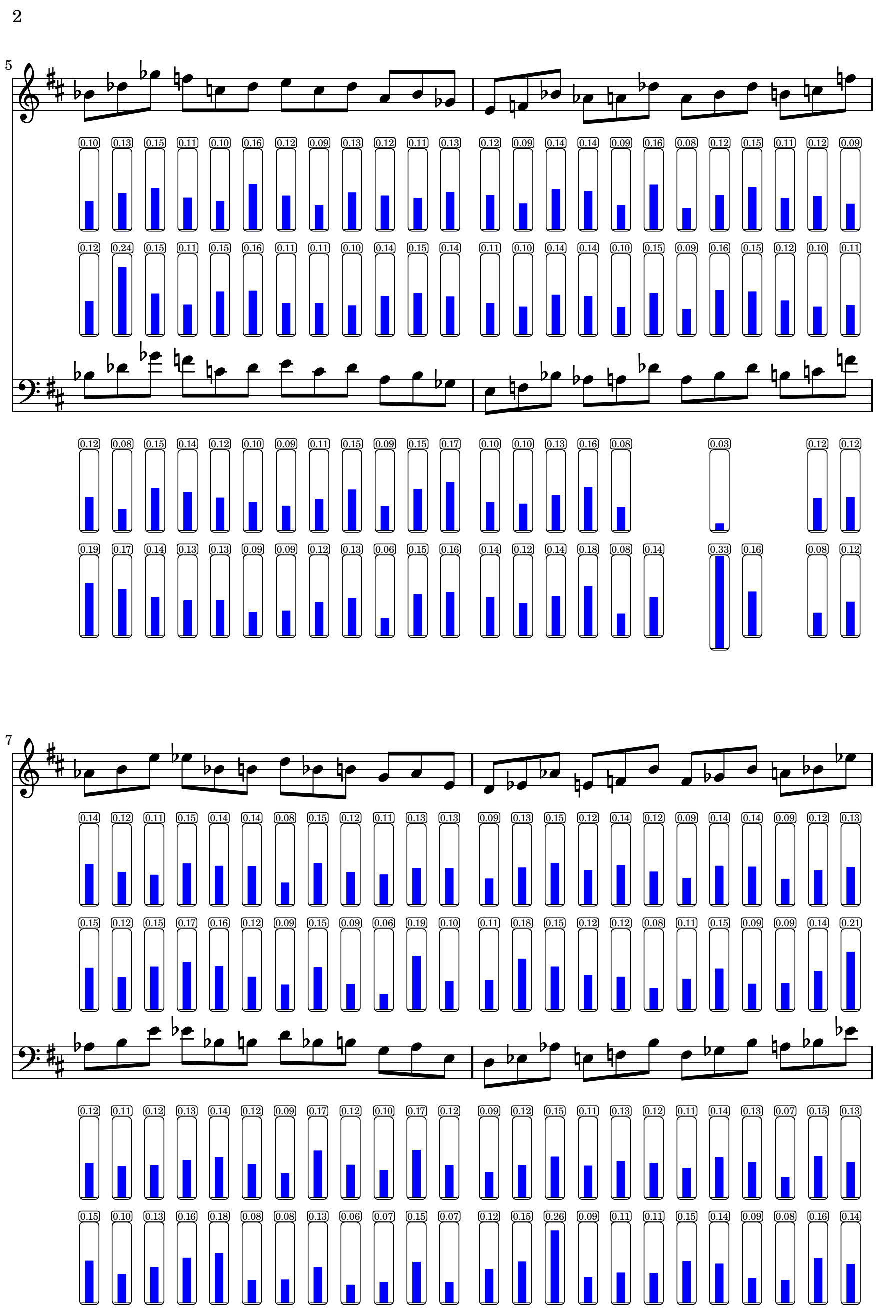


Figure 9.20: Database Results Page 2 for Lauren Hibberd, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

from the keyboard. Plotting these results as box-plots for each performer, this allows observation on each performer's use of these parameters throughout each piece. Each measurement is normalised for each performer.

The tempo for performances of the prelude tends not to fluctuate too wildly as each performer has a fairly limited spread of results as seen in Figure 9.21. Motion is varied more often in general than tempo and dynamics, however, Martin Jones and Simon Coverdale appear to have a skewed distribution. For dynamics however, Martin Jones and Lauren Hibberd have a more normal distribution compared to the other performers. Performances of the finale invoke a more similar use of performance parameters across performers, as seen in Figure 9.22. Notably one would expect the prelude to have far more expressive movement which may be true considering the release of notes, but this is not captured in this extracted dataset.

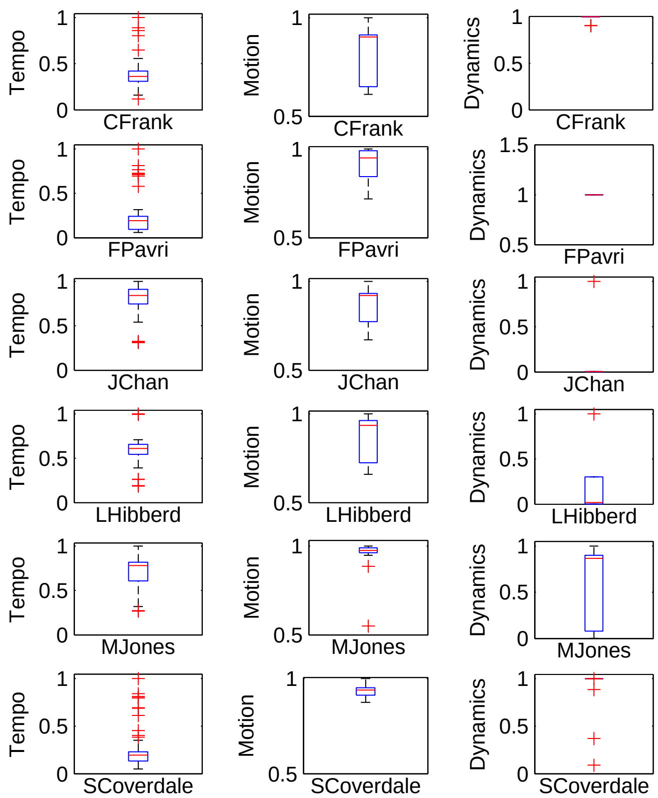


Figure 9.21: Box-plots for all Six Performers measuring Tempo, Motion norm and Dynamics used in Performances of Chopin's A major Prelude

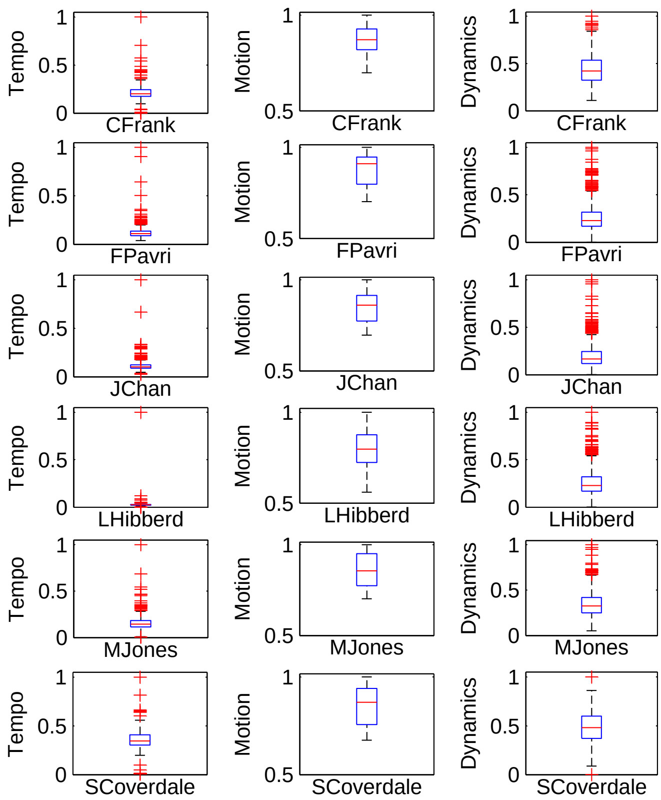


Figure 9.22: Box-plots for all Six Performers measuring Tempo, Motion norm and Dynamics used in Performances of Chopin's B flat minor sonata finale

To produce a representation of how these parameters are used in accenting particular notes, the outliers for each dataset for below the 5th percentile and above the 95th percentile are highlighted in the following scatter plots in Figures 9.23 ,9.25 and 9.27, and then plotted on the appropriate place in the score in Figures 9.24, 9.26 and 9.28. These measurements are only taken for the first five bars of the finale.

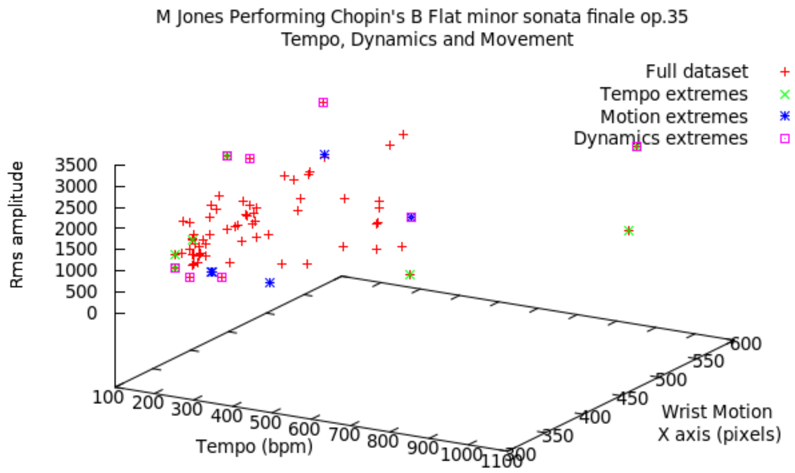


Figure 9.23: Scatter Plot Showing Extremes in Tempo, Dynamics and Motion for Martin Jones performing the Chopin finale

The scatter plot for Martin Jones' performance of the finale (seen in Figure 9.23) shows quite a few combinations in motion(blue) and dynamics(pink) as well as tempo(green) and dynamics(blue). These translate into the annotated score shown in Figure 9.24 by highlighting the halfway point in each bar, particularly in the left hand. These accents appear to be more rhythmical than anything entirely structural.

Jessica Chan's performance is characterised by the scatter plot shown in Figure 9.25. An interesting point to note is the location of the minima and maxima of the motion parameter. The minima tend to occur in the lower half of the tempo range, whilst the maxima appear to occur within the upper half. This upper half are also characterised by larger rms amplitude values than the minima. These measurements may reflect the grouping wrist and tempo movements seen in the earlier time graphs. The location of these maxima and minima in correspondence with the musical score is seen in Figure 9.26. Again the beginning of the finale is well accented across tempo, dynamics and motion, with further combinations of

Music engraving by LilyPond 2.12.2—www.lilypond.org

Figure 9.24: Annotated Score for Martin Jones' performance of the Chopin finale, noting extremes in tempo(T), dynamics(D) and motion(M)

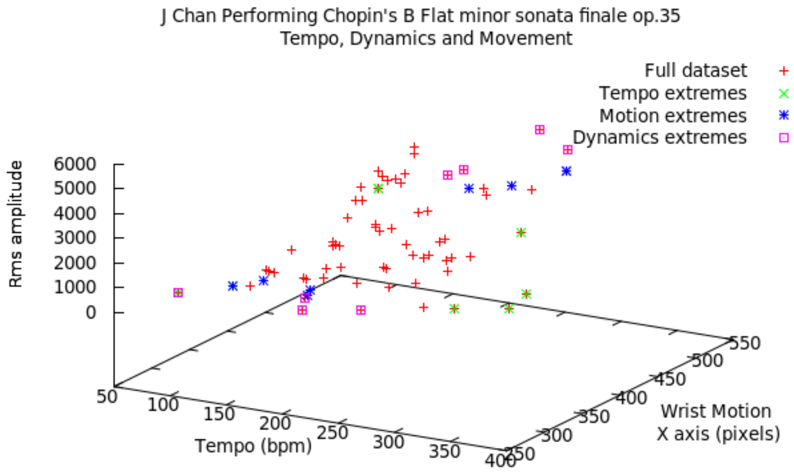
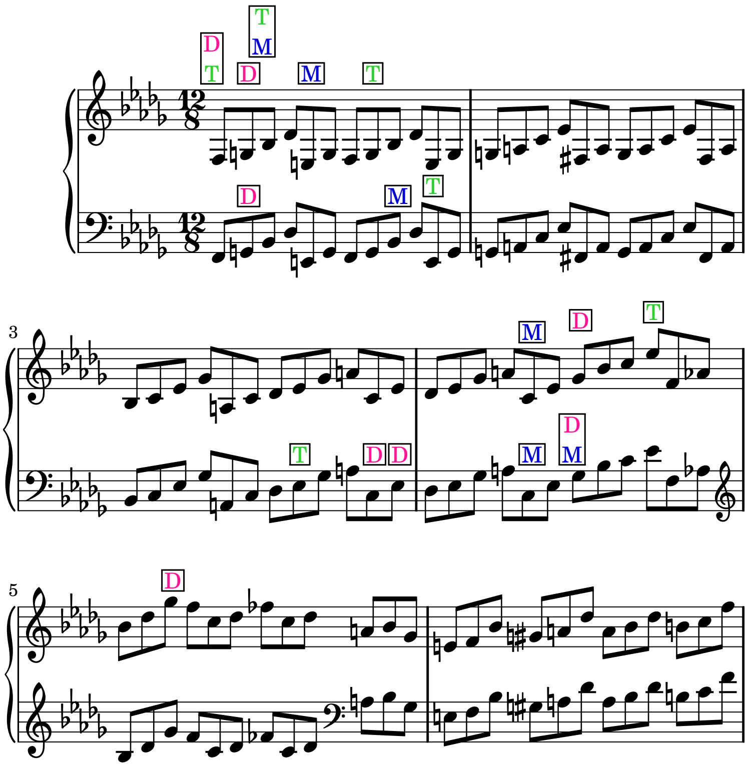


Figure 9.25: Scatter Plot Showing Extremes in Tempo, Dynamics and Motion for Jessica Chan performing the Chopin finale

parameters occurring halfway through bar 4.

The scatter plot for Lauren Hibberd's performance seen in Figure 9.27, shows combinations of extremes in parameters occurring particularly between motion and dynamics with the odd pairing between tempo and dynamics extremes. There are no evident clusters of these maxima and minima in terms of all motion, dynamics and tempo and so the location of these noted in the scatter plot is now considered in the annotated score in Figure 9.28. The beginning of the finale is heavily accented with extremes in tempo, dynamics and motion and again in the second half of bar 4.

Each of the scores show accents at the beginning of the piece and some near the end of bar 4 which would suggest the beginning of a new phrase at bar 5, however as this is not a strong result, it suggests that these parameters reflect the change in composition and not in fact a strong phrase boundary.



Music engraving by LilyPond 2.12.2—www.lilypond.org

Figure 9.26: Annotated Score for Jessica Chan's performance of the Chopin finale, noting extremes in tempo(T), dynamics(D) and motion(M)

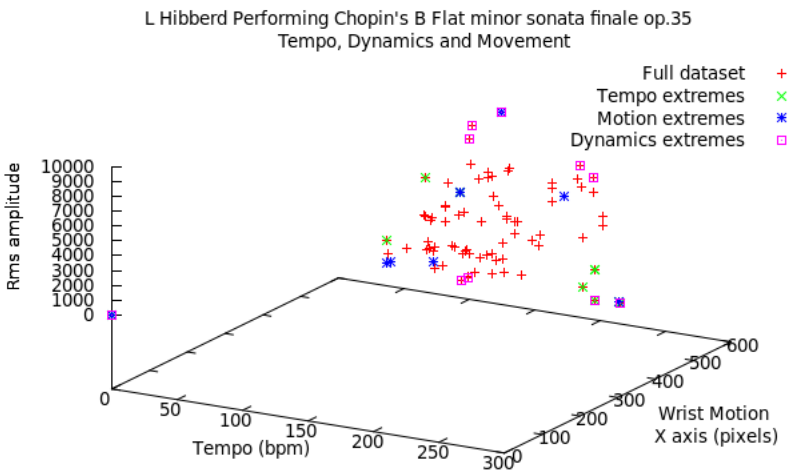
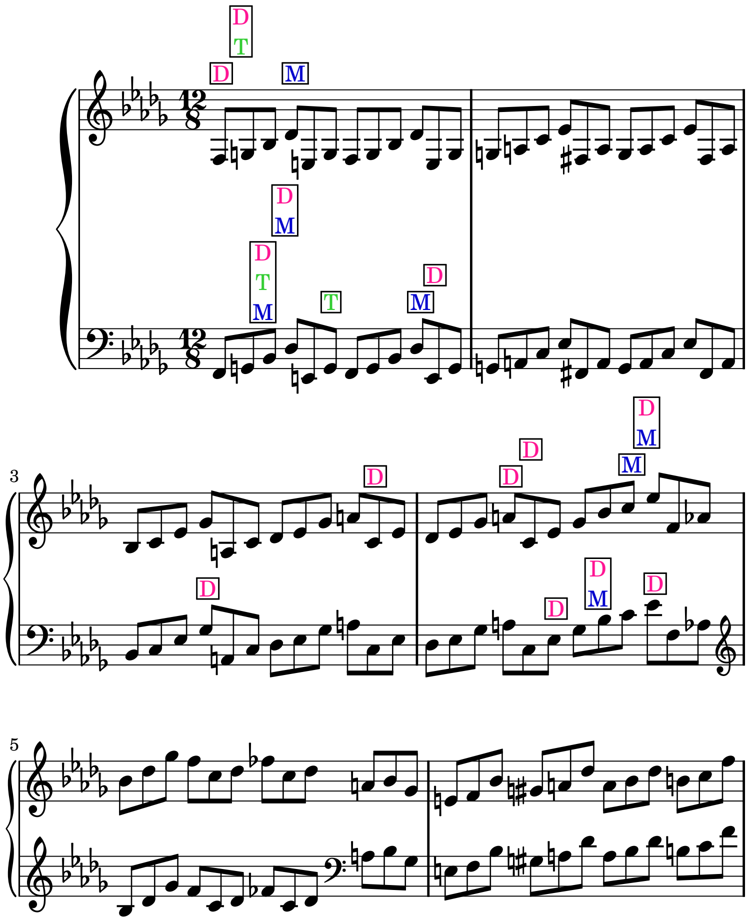


Figure 9.27: Scatter Plot Showing Extremes in Tempo, Dynamics and Motion for Lauren Hibberd performing the Chopin finale



Music engraving by LilyPond 2.12.2—www.lilypond.org

Figure 9.28: Annotated Score for Lauren Hibberd's performance of the Chopin finale, noting extremes in tempo(T), dynamics(D) and motion(M)

9.3 Exploring Finger Curvature

An advantage of using the finger motion capture system is that we can also examine curvature of fingers as they are used to play each note. For this particular question, the curvature of the thumb and the second finger are examined for the finale. These are calculated as distances between the x,y coordinates of the metacarpophalangeal and the proximal phalanx for the thumb and first finger, and the distance from the proximal to the distal phalanx of the first finger. In the graphs for each performer in Figure 9.31 for Lauren Hibberd, Figure 9.30 for Martin Jones and Figure 9.29 for Jessica Chan, an increase in each of the three graphs for curvature indicates that the finger is becoming flatter and parallel to the keyboard. A decrease indicates that the finger is becoming more curved.

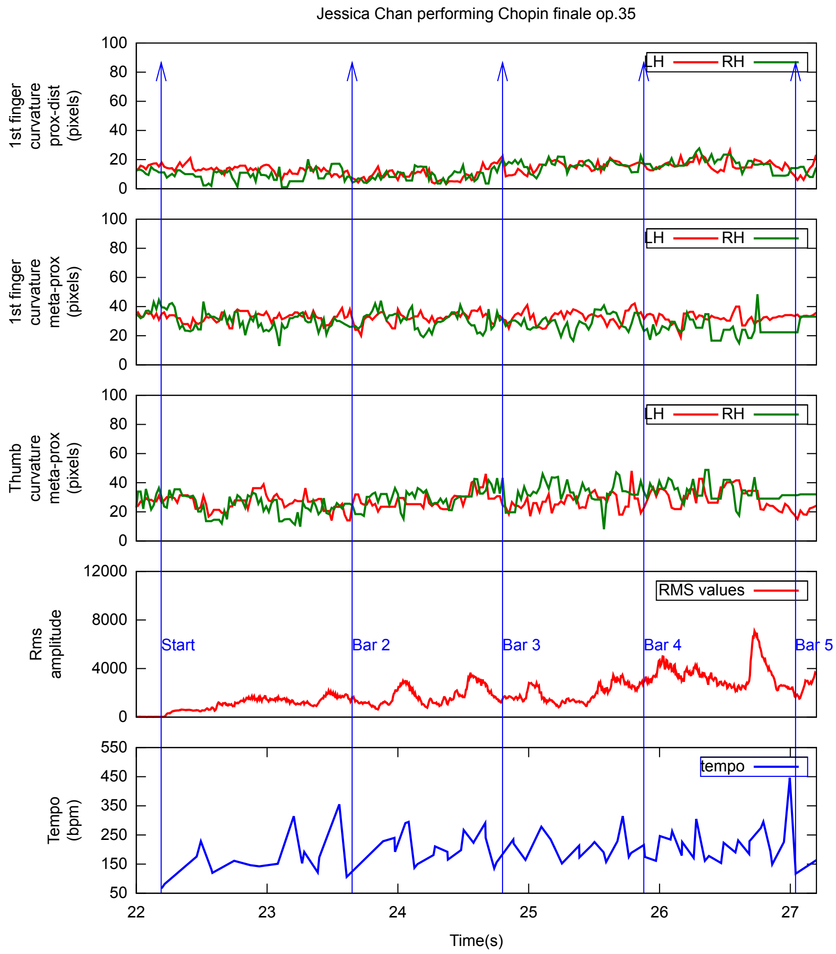


Figure 9.29: Finger Curvature, Tempo and Dynamics for Jessica Chan, performing the Chopin finale

Jessica Chan demonstrates a style of playing in which she moves her hands around in each of the three axes extraneously to the movement required to physically play each note. Seen in the previous graphs marking the coordinates of wrist markers, a 'releasing' action is seen often in the prelude, and this is carried into the finale despite the dramatically different tempo. For this performance as seen in Figure 9.29, the thumb curvature characterises this movement in the first bar, where a repeating pattern is seen for the twelve quavers, separating into two groups of six.

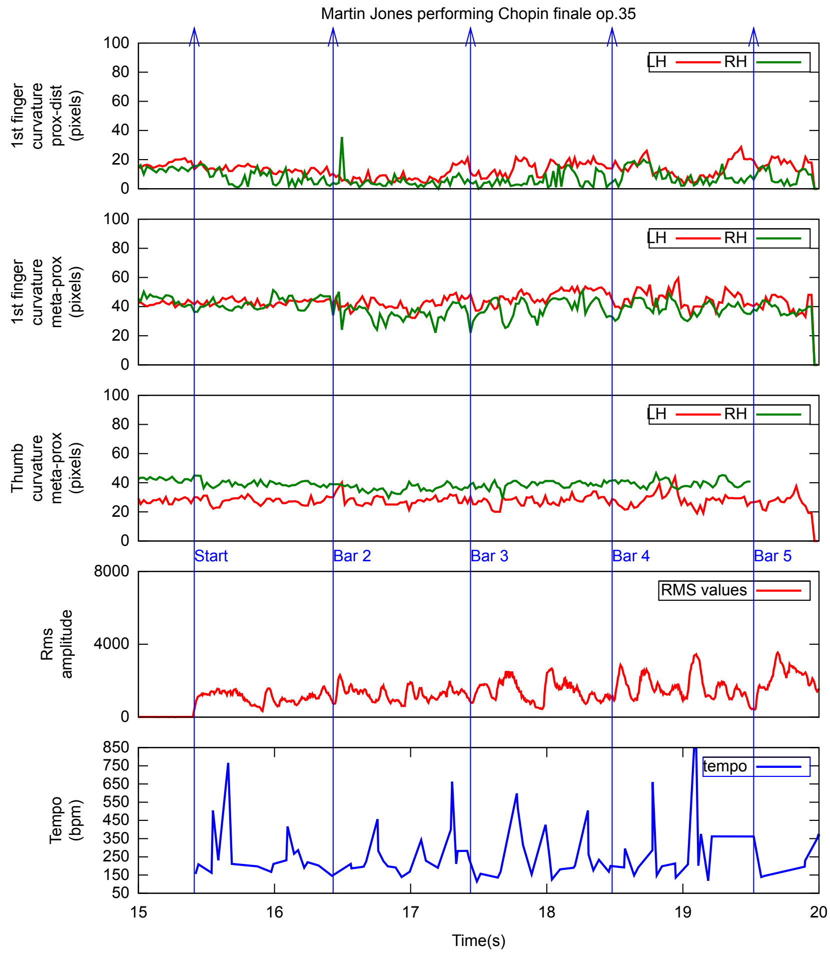


Figure 9.30: Finger Curvature, Tempo and Dynamics for Martin Jones, performing the Chopin finale

In contrast, Martin Jones keeps his fingers flat whilst playing the first bar which is demonstrated in Figure 9.30 by negligible differences in curvature. The thumb is kept mainly flat for the next few bars, whilst the curvature for the first finger shows clearly where notes are performed using this particular finger. The differences in curvature for these performed notes are negligible suggesting that he uses his first finger in the same way for each note. Using a flat thumb and a curved first finger suggests Martin Jones may be using the right hand thumb to emphasise the first and fifth quaver in each group of six as an underlying melody.

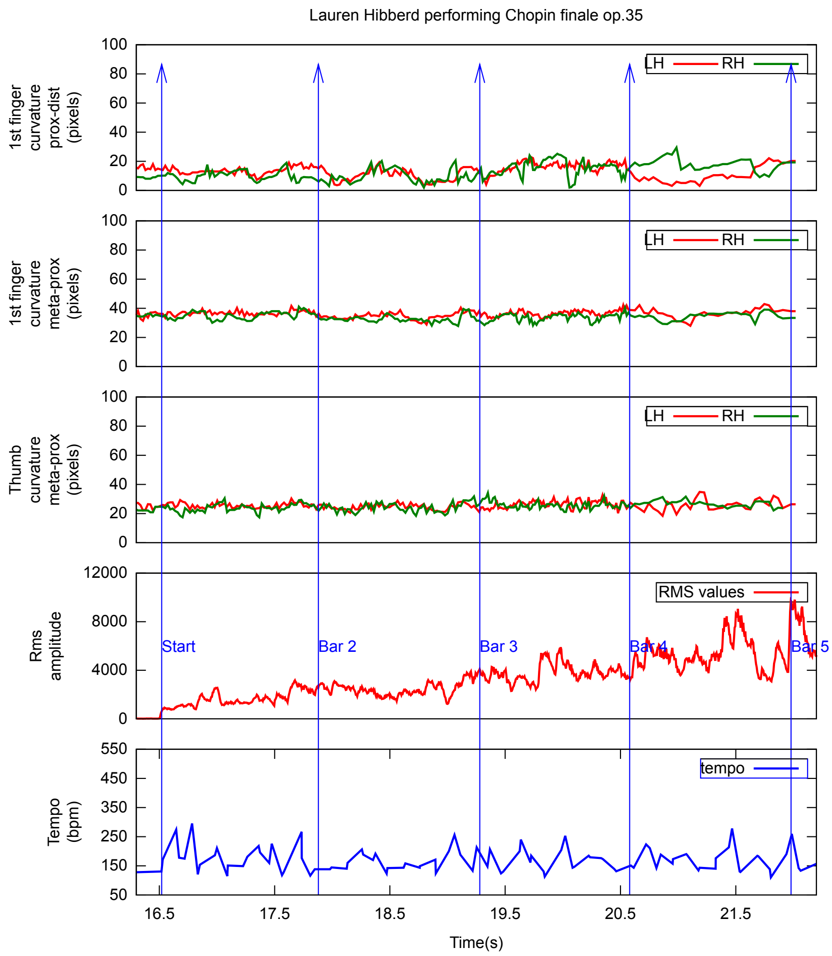


Figure 9.31: Finger Curvature, Tempo and Dynamics for Lauren Hibberd, performing the Chopin finale

Lauren Hibberd is another performer that keeps her fingers relatively flat whilst performing the finale, again which can be seen by the curvature plotted in Figure 9.31. The fingers remain fairly flat throughout the piece whereas the tips of the first finger between the proximal and distal interphalangeal change in curvature for where the notes need to be performed. An interesting point to note is that the curvature of the fingers remains constant throughout the crescendo in amplitude of the sound wave suggesting that curvature is not a direct factor for loudness of each note.

These results demonstrate the ability of the finger tracking system to glean information on performer playing styles and also structural information intended by the performer. The differences between each performer is clearly visible in the changes in curvature for each finger. Further investigation would involve each of the fingers' curvature and attempt to align them to the performed notes.

9.4 Conclusions

Quantitative measurements of aural and visual parameters in performances of both Chopin's Prelude in A major Op.28 No.7 and B flat minor sonata finale movement Op.35 reveal structural information from the manipulation of tempo, dynamics and finger movement. This is used to analyse a point of disagreement amongst traditional analysis on the importance of bar 5 in the finale as either a continuation of the initial theme starting at bar 1 or the beginning of a new phrase marking the first four bars as simply an introduction.

The beginning of this chapter again stated some hypotheses relating to how performers used these parameters of tempo, dynamics and finger motion to project structural ideas. Hypothesis 9.1 stated that trajectories of finger motion in the x, y and z axis would reflect expressive accents within the phrase and Hypothesis 9.2 stated that it was expected that wrist motion that reflects movements toward the soundboard of the keyboard, and movements toward the key-bed will produce high values of rms amplitude.

Continuous measurement and display of these parameters against time allows closer observation of fluctuations at particular structural points in each piece. From the multi-modal graphs of wrist motion, dynamics and tempo, trajectories of the y and z axis components reveals information about note groupings and general phrasing. For all recorded performances of the finale, it is evident that each performer groups the quavers into sixes. For particular performances such as Martin Jones', the change in composition halfway through bar 3 where the pitch changes every six quavers instead of twelve in bars 1-2, is marked by accents in tempo and dynamics. This confirms that Hypothesis 9.1 is correct. From comparing performances of the prelude and the finale, results show that five out of the six performers suggest that there is a boundary at bar 5, however, it is not a highly important one in terms of structure.

Hypothesis 9.2 appears to be rejected as the observations for performers' wrist movements moving towards the keyboard do not seem to coincide with increases in rms amplitude. As the estimations for the z axis were not entirely accurate but more a reflection of the height movement of each finger, a confirmation of the hypothesis in this respect would have been speculative. However, further investigation is warranted into the expressive movements of fingers throughout piano performance and their close relationship with audio parameters.

From statistical analysis we see that the rarest occurring values for each parameter occur at specific points in the phrasing structure, which when applied to performances of the finale, mark bar 5 as a change in the composition, but not a complete change in phrase as would be expected for the introduction of a new theme. Comparing the results from these different types of analysis confirms the interpretation of phrasing structure.

The methodology used shows that very intricate details regarding how performers play each note can be extracted from performances and used to indicate structure even in pieces where the structure is ambiguous.

Improvements on this system could be made in the alignment of the video parameters to the audio stream. The raw output video from the capture camera could be altered to include time-stamping allowing more accurate alignment of gestures to notes. Viewing the curvature of each finger, automatic detection of notes being played could be programmed in order to better align the gesture with the beginning and end of each note. This small scale analysis performed here could be run for the entire piece, for many more performers and many more pieces used for control. In comparison with the statistical analysis for particular accented notes, structure can be more easily detected. Predicting structure in unknown pieces is therefore possible.

Chapter 10

Discussion

In performing the research undertaken in this thesis, a number of issues concerning work of such an interdisciplinary nature have been noted. A major issue in this kind of research comes from the collaboration between engineers and musicians. Some artists are of the opinion that engineers work solely in numbers and so any analysis is reducing 'art' to streams of data. Although we can measure these parameters, the Gestalt theory that these things acting together are larger than the sum of their parts, would imply that we can never truly measure the 'essence' of music. Music is a phenomena that affects us all in different ways and in a way is completely subjective. However, just as fine art can be described in terms of form, line, colour and other parameters, music can also be described in terms of harmony, rhythm, pitch, timbre etc. There are ways to break down these forms of art. Measuring timings etc. is a lower level descriptor of 'performance' but also allows the measurements to be taken of an instance of music rather than a score.

Essentially I am applying empirical measurement techniques to musical data informing comment on attributes that can be of an entirely subjective nature. Particularly in classical music where structure can be complex and hierarchical, the phrasing in some compositions cannot be entirely agreed upon by human experts never mind by computational methods. This may be the reason behind moves in the machine learning field into analysing popular music for structure, where a verse and chorus construction may be easier to distinguish by all parties. However, the classical music genre still represents an example of how complex musical structure can be and the methods coined for popular music may be unlikely to be successfully applied in all cases.

The methods of analysis used in this thesis for such a purpose as determining structure in these more complex cases aim to balance an empirical scientific approach with the subjective musical context. This has proved particularly important considering the subjectiveness with which fluctuations in aural and visual parameters are produced by the performers. As quoted in Chapter 2.1, Eric Clarke observes that fluctuations in tempo can be used for different purposes depending on the structure of the piece. Visual parameters are also affected by the movements necessary for basic note production. Separating movements in terms of function is complicated considering some gestures may be multi-functional. Therefore, the analysis has aimed to not discard any information which may pertain to the motor movements required for note production, but to include them in the analysis. In terms of phrasing, this would lessen the chance of movements lining up exactly with these larger chunks of notes instead of individual notes or chords, and so the results found will likely be more accurate.

When dealing with the multi-modal streams of data produced by the recordings, care is taken when choosing methods of measuring the relationships between them. A lot of statistical tests determine whether data is related in direct ways such as increasing tempo when there is increasing dynamics. However, with the subjective nature of performance, and the manipulation of parameters changing differently depending on the structural function, these factors will not always be changing in the same way over time. To compensate for this, the tests used include determining how regularly troughs in motion norm occur close to a phrasing boundary, and more emphasis has been placed on graphs of the multi-modal parameters plotted in time. What is required are methods of analysis which can take into account the large variability of each of the parameters but still recognising that there are certain fixed parameters such as pitch and structure.

An issue in including the measurement of physical gestures alongside aural parameters occurs in the alignment with the notes from the original score. As gestures are multi-functional, and often can be for necessary purposes as well as for expression, it is difficult to separate these from purely expressive gestures. For this reason, it is also difficult to align gestures to singular notes. This can make direct comparison throughout aural and visual domains complicated.

Drawing back from analysis, a bigger question to ask is whether performers actually intend the manipulation of parameters such as tempo, dynamics and movement as a communication of musical structure for themselves and/or the audience. In an attempt to examine the differences between phrasing boundaries highlighted by changes in performance parameters, and phrasing boundaries identified by audience judges, videos created from the first experiment were used as stimuli and audience judges were asked to denote phrase shaping by moving a slider. There seemed to be very little difference in boundaries for performances of the Prelude in A major op.28 no.7 but this could be partially explained by its strict explicit structure. We cannot tell exactly how these audience judges are making their judgements, particularly in audio-only presentations. Detecting phrasing computationally then becomes more a question of how structure is reflected by these certain parameters instead of trying to imitate how audiences perceive it.

The experiments in this thesis are a unique comparison between various composed pieces. A control piece has been used in an effort to benchmark variations in each performer's style of playing. Examination of solely the control piece has demonstrated various styles of performer playing even when conveying the exact same structure. This has been confirmed by multi-modal explorations of tempo, dynamics and movement trajectories which show decreases and increases in varying combinations at phrasing boundaries. This, along with the examination of the extremes of these trajectories at their corresponding occurrence within the score shows the fastest/slowest tempi reserved for particularly important structural points for the example pianists. These example performance measurements reflect completely different performances of the same piece, yet when looking at these particularly highlighted points in the score, there are many agreements. These results coincide with suggestions by Repp [100] that different expressive strategies are not necessarily produced by different structural interpretations, and this has been seen within the research in this thesis to extend for physical gestures.

The comparison between two pieces in Chapter 8 shows that certain elements of performer style are carried over from the control piece, yet there are also many differences. The two pieces are composed by Chopin in a similar style with similar rhythmic repetition, albeit dissimilar rhythms. Differences are evident in motion norm between pieces for the same performer with the leading markers also changing between pieces. This suggests the rhythmic make-up of the phrase may have far more influence on the motion trajectory, something previously stated by Wanderley [125]. Similarities still occur with motion peaks and troughs occurring at phrasing boundaries but for these longer, expanded phrases in Prelude 6, subchunking is sometimes present. In both of these pieces, it is evident that whilst performer movement style can be widely differing, the underlying motion norms conform to the same structure, confirming Hypothesis 8.1. Using this method of multi-modal analysis for tempo, dynamics and motion, all nine performers in this experiment produce a similar structural interpretation of eight phrases for Prelude 7 and the first five phrases in Prelude 6. The second part of Prelude 6 cannot be subjected to direct comparison between performers. Progressing from these results, further studies could include examining these different interpretations, however, a method of accurately determining these interpretations from the performers themselves must be developed. Within these phrasing shapes of tempo, dynamics and motion, there are different sub-shapes which may be reliant on how each performer is accenting the notes within the phrase. Examples of sub-phrase analysis are seen in the statistical analysis of the extremes of each parameter. From deducing their position on the score, all nine performers use local maxima and minima to determine the accents within each phrase. Particular points of interest are characterised by extremes in all three parameters. Further exploration into other parameters such as articulation and timbre would be expected to produce similar results. These suggest that each performer draws our attention to interesting points in structure in a form that is comparable across a number of pieces.

The exploration into interpretations of the finale in Chapter 9, uses more intricate detail to determine the importance of each note as it is played in the overall picture of the opening bars, but works on the theory that performer 'styles' can be used to discover structure. This in a sense could be done for the second half of Prelude 6 in the previous experiment. For each of the six pianists, wrist and thumb motion is examined in all three axes alongside tempo and dynamics. This produces results which can determine phrase shaping and even note 'groupings'. Also evident from performances of the prelude is the 'releasing' action with which pianists tend to play loud chords. Accents in note duration and inter-onset interval are apparent at the beginning of the finale, however, in most of the performances, we do not see these accents repeated at the beginning of bar 5. This suggests that bar 5 is regarded as not the entry of a new theme but the continuation of the theme beginning at bar 1. These results demonstrate a method of detecting structure purely from performance parameters that could be used without a priori understanding of the musical structure itself.

Finally, the study of finger curvature enabled by the use of the finger tracking system FingerDance 4, produce results which correctly identify the style of 'touch' used by each performer in performances of the prelude and the finale. The example of Lauren Hibberd and Martin Jones who both use a flat fingered approach to the finale against Jessica Chan's more curved approach immediately allow us to examine the differences between different touches and the different accents they produce.

Chapter 11

Final Conclusions

The two main aims produced at the start of this thesis were to

Aim 1: design capture systems, storage and visualisation formats that allow accurate and robust methods of recording live performances and display the results in a useful way for musicological analysis.

Aim 2: to determine whether structure can be elucidated from the empirical analysis of multi-modal performance parameters.

The first aim was satisfied in the first half of the thesis which detailed the design of multi-modal systems from a selection of proprietary products as well as specially designed ones. In order to satisfy the need for a cheap, portable and inexpensive motion capture system, Chapter 4 detailed an accurate finger motion capture system which operated with the least disturbance to the performer. This used UV paint dots as passive markers in an image processing based system. This system estimates 3D positioning within a margin of 1.66mm error and can also provide information on finger curvature. Chapters 6 and 5 demonstrate how the movement capture system alongside other multi-modal capture systems can be used to produce and store multi-modal information and display queries above a musical score. This is in such a format as to be incredibly useful for musicological analysts.

The experiments that followed in Section III used these tools to highlight how structure can be detected and in some cases predicted from the fluctuations in performance parameter data, thus satisfying the second main aim of the thesis.

Within these experiments, the following hypotheses were made:

Hypothesis 8.1 Regardless of the subjective and personal nature of physical gesture in relation to musical structure, there will exist an underlying pattern that is related to phrasing and is common across all performers.

Hypothesis 8.2 The underlying motion profile of the performer related to phrasing will be the same across pieces.

Hypothesis 8.3 When investigating the role of gesture in multi-modal detection of phrasing, a combination of aural and visual parameters will provide the most accurate indicator of phrasing.

Hypothesis 8.4 Where combinations of global maxima and minima occur in both aural and visual streams of data, these will be related to the most important structural features of the composition.

Hypothesis 9.1 Trajectories of finger motion in the x, y and z axis will reflect expressive accents within the phrase.

Hypothesis 9.2 It is expected that wrist motion that reflects movements toward the soundboard of the keyboard, and movements toward the key-bed will produce high values of rms amplitude.

From the gestural motion studies conducted from performances of two Chopin Preludes in Chapter 8, it was shown that despite the idiosyncratic nature of the performers' gestures in performances of both preludes, the underlying motion norm suggested the same phrasing structure. This was confirmed by measuring the local maxima of the motion profile between phrases for each pianist. These local maxima occurred reliably at the same point for each phrase for each performer. These patterns were evident across all performers despite their background and ideas on movement within performance. This confirms Hypothesis 8.1. Correlating each performer's patterns of motion profile across their performances of the Preludes, it is shown that few result in a high correlation. Some even result in negative correlations. This suggests that the motion profile for each performer changes depending on which piece they are performing. This rejects Hypothesis 8.2. Factors for this may be due to changes in rhythm, melody or harmony, however, seeing as the rhythmically repeating phrases of Prelude 7 tend to produce similar motion patterns for each performer, it suggests that motion may be highly linked to rhythm.

As each performer 'style' is different and the use of these parameters can be varied according to the position on the score, it becomes apparent through observation of the multi-modal graphs that a combination of parameters indicates phrasing boundaries. An example of this is clearest at the harmonic arrival between phrases 5 and 6 where global maxima and minima in motion, dynamics and tempo coincide. This suggests that Hypothesis 8.3 is correct. When examining the maxima and minima of the dataset and their occurrence in the musical score, it is clear that performers tend to use combinations of these extremes at important points in structure, suggesting that Hypothesis 8.4 is correct. The location of these extremes in motion dynamics and tempo occur at particular accents of harmony, melody and rhythm set out by Parncutt.

Using this knowledge to then try and predict musical structure from performance nuances, the second experiment analyses professional performances of Chopin's B flat minor sonata op.35 finale movement. Looking at intricate finger movement (as the piece is performed at the fastest limits of technical ability), we can see patterns of how notes are grouped and accented. When added to information on tempo and dynamics, this provides an interpretation from which we can glean structural issues such as the interpretation of bar 5 as not the introduction of a new theme but the continuation of the main theme introduced at bar 1.

Measurements of wrist movement in the x, y and z axes, throughout performances of the finale show certain accents defined by peaks and troughs that occur simultaneously with accents in tempo and dynamics. When located on the score of the performance, these appear at points which reflect particular harmonic and structural changes. This confirms Hypothesis 9.1. Measured movements of the wrist towards and away from the keyboard do not necessarily coincide with increases in rms amplitude and as the z axes is a estimation, Hypothesis 9.2 cannot be confirmed. These measurements of motion, tempo and dynamics provide insight into the structural choices of the six professional performers when considering the finale, and allow the conclusions to be drawn on the particular ambiguous boundary of bar 5, something which cannot be achieved by traditional score analysis alone.

Such research into how performers highlight structure with these parameters has major benefits for piano pedagogy and implications for computational methods of detecting structure such as in the field of music information retrieval.

As noted in Section 5, there have been many developments noted for these systems and for the multi-modal analysis techniques. The most pertinent of these I believe lie with the development of the finger tracking system and the alignment of physical gestures with aural parameters, for both analysis and visualisation, and to investigate more thoroughly the role of physical gestures in music performance.

A long debate has been waged between scientists and musicians over how the finger strikes the key manipulates the resulting sound. Contrary to the belief that the only variable can be key velocity, a direct measure of the force applied to the key, pianists claim that the shape of the hand i.e. flat versus curved fingers alters not just the loudness but also timbre. The Fingerdance software in its developed form could be pivotal in answering these questions alongside physical modelling of the piano itself.

The development of methods such as these for automatically detecting structure must be cultured in a way which respects the context of the music being analysed and the subjectivity of the performances. A highly-integrated approach to computational methods is required, which constantly refer to musicians' interpretations and analyses of structure. Only in this way will automatic detection be completely valid in all disciplines, and be useful in performing functions pertaining to the analysis of music.

As well as determining that musical structure can be measured from quantifiable expressive parameters, this study has further implications for assisting computational music analysis as well as music information retrieval. Implications for piano pedagogy arise from relating body movement to underlying musical structure as well as the study of the relationship between finger curvature and the resultant acoustic sound. Examining this first step in the communication of musical information from composer through the performer to the audience can also reveal what is conveyed in a musical performance so we can ultimately understand what is being perceived and how.

Bibliography

[1] Ardour - the new digital audio workstation. http://www.ardour.org.

[2] Audacity. http://audacity.sourceforge.nets.

[3] Lilypond - music notation for everyone. http://lilypond.org/.

[4] Mencoder. http://www.mplayerhq.hu/design7/news.html.

[5] Moog piano bar. http://www.moogmusic.com/detail.php.

[6] Pca module for python. http://folk.uio.no/henninri/pcamodule/.

[7] Performance markup language. http://www.n-ism.org/Projects/pml.php.

[8] Rosegarden: music software for linux. http://www.rosegardenmusic.com.

[9] World science festival 2009: Notes and neurons. http://www.worldsciencefestival/2009/notes-and-neurons.

[10] Coriander - gui for firewire camera control and capture, 2010. http://damien.douxchamps.net/ieee1394/coriander/index.php.

[11] Opencv library - library of programming functions aimed at computer vision, 2010. http://opencvlibrary.sourceforge.net/.

[12] Optotrak certus full motion capture system, 2010. http://www.ndigital.com/lifesciences/products-motioncapturesystems.php.

[13] D. Bachmann, M. translated by Parlett. *Dalcroze today: An Education Through and Into Music*. Clarendon Press : Oxford, 1991.

[14] D. G. Barolsky. The performer as analyst. *Music Theory Online*, 13 No.1, 2007.

[15] E. Bisesi and R. Parncutt. An accent-based approach to automatic rendering of piano performance. In W. Goebl, editor, *Proceedings of the Second Vienna Talk on Music Acoustics, Vienna, 19-21 September, 2010*, pages 26–30. Vienna: Institute of Musical Acoustics (Wiener Klangstil), 2010.

[16] P. Brossier. Aubio: a library for audio labelling. http://aubio.org.

[17] B. Buck, J. MacRitchie, and L. Piwek. Perceptual recognition of embodied musical structure. In *Abstract Proceedings of The Musical Body: Gesture, Representation and Ergonomics in Musical Performance, 22nd - 24th April, London, UK*, 2009.

[18] J. Bullock. Libxtract: A lightweight library for audio feature extraction. In *Proceedings of the International Computer Music Conference*, 2005.

[19] A. Burns and B. Mazzarino. *Gesture in Human-Computer Interaction and Simulation*, chapter Finger Tracking Methods Using EyesWeb, pages 156–167. Springer-Verlag, 2006.

[20] A.-M. Burns and M. M. Wanderley. Visual methods for the retrieval of guitarist fingering. In N. Schnell, F. Bevilacqua, M. J. Lyons, and A. Tanaka, editors, *NIME*, pages 196–199. IRCAM - Centre Pompidou in collaboration with Sorbonne University, 2006.

[21] C. Cadoz and M. Wanderley. *Trends in Gestural Control of Music*, chapter Gesture-Music. Iracam - Centre Pompidou, 2000.

[22] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe. *Gesture-based Communication in Human-Computer Interaction*, chapter Multimodal analysis of expressive gesture in music and dance performers. Springer-Verlag, 2004.

[23] A. Camurri, B. Mazzarino, and G. Volpe. Analysis of expressive gestures in human movement: the eyesweb expressive gesture processing library. In *Proc. XIV Colloquium on Musical Informatics, Firenze, Italy*, 2003.

[24] C. Cannam, C. Landone, M. Sandler, and J. Bello. The sonic visualiser :a visualisation platform for semantic descriptors from audio signals. In *Proceedings of the International*, 2006.

[25] E. Clarke and J. Davidson. *Composition-Performance-Reception : Studies in the Creative Process in Music*, chapter The Body in performance, pages 74–92. Ashgate Publishing Ltd, 1998.

[26] E. F. Clarke. Mind the gap: formal structures and psychological processes in music. *Contemporary Music Review*, 3 part 1:1–13, 1989.

[27] E. F. Clarke. The semiotics of expression in musical performance. *Contemporary Music Review*, 17, Part 2:87–102, 1998.

[28] H. Conklin Jnr, A. Askedfelt, E. Janson, D. Hall, G. Weinreich, and K. Wogram. *Five Lectures on the Acoustics of the Piano*. Royal Swedish Academy of Music, 1990.

[29] N. Cook. Structure and performance timing in bach's c major prelude (wtci): An empirical study. *Music Analysis*, 6 No.3:257–272, 1987.

[30] A. Cutti, G. Paolini, M. Troncossi, A. Cappello, and A. Davalli. Soft tissue artefact assessment in humeral axial rotation. *Gait and Posture*, 21:341–349, 2005.

[31] S. Dahl and A. Friberg. Visual perception of expressiveness in musicians' body movements. *Music Perception*, 24(5):433–454, 2007.

[32] J. Davidson. Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, 21(2):103–113, 1993.

[33] J. Davidson. *Musical Performance: A Guide to Understanding*, chapter Communicating with the body in performance, pages 144–152. Cambridge University Press, 2002.

[34] J. W. Davidson. Qualitative insights into the use of expressive body movement in solo piano performance: a case study approach. *Psychology of Music*, 35(3):381–401, 2007.

[35] J. W. Davidson and J. S. Correia. *The Science and Psychology of Music Performance*, chapter Body Movement, pages 237–250. Oxford University Press, 2002.

[36] J. W. Davis. Recognizing movement using motion histograms. Technical report, MIT Media Laboratory, 1998.

[37] P. de Alcantara. *Indirect Procedures : A Musician's Guide to the Alexander Technique*. Oxford University Press, 1997.

[38] F. Delalande. *Glenn Gould,Pluriel*, chapter La gestique de Gould:elements pour un semiolgie du geste musical, pages 83–111. Louise Corteau Editrice Inc, 1988.

[39] S. Dixon. An interactive beat tracking and visualisation system. In *Proceedings of the International Computer Music Conference*, 2001.

[40] S. Dixon. On the analysis of musical expression in audio signals. *Storage and Retrieval for Media Databases, SPIE-IS&T Electronic Imaging*, 5021:122–132, 2003.

[41] S. Dixon and G. Widmer. Match: A musical alignment toolchest. In *Proceedings of the 6th International Conference on Music Information Retrieval*, 2005.

[42] C. Drake and C. Palmer. Accent structures in music performance. *Music Perception*, 10:3:343–378, 1993.

[43] A. Friberg and J. Sundberg. Does music performance allude to locomotion? a model of final ritardandi derived from the measurements of stopping runners. *Journal of Acoustical Society of America*, 105(3):1469–1484, 1999.

[44] A. Gabrielsson. Music performance research at the millenium. *Psychology of Music*, 31:3:221–272, 2003.

[45] B. Gingras, S. McAdams, and P. Schubert. The performer as analyst: A case study of j.s. bach's "dorian" fugue (bwv 538). *Journal of New Music Research*, 2009.

[46] J. Ginsborg and E. King. Gestures and glances: The effects of familiarity and expertise on singers' and pianists' bodily movements in ensemble rehearsals. In *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyvaskyla, Finland*, 2009.

[47] R. Godoy, E. Haga, and A. Jensenius. Exploring music-related gestures by sound tracing . a preliminary study. *2nd International Symposium on Gesture Interfaces for Multimedia Systems (GIMS2006)*, 2006.

[48] R. Godoy and A. Jensenius. Body movement in music information retrieval. In *Proceedings of the International Society for Music Information Retrieval*, 2009.

[49] W. Goebl, R. Bresin, and A. Galembo. The piano action as the performer's interface: Timing properties,. *Proceedings of the Stockholm Music acoustics Conference*, 2003.

[50] W. Goebl, S. Dixon, G. DePoli, A. Friberg, R. Bresin, and G. Widmer. *Sound to Sense, Sense to Sound: A State-of-the-Art in Sound and Music Computing*, chapter "Sense" in Expressive Music Performance: Data Acquisition, Computational Studies and Models, pages 63–103. Berlin: Logos Verlag., 2005.

[51] W. Goebl and C. Palmer. Tactile feedback and timing accuracy in piano performance. *Experimental Brain Research*, 2007.

[52] W. Goebl and C. Palmer. Finger motion in piano performance: Touch and tempo. In *Proceedings of the International Symposium on Performance Science*, 2009.

[53] W. Goebl and C. Palmer. Synchronization of timing and motion among performing musicians. *Music Perception*, 26(5):427–438, 2009.

[54] M. Good. *The Virtual Score:Representation, Retreival, Restoration: Vol.9.*, chapter MusicXML for Notation and Analysis, pages 113–124. MIT Press, 2001.

[55] D. O. Gorodnichy and A. Yogeswaran. Detection and tracking of pianist hands and fingers. In *Proceedings of the Canadian Conference of Computer and Robot Vision*, 2006.

[56] M. Grachten and G. Widmer. Towards phrase structure reconstruction from expressive performance data. In *The inaugural International Conference on Music Communication Science, Sydney, Australia*, 2007.

[57] M. Grachten and G. Widmer. Who is who in the end? recognizing pianists by their final ritardandi. In *Proceedings of the International Society for Music Information Retrieval*, 2009.

[58] H. Guan, C.-S. Chua, and Y.-K. Ho. 3d hand pose retrieval from a single 2d image. In *International Conference on Image Processing 2001, Proceedings*, volume 1, pages 157–160, 2001.

[59] H. Honing. The final ritard: On music, motion, and kinematic models. *Computer Music Journal*, 27:3:66–72, 2003.

[60] D. Huron. Music information processing using the humdrum toolkit: Concepts, examples, and lessons. *Computer Music Journal*, 26(2):11–26, 2002.

[61] A. Jensenius. Using motiongrams in the study of musical gestures. In *Proceedings of the 2006 International Computer Music Conference*, 2006.

[62] A. Jensenius. *Action-Sound: Developing Methods and Tools to Study Music-Related Body Movement*. PhD thesis, Department of Musicology, University of Oslo, 2007.

[63] A. Jensenius, N. Castagne, A. Camurri, E. Maestre, J. Malloch, and D. McGilvray. A summary of formats for streaming and storing music-related movement and gesture data. In *Proceedings of the 4th International Conference on Enactive Interfaces*, 2007.

[64] T. Jensenius, A. R.and Kvifte and R. I. Godoy. Towards a gesture description interchange format. *Proceedings of New Interfaces for Musical Expression, NIME 06, IRCAM-Centre Pompidou, Paris, France*, pages 176–179, 2006.

[65] V. Kofi Agawu. Concepts of closure and chopin's opus 28. *Music Theory Spectrum*, 9:1–17, 1987.

[66] J. Langner and W. Goebl. Representing expressive performance in tempo-loudness space. *ESCOM Creativity, Liege, April 2002*, 2002.

[67] H. Lehtonen, A. Askenfelt, and V. Valimaki. Analysis of the part-pedaling effect of the piano. *Journal of Acoustical Society of America*, 126 (2):EL49–EL54, 2009.

[68] H. Lehtonen, H.M.and Penttinen, J. Rauhala, and V. Valimaki. Analysis of the sustain-pedal effect in the grand piano. In *Proceedings of 19th International Congress on Acoustics, Madrid, Spain*, 2007.

[69] M. Leman. *Embodied Music Cognition and Mediation Technology*. MIT Press, 2008.

[70] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, 1983.

[71] J. Lester. *The Practice of Performance: Studies in Musical Interpretation*, chapter Performance and Analysis: Interaction and Interpretation. Cambridge, CUP, 1995.

[72] C.-C. Lien and C.-L. Huang. Model-based articulated hand motion tracking for gesture recognition. *Image and Vision Computing*, 16:121–134, 1998.

[73] E. Lin, A. Cassidy, D. Hook, A. Baliga, and T. Chen. Hand tracking using spatial gesture modeling as visual feedback. *Multimodal Interfaces,2002. proceedings. Fourth IEEE International Conference*, pages 197–202, 2002.

[74] I. Lindstedt. Some remarks on computer-assisted analysis of the finale of chopin's piano sonata in b flat minor op.35. *Analytical Perspectives on the Music of Chopin*, pages 203–223, 2004.

[75] A. Luciani, M. Evrard, N. Castagne, C. Cadoz, J.-L. Florens, and D. Courousse. A basic gesture and motion format for virtual reality multisensory applications. In *Proceedings of the 1st international conference on Computer Graphics Theory and Applications*, 2006.

[76] J. MacRitchie, N. Bailey, and G. Hair. Multi-modal acquisition of performance parameters for analysis of chopin's b flat minor piano sonata finale op.35. In *Abstract Proceedings of DMRN+1 Workshop, Queen Mary University, London*, 2006.

[77] J. MacRitchie, B. Buck, and N. Bailey. Gestural communication: Linking the multi-modal analysis of performance to perception of musical structure. In *Proceedings of the 2nd International Symposium of Performance Science, Auckland, New Zealand*, 2009.

[78] J. MacRitchie, B. Buck, and N. Bailey. Visualising musical structure through performance gesture. In *Proceedings of the 10th International Society for Music Information Retrieval Conference, 26th-30th October, Kobe, Japan*, 2009.

[79] J. MacRitchie, G. Hair, N. Bailey, and S. Pullinger. Extracting musical structure from multi-modal performance analysis. In *Proceedings of the Fourth Conference on Interdisciplinary Musicology, 2nd-6th July, Thessaloniki, Greece*, 2008.

[80] J. MacRitchie, S. Pullinger, N. Bailey, and G. Hair. Communicating phrasing structure with multi-modal expressive techniques in piano performance. In *Proceedings of The Second International Conference on Music Communication Science, 3-4 December Sydney, Australia*, 2009.

[81] R. A. d. S. Marranita. Visual tracking of articulated objects: An application to the human hand. Master's thesis, Universidade tecnica de lisboa : Instituto superior tecnico, 2005.

[82] D. McGilvray. *On the Analysis of Performance By Computer*. PhD thesis, University of Glasgow, 2008.

[83] D. McNeill. *Language and Gesture*. Cambridge University Press, 2000.

[84] M. Nusseck and M. Wanderley. Music and motion - how music-related ancillary body movements contribute to the experience of music. *Music Perception*, 26:4:335–353, 2009.

[85] O. Ortmann. *Physiological Mechanics of Piano Technique*. Da Capo Press, 1981.

[86] C. Palmer. Music performance. *Annual Reviews in Psychology*, 48:115–38, 1997.

[87] C. Palmer, C. Carter, E. Koopmans, and J. Loehr. Movement, planning and music: Motion coordinates of skilled performance. In *The inaugural conference on music communication science*, 2007.

[88] C. Palmer and S. Dalla Bella. Movement amplitude and tempo change in piano performance. *Journal of Acoustical Society of America*, 115:5, 2004.

[89] C. Palmer, E. Koopmans, J. Loehr, and C. Carter. Movement-related feedback and temporal accuracy in clarinet performance. *Music Perception*, 26(5):439–450, 2009.

[90] R. Parncutt. *Perspektiven und Methoden einer Systemischen Musikwissenschaft*, chapter Accents and Expression in Piano Performance, pages 163–185. Germany: Peter Lang, 2003.

[91] R. Parncutt. *Perspektiven und Methoden einer Systemischen Musikwissenschaft*. Festchrift Frike, 2003. Accents and expression in piano performance chapter in book.

[92] K. Pearson. On the lines and planes of closest fit to systems of points in space. *Philisophical Magazine*, 2:6:559–572, 1901.

[93] R. Perry. The music encoding initiative (mei). *Proceedings of the First International Conference on Musical Applications Using XML*, pages p55–59, 2002.

[94] S. Pullinger. *A System for the Analysis of Musical Data*. PhD thesis, University of Glasgow, 2010.

[95] S. Pullinger, N. J. Bailey, D. McGilvray, and J. MacRitchie. A workflow for the computer representation of musical performance. University of Glasgow, Technical Report, 2007.

[96] S. Pullinger, D. McGilvray, and N. Bailey. Music and gesture file: Performance visualisation, analysis, storage and exchange. In *Proceedings of the International Computer Music Conference*, 2008.

[97] J. Ramsay and B. Silverman. *Functional Data Analysis*. Springer, New York, 2005.

[98] B. Repp. Expressive timing in schumann's "traumerei": An analysis of performances by graduate student pianists. *Journal of Acoustical Society of America*, 98(5):2413–2427, 1995.

[99] B. Repp. Expressive timing in a debussy prelude : A comparison of student and expert pianists. *Musicae Scientiae*, 1(2):257–268, 1997.

[100] B. Repp. A microcosm of musical expression: Ii. quantitative analysis of pianists' dynamics in the initial measures of chopin's etude in e major. *Journal of Acoustical Society of America*, 105:3:1972–1988, 1999.

[101] B. H. Repp. The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception*, 14, No.4:419–444, 1997.

[102] B. H. Repp. A microcosm of musical expression. i. quantitative analysis of pianists' timing in the initial measures of chopin's etude in e major. *Journal of Acoustical Society of America*, 104(2):1085–1100, 1998.

[103] H. Rijpkema and M. Girard. Computer animation of knowledge-based human grasping. In *Proceedings of the 18th Annual Conference on Computer Graphics and Interactive Techniques*, 1991.

[104] J. Rink. *Musical Performance: A guide to understanding*, chapter Analysis and (or?) Performance, pages 35–58. Cambridge University Press, 2002.

[105] M. Rodger. The acquisition of musical body movement in relation to acquired skill in expert clarinet performance. In *Abstract Proceedings of The*

Musical Body: Gesture, Representation and Ergonomics in Musical Performance, London, UK, 2009.

[106] C. Rosen. *The Romantic Generation*. Fontana Press, 1999.

[107] D. Rumelhart and D. Norman. Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science*, 6:1–36, 1982.

[108] C. Seashore. *Psychology of Music*. McGraw-Hill, 1938.

[109] O. Senn, L. Kilchenmann, and M. Camp. Expressive timing: Martha argerich plays chopin's prelude op.28/4 in e minor. In *Proceedings of Intl Symposium on Performance Science*, pages 107–112, 2009.

[110] H. Shoda and M. Adachi. Time-series analysis of relationships between the pianist's body movement and the temporal expression for romantic pieces. In *Proceedings of the Second International Conference on Music Communication Science, Sydney, Australia*, 2009.

[111] J. Sloboda and A. Lehmann. Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a chopin piano prelude. *Music Perception*, 19(1):87–120, 2001.

[112] N. Spiro, N. Gold, and J. Rink. Performance motives: Analysis and comparison of performance timing repetitions using pattern matching and formal concept analysis. In *Proceedings of International Symposium on Performance Science*, 2007.

[113] A. Stulov and D. Kartofelev. Piano hammer-string interaction: The influence of the elastic parameters of bass hammers on the contact time duration. In *Proceedings of the Acoustics High Tatras 2009 '34th International Acoustical Conference - EAA Symposium'*, 2009.

[114] H. Suzuki. Spectrum analysis and tone quality evaluation of piano tone with hard and soft touches. *Acoustical Science and Technology*, 28(1):1–6, 2007.

[115] M. Talbot. *The Finale in Western Instrumental Music*. Oxford University Press, 2001.

[116] R. Taniskin. *The Oxford History of Western Music, Vol3:The 19th Century*. Oxford University Press, 2005.

[117] M. Thompson and G. Luck. Effect of pianists' expressive intention on amount and type of body movement. In *Proceedings of the 10th International conference on music perception and cognition*, 2008.

[118] M. Thompson and G. Luck. Exploring relationships between expressive and structural elements of music and pianist's gestures. In *Proceedings of Fourth Conference on Interdisciplinary Musicology (CIM08) Thessaloniki, Greece, 3-6 July*, 2008.

[119] W. Thompson and B. Robitaille. Can composers express emotions through music? *Empirical Studies of the Arts*, 10:1:79–89, 1992.

[120] G. Todd. Listener habits and choices–and their implications for public performance venues. *Journal of Sound and Vibration*, 239:589–606, 2001.

[121] N. Todd. A model of expressive timing in tonal music. *Music Perception*, 3:1:33–58, 1985.

[122] A. Tuchowski. *Integracja strukturalna w ÅŻwietle przemian stylu Chopina [Structural coherence in view of the transformational changes of Chopins style]*. KrakÃşw: Musica Iagellonica, 1996.

[123] B. Vines, C. Krumhansl, M. Wanderley, and D. Levitin. Cross-modal interactions in the perception of musical performance. *Cognition*, 101:80–113, 2006.

[124] B. Vines, M. Wanderley, C. Krumhansl, R. Nuzzo, and D. Levitin. *GW 2003*, chapter Performance Gestures of Musicians : What Structural and Emotional Information do they convey?, pages 468–478. Springer-Verlag, 2004.

[125] M. Wanderley, B. Vines, N. MIddleton, C. McKay, and W. Hatch. The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, 34:97–113, 2005.

[126] M. M. Wanderley. *Gesture and Sign Language in Human-Computer Interaction*, chapter Quantitative Analysis of Non-Obvious Performer Gestures, pages 241–253. Springer-Verlag, 2002.

[127] Welch. Motion tracking: No silver bullet, but a respectable arsenal. *IEEE Computer Graphics and Applications*, 22 :6:24–38, 2002.

[128] H. Wold. *Multivariate Analysis*, chapter Estimation of principal components and related models by iterative least squares, pages 391–420. Academic Press, NY, 1966.

Appendices

Appendix A
Expanded loadings tables for three performers from Chapter 8

Marker	PC1	PC2	PC3	PC4	PC5	PC6
C7:X	0.09	-0.01	0.28	-0.01	0.05	-0.04
C7:Y	-0.11	0.15	0.01	0	0.1	0.09
C7:Z	-0.13	-0.07	-0.06	-0.03	-0.15	0.15
T10:X	0.12	0.02	0.2	-0.02	0.08	-0.05
T10:Y	-0.13	0.11	0	-0.02	0.1	0.04
T10:Z	-0.13	0.03	-0.15	0.04	-0.08	0.09
CLAV:X	0.1	0.03	0.26	0	0.05	-0.04
CLAV:Y	-0.1	0.17	0.03	0	0.06	0.06
CLAV:Z	-0.08	-0.11	0.17	0	-0.22	0.13
STRN:X	0.11	0.06	0.2	0	0.06	-0.03
STRN:Y	-0.1	0.17	0.05	-0.02	0.03	0.02
STRN:Z	0.07	-0.11	0.25	0.02	-0.06	-0.05
LSHO:X	0.08	-0.04	0.27	-0.05	0.07	0
LSHO:Y	-0.1	0.16	0.03	0.02	0.07	0.06
LSHO:Z	-0.13	0.02	0.11	0.05	-0.11	0.19
LUPA:X	0.12	-0.05	0.15	-0.11	0.09	0.02
LUPA:Y	-0.11	0.15	0.06	0.08	0.03	0
LUPA:Z	-0.13	0.02	0.11	-0.03	-0.08	0.27
LUPB:X	0.14	-0.01	0.05	-0.09	0.08	0.02
LUPB:Y	-0.13	0.12	0.07	0.07	0.05	-0.03
LUPB:Z	-0.09	0.09	0.06	-0.02	-0.15	0.37
LUPC:X	0.14	0.04	0.1	0.01	-0.01	-0.01
LUPC:Y	-0.11	0.12	0.09	0.16	-0.02	-0.08
LUPC:Z	-0.13	-0.03	0.15	0	-0.04	0.14
LELB:X	0.14	0.08	-0.02	0.04	-0.05	0
LELB:Y	-0.12	0.07	0.1	0.17	0	-0.12
LELB:Z	-0.1	0.05	0.13	0	-0.17	0.29
LMEP:X	0.15	0.03	-0.01	-0.03	0.01	0.03
LMEP:Y	-0.12	0.09	0.1	0.18	-0.03	-0.11
LMEP:Z	-0.11	-0.03	0.13	-0.08	-0.06	0.27
LWRA:X	0.09	0.15	0	0.15	-0.14	-0.03
LWRA:Y	-0.08	0.13	0.07	0.24	-0.09	-0.12
LWRA:Z	-0.13	-0.04	0.14	0.04	0.07	-0.02
LWRB:X	0.11	0.13	-0.02	0.11	-0.13	0
LWRB:Y	-0.07	0.14	0.07	0.24	-0.1	-0.12
LWRB:Z	-0.14	-0.06	0.12	0	0.07	0
LFRA:X	0.14	0.08	-0.01	0.04	-0.06	0.02
LFRA:Y	-0.1	0.12	0.08	0.22	-0.08	-0.12
LFRA:Z	-0.13	-0.05	0.12	-0.03	0.02	0.12
LFIN:X	0.1	0.14	-0.02	0.13	-0.15	-0.01
LFIN:Y	-0.07	0.15	0.06	0.24	-0.12	-0.11
LFIN:Z	-0.13	-0.08	0.07	-0.06	0.12	-0.02
RSHO:X	0.1	0.05	0.24	-0.01	0.03	-0.02
RSHO:Y	-0.1	0.17	0.01	0	0.11	0.08
RSHO:Z	-0.09	-0.15	0	-0.01	-0.24	0.01
RUPA:X	0.11	0.12	0.14	-0.1	-0.02	0
RUPA:Y	-0.1	0.16	0	-0.1	0.05	0
RUPA:Z	-0.12	-0.08	0.01	-0.12	-0.25	-0.11
RUPB:X	0.11	0.13	0.05	-0.11	-0.06	0.04
RUPB:Y	-0.11	0.14	0.01	-0.15	0	-0.05
RUPB:Z	-0.09	-0.03	-0.04	-0.17	-0.37	-0.08
RUPC:X	0.11	0.13	0.08	-0.1	-0.06	0.03
RUPC:Y	-0.09	0.16	-0.01	-0.19	0	-0.06
RUPC:Z	-0.11	-0.13	0.09	0.05	-0.05	-0.11
RELB:X	0.11	0.14	-0.01	-0.13	-0.09	0.06
RELB:Y	-0.09	0.14	0	-0.23	-0.04	-0.09
RELB:Z	-0.12	-0.07	0.06	-0.07	-0.19	-0.16
RMEP:X	0.1	0.15	0.01	-0.15	-0.08	0.03
RMEP:Y	-0.09	0.15	-0.01	-0.22	-0.02	-0.09
RMEP:Z	-0.13	-0.07	0.07	-0.08	-0.1	-0.16
RWRA:X	0.11	0.14	-0.01	-0.1	-0.1	0.08
RWRA:Y	-0.11	0.1	0.01	-0.22	0.02	-0.2
RWRA:Z	-0.12	-0.1	0.09	0.01	0.06	-0.07
RWRB:X	0.11	0.14	-0.01	-0.12	-0.1	0.04
RWRB:Y	-0.11	0.09	0.01	-0.21	0.02	-0.19
RWRB:Z	-0.13	-0.09	0.09	-0.02	0.05	-0.07
RFRA:X	0.1	0.15	0.01	-0.14	-0.08	0.05
RFRA:Y	-0.1	0.13	0	-0.22	0	-0.14
RFRA:Z	-0.13	-0.08	0.09	-0.04	-0.01	-0.11
RFIN:X	0.11	0.14	-0.01	-0.11	-0.09	0.06
RFIN:Y	-0.12	0.08	0.02	-0.2	0.05	-0.21
RFIN:Z	-0.13	-0.09	0.08	-0.06	0.1	-0.07
RFHD:X	0.12	0.06	0.14	0.06	-0.1	-0.08
RFHD:Y	-0.09	0.17	0.05	0.06	0.13	0.05
RFHD:Z	0.07	-0.07	0.12	-0.04	-0.29	-0.1
LFHD:X	0.12	-0.01	0.2	-0.01	0	-0.03
LFHD:Y	0	0.21	0.01	0.1	-0.03	0.02
LFHD:Z	0.07	-0.1	0.14	-0.05	-0.09	0.05
LBHD:X	0.04	-0.09	0.24	-0.1	0.17	0.02
LBHD:Y	-0.06	0.19	0	0.04	0.1	0.11
LBHD:Z	-0.08	-0.12	0.1	-0.07	0.07	0.21
RBHD:X	0.11	0.03	0.22	0.04	-0.03	-0.08
RBHD:Y	-0.12	0.05	0.04	-0.05	0.23	0.11
RBHD:Z	-0.11	-0.07	0.04	-0.05	-0.18	-0.01

Figure 11.1: Loadings for the First Six Principal Components, Performer 1, Prelude 7 with top ten loadings in the first component highlighted in red and the second component in blue

Marker	PC1	PC2	PC3	PC4	PC5	PC6
C7:X	0.1	-0.14	0.13	-0.06	0.02	-0.02
C7:Y	0.15	0.1	-0.03	-0.01	-0.05	0.05
C7:Z	0.12	0.07	0.09	-0.07	-0.08	-0.05
T10:X	0.09	-0.16	0.08	-0.04	0.08	0.01
T10:Y	0.16	0.09	0.01	0.01	-0.04	0.06
T10:Z	0.01	0.18	-0.09	0.03	-0.07	0
CLAV:X	0.11	-0.14	0.1	-0.04	0.05	-0.02
CLAV:Y	0.16	0.09	-0.02	0	-0.04	0.03
CLAV:Z	0.06	-0.09	0.24	-0.1	-0.04	-0.08
STRN:X	0.11	-0.14	0.05	-0.01	0.1	0
STRN:Y	0.16	0.08	0	0.03	-0.03	0.03
STRN:Z	0.05	-0.12	0.23	-0.08	0	-0.07
LSHO:X	0.08	-0.15	0.13	-0.09	0.06	-0.01
LSHO:Y	0.16	0.1	-0.03	0	-0.04	0.03
LSHO:Z	0.13	0.1	0.06	-0.1	-0.1	-0.03
LUPA:X	0.07	-0.16	0.11	-0.07	0.03	0.08
LUPA:Y	0.16	0.09	-0.02	0.05	-0.03	-0.01
LUPA:Z	0.13	0.08	0.05	-0.19	-0.01	-0.1
LUPB:X	0.07	-0.16	0.1	-0.05	-0.03	0.19
LUPB:Y	0.16	0.09	0	0.05	-0.03	0.02
LUPB:Z	0.13	0.1	0	-0.15	-0.05	-0.05
LUPC:X	0.09	-0.16	0.07	0.02	-0.02	0.1
LUPC:Y	0.16	0.07	-0.01	0.11	-0.03	-0.05
LUPC:Z	0.12	0.07	0.09	-0.19	0.03	-0.15
LELB:X	0.06	-0.12	0.02	0.14	-0.16	0.24
LELB:Y	0.14	0.06	0.04	0.16	-0.05	-6.050434626
LELB:Z	0.13	0.07	0.03	-0.15	0.01	-0.15
LMEP:X	0.05	-0.11	0.04	0.07	-0.16	0.33
LMEP:Y	0.14	0.05	0.02	0.18	-0.03	-0.07
LMEP:Z	0.11	0.08	0.04	-0.23	0.06	-0.14
LWRA:X	0.08	-0.11	-0.04	0.19	0.03	-0.25
LWRA:Y	0.12	0.05	0	0.21	-0.01	-0.22
LWRA:Z	-0.03	0.15	0.07	-0.19	0.15	-0.03
LWRB:X	0.08	-0.12	-0.04	0.18	0.01	-0.19
LWRB:Y	0.12	0.04	0	0.21	0.02	-0.26
LWRB:Z	-0.02	0.15	0.07	-0.2	0.14	-0.05
LFRA:X	0.08	-0.14	0	0.14	-0.09	0.11
LFRA:Y	0.13	0.04	0.01	0.2	0.01	-0.2
LFRA:Z	0.03	0.13	0.06	-0.25	0.13	-0.09
LFIN:X	0.07	-0.11	-0.05	0.17	0.05	-0.26
LFIN:Y	0.11	0.03	-0.01	0.21	0.03	-0.29
LFIN:Z	-0.04	0.16	0.08	-0.16	0.1	0.02
RSHO:X	0.12	-0.14	0.09	-0.03	0.04	-0.02
RSHO:Y	0.16	0.09	-0.03	-0.01	-0.06	0.05
RSHO:Z	-0.13	0.03	0.15	0.08	0.06	-0.07
RUPA:X	0.13	-0.13	0.01	-0.03	0.13	0.01
RUPA:Y	0.16	0.1	0.01	0.01	-0.01	0.07
RUPA:Z	-0.12	0.03	0.14	0.13	0.17	7.5376660128
RUPB:X	0.11	-0.13	-0.09	-0.06	0.17	0.01
RUPB:Y	0.15	0.1	0.05	0.05	0.05	0.1
RUPB:Z	-0.12	0.03	0.13	0.14	0.17	-0.02
RUPC:X	0.13	-0.13	-0.02	-0.04	0.15	0.01
RUPC:Y	0.15	0.1	0.01	0.02	0.03	0.08
RUPC:Z	-0.09	0.03	0.24	0.15	0.07	-0.02
RELB:X	0.09	-0.12	-0.14	-0.06	0.19	0.01
RELB:Y	0.13	0.09	0.07	0.09	0.13	0.14
RELB:Z	-0.09	0.03	0.23	0.16	0.12	-0.01
RMEP:X	0.1	-0.1	-0.13	-0.06	0.23	0.04
RMEP:Y	0.13	0.1	0.06	0.07	0.11	0.13
RMEP:Z	-0.09	0.04	0.22	0.16	0.15	0.04
RWRA:X	0.1	-0.12	-0.13	-0.04	0.19	-0.03
RWRA:Y	0.1	0.12	0.07	0.1	0.11	0.18
RWRA:Z	-0.04	0.13	0.16	0	0.12	-0.03
RWRB:X	0.1	-0.12	-0.14	-0.03	0.19	0.01
RWRB:Y	0.1	0.12	0.07	0.1	0.11	0.18
RWRB:Z	-0.04	0.13	0.17	0.01	0.12	-0.01
RFRA:X	0.1	-0.09	-0.12	-0.06	0.24	0.03
RFRA:Y	0.13	0.11	0.06	0.07	0.11	0.14
RFRA:Z	-0.06	0.11	0.21	0.07	0.15	0.01
RFIN:X	0.09	-0.12	-0.14	-0.04	0.18	-0.01
RFIN:Y	0.08	0.13	0.07	0.09	0.07	0.2
RFIN:Z	-0.03	0.13	0.14	-0.04	0.13	-0.02
RFHD:X	0.08	-0.14	0.13	-0.02	-0.09	0
RFHD:Y	0.14	0.1	-0.04	-0.02	-0.11	0.03
RFHD:Z	0.06	-0.11	0.2	0.04	-0.05	0.05
LFHD:X	0.08	-0.14	0.17	-0.05	-0.09	-0.04
LFHD:Y	0.15	0.08	-0.06	0	-0.11	0.05
LFHD:Z	0.01	-0.06	0.22	-0.03	-0.25	-0.07
LBHD:X	0.07	-0.11	0.19	-0.09	-0.06	-0.07
LBHD:Y	0.15	0.1	-0.04	-0.02	-0.11	0.02
LBHD:Z	0.01	-0.02	0.19	-0.12	-0.28	-0.13
RBHD:X	0.09	-0.14	0.15	-0.03	-0.08	-0.01
RBHD:Y	0.13	0.12	-0.03	-0.04	-0.09	0
RBHD:Z	0.09	-0.09	0.12	0.04	0.13	0.12

Figure 11.2: Loadings for the First Six Principal Components, Performer 2, Prelude 7 with top ten loadings in the first component highlighted in red and the second component in blue

Marker	PC1	PC2	PC3	PC4	PC5	PC6
C7:X	0.15	0	0.05	-0.09	0.06	-0.08
C7:Y	0.16	0.01	-0.08	0.02	0.01	0.01
C7:Z	0.1	-0.12	-0.08	-0.01	-0.07	-0.23
T10:X	0.15	0.07	0.03	-0.07	0.02	0.03
T10:Y	0.16	-0.01	-0.09	-0.02	0.01	0.01
T10:Z	0.11	-0.05	-0.15	0.08	-0.14	-0.11
CLAV:X	0.15	0.06	0.09	-0.06	0.04	-0.07
CLAV:Y	0.16	0.03	-0.05	0.03	0	0.01
CLAV:Z	0.06	-0.16	0.01	-0.11	-0.01	-0.28
STRN:X	0.13	0.12	0.11	-0.02	-0.02	0.03
STRN:Y	0.16	0.03	-0.02	0.02	-0.02	0.02
STRN:Z	0.07	-0.16	0.02	-0.1	0.02	-0.25
LSHO:X	0.15	-0.03	0.01	-0.1	0.1	-0.04
LSHO:Y	0.16	0.02	-0.05	0.04	-0.01	0
LSHO:Z	0.13	-0.1	-0.04	0.02	0.01	-0.2
LUPA:X	0.1	-0.04	-0.05	-0.25	0.21	0.01
LUPA:Y	0.16	0.05	-0.04	0.05	-0.07	-0.01
LUPA:Z	0.13	-0.12	0.02	0.09	0.06	-0.08
LUPB:X	0.05	0.03	-0.11	-0.34	0.18	0.02
LUPB:Y	0.16	0.06	-0.05	0.03	-0.08	-0.02
LUPB:Z	0.13	-0.11	0.02	0.1	0.05	-0.08
LUPC:X	0.09	0.09	-0.08	-0.27	0.08	-0.02
LUPC:Y	0.15	0.08	-0.03	0.07	-0.14	-0.03
LUPC:Z	0.11	-0.12	0.07	0.14	0.07	-0.08
LELB:X	-0.01	0.12	-0.14	-0.3	0.03	-0.03
LELB:Y	0.14	0.09	-0.05	0.03	-0.15	-0.02
LELB:Z	0.11	-0.12	0.06	0.13	0.04	-0.06
LMEP:X	0	0.09	-0.1	-0.33	0.13	0.01
LMEP:Y	0.14	0.09	-0.03	0.07	-0.17	-0.03
LMEP:Z	0.1	-0.14	0.07	0.13	0.09	-0.03
LWRA:X	0.05	0.16	-0.08	-0.09	-0.22	0.02
LWRA:Y	0.13	0.08	-0.04	0.08	-0.21	0
LWRA:Z	0.04	-0.18	0.04	0.06	0.11	0.09
LWRB:X	0.03	0.17	-0.09	-0.15	-0.15	-0.01
LWRB:Y	0.13	0.07	-0.04	0.09	-0.22	0
LWRB:Z	0.06	-0.18	0.05	0.09	0.08	0.1
LFRA:X	0	0.14	-0.12	-0.28	0.02	-0.01
LFRA:Y	0.13	0.09	-0.04	0.08	-0.2	-0.02
LFRA:Z	0.09	-0.16	0.07	0.12	0.09	0.01
LFIN:X	0.05	0.16	-0.08	-0.09	-0.21	0.01
LFIN:Y	0.13	0.07	-0.03	0.09	-0.23	0.01
LFIN:Z	0.03	-0.17	0.02	0.03	0.1	0.12
RSHO:X	0.12	0.08	0.19	-0.06	0.05	-0.09
RSHO:Y	0.16	0.01	-0.09	0.02	0.02	-0.01
RSHO:Z	-0.05	-0.13	0.06	-0.13	-0.14	-0.27
RUPA:X	0.07	0.13	0.23	-0.01	0.11	-0.06
RUPA:Y	0.17	0	-0.05	0.01	0.03	0.05
RUPA:Z	-0.02	-0.14	0.15	-0.15	-0.17	0.02
RUPB:X	0.02	0.17	0.15	0.05	0.16	-0.11
RUPB:Y	0.16	-0.01	-0.01	0	0.03	0.12
RUPB:Z	-0.01	-0.13	0.16	-0.14	-0.2	0.01
RUPC:X	0.05	0.15	0.22	6.3081061231	0.11	-0.06
RUPC:Y	0.16	0	-0.01	0.01	0.05	0.09
RUPC:Z	0.02	-0.16	0.13	-0.16	-0.2	0.04
RELB:X	0	0.17	0.13	0.07	0.16	-0.11
RELB:Y	0.15	-0.02	0.05	-0.01	0.06	0.19
RELB:Z	0.01	-0.08	-0.02	-0.05	-0.09	0.03
RMEP:X	0.01	0.17	0.18	0.05	0.15	-0.07
RMEP:Y	0.15	-0.01	0.04	-0.01	0.06	0.17
RMEP:Z	0.03	-0.15	0.15	-0.14	-0.16	0.13
RWRA:X	0.01	0.14	0.25	0	-0.02	-0.03
RWRA:Y	0.14	-0.01	0.05	-0.02	0.08	0.19
RWRA:Z	0.05	-0.12	0.17	-0.12	-0.14	0.15
RWRB:X	0	0.16	0.23	0.02	0	-0.06
RWRB:Y	0.14	-0.02	0.01	-0.01	0.11	0.19
RWRB:Z	0.04	-0.13	0.18	-0.13	-0.12	0.17
RFRA:X	0	0.17	0.19	0.04	0.09	-0.07
RFRA:Y	0.15	-0.02	0.04	-0.01	0.08	0.19
RFRA:Z	0.04	-0.14	0.18	-0.14	-0.15	0.16
RFIN:X	0.01	0.14	0.25	0	-0.04	-0.04
RFIN:Y	0.14	-0.02	0.01	-0.02	0.12	0.19
RFIN:Z	0.03	-0.1	0.21	-0.13	-0.11	0.16
RFHD:X	0.06	0.1	0.19	-0.02	-0.1	-0.13
RFHD:Y	0.16	0.02	-0.07	0.03	0.01	0.01
RFHD:Z	-0.03	-0.13	0.09	-0.05	-0.07	-0.19
LFHD:X	0.14	0.02	0.09	-0.09	0.04	-0.19
LFHD:Y	0.15	0.06	-0.04	0.04	-0.04	-0.01
LFHD:Z	0.04	-0.16	-0.03	-0.08	0.01	-0.25
LBHD:X	0.13	-0.08	-0.01	-0.1	0.11	-0.09
LBHD:Y	0.16	0.02	-0.08	0.02	0.02	-0.02
LBHD:Z	0.08	-0.15	-0.12	-0.06	0.07	-0.11
RBHD:X	0.11	0.03	0.18	-0.07	-0.03	-0.12
RBHD:Y	0.16	-0.02	-0.1	0	0.05	0
RBHD:Z	0.01	-0.18	-0.01	-0.03	-0.02	-0.08

Figure 11.3: Loadings for the First Six Principal Components, Performer 3, Prelude 7 with top ten loadings in the first component highlighted in red and the second component in blue

Marker	PC1	PC2	PC3	PC4	PC5	PC6
C7:X	0.07	-0.06	0.25	0.02	-0.09	0.04
C7:Y	0.16	0.08	-0.02	-0.05	-0.03	-0.09
C7:Z	-0.01	0.15	0.05	0.05	0.2	-0.16
T10:X	0.06	-0.12	0.18	0.03	-0.17	0.05
T10:Y	0.15	0.1	-0.03	0.02	-0.04	-0.15
T10:Z	0.01	0.14	-0.16	-0.01	0.15	-0.05
CLAV:X	0.1	-0.07	0.21	0.03	-0.11	0.07
CLAV:Y	0.17	0.07	-0.02	-0.02	-0.01	-0.04
CLAV:Z	-0.02	0.1	0.21	0.09	0.14	-0.09
STRN:X	0.08	-0.11	0.13	0.03	-0.2	0.13
STRN:Y	0.16	0.07	-0.03	0.05	-0.01	-0.04
STRN:Z	-0.01	-0.01	0.26	0.11	0.02	0.01
LSHO:X	0.01	-0.06	0.27	0.01	-0.13	-0.02
LSHO:Y	0.16	0.08	-0.02	-0.04	-0.01	-0.03
LSHO:Z	0.08	0.13	0.11	-0.06	0.11	-0.09
LUPA:X	-0.04	-0.12	0.19	0	-0.14	-0.18
LUPA:Y	0.16	0.08	-0.02	0.01	-0.03	0.02
LUPA:Z	0.06	0.13	0.13	-0.15	0.16	-0.07
LUPB:X	-0.02	-0.16	0.08	0.02	-0.13	-0.28
LUPB:Y	0.16	0.09	-0.02	0.05	-0.07	0.04
LUPB:Z	0.09	0.11	0.1	-0.17	0.21	-0.14
LUPC:X	0.04	-0.16	0.1	0.08	-0.11	-0.19
LUPC:Y	0.16	0.08	-0.01	0.08	-0.06	0.11
LUPC:Z	0.03	0.15	0.15	-0.09	0.1	0.09
LELB:X	0.05	-0.15	-0.03	0.1	-0.07	-0.26
LELB:Y	0.14	0.08	-0.02	0.14	-0.16	0.13
LELB:Z	0.06	0.12	0.14	-0.12	0.23	0.03
LMEP:X	0.02	-0.16	-0.02	0.03	-0.09	-0.32
LMEP:Y	0.15	0.08	-0.01	0.14	-0.1	0.15
LMEP:Z	0	0.12	0.16	-0.19	0.19	0.02
LWRA:X	0.12	-0.1	-0.01	0.19	0.07	-0.08
LWRA:Y	0.15	0.07	0	0.18	-0.04	0.17
LWRA:Z	-0.01	0.16	0.06	-0.08	-0.13	0.06
LWRB:X	0.11	-0.13	-0.01	0.16	0.07	-0.13
LWRB:Y	0.15	0.06	0	0.19	-0.02	0.17
LWRB:Z	-0.03	0.16	0.06	-0.11	-0.1	0.07
LFRA:X	0.06	-0.16	-0.02	0.09	-0.03	-0.26
LFRA:Y	0.15	0.07	-0.01	0.16	-0.06	0.17
LFRA:Z	-0.02	0.15	0.12	-0.16	0.06	0.06
LFIN:X	0.12	-0.11	-0.01	0.18	0.09	-0.08
LFIN:Y	0.15	0.06	0.01	0.17	0.01	0.19
LFIN:Z	-0.05	0.16	0.03	-0.07	-0.15	0.07
RSHO:X	0.13	-0.06	0.17	0.02	-0.06	0.1
RSHO:Y	0.17	0.07	-0.03	-0.04	-0.02	-0.08
RSHO:Z	-0.14	0.08	0.06	0.18	0.11	-0.02
RUPA:X	0.14	-0.09	0.08	0	0.01	0.1
RUPA:Y	0.16	0.08	-0.03	0	-0.01	-0.1
RUPA:Z	-0.14	0.07	0.05	0.22	0.09	-0.04
RUPB:X	0.12	-0.13	0.01	-0.01	0.07	0.09
RUPB:Y	0.15	0.1	-0.02	0.06	-0.01	-0.12
RUPB:Z	-0.14	0.04	0.02	0.24	0.16	-0.05
RUPC:X	0.13	-0.12	0.04	0	0.06	0.08
RUPC:Y	0.16	0.07	-0.05	0.02	0.03	-0.09
RUPC:Z	-0.1	0.12	0.12	0.16	-0.04	-0.02
RELB:X	0.11	-0.14	-0.04	-0.02	0.13	0.06
RELB:Y	0.14	0.09	-0.04	0.12	0.04	-0.12
RELB:Z	-0.12	0.08	0.09	0.22	0.05	-0.04
RMEP:X	0.11	-0.13	-0.03	-0.02	0.12	0.07
RMEP:Y	0.14	0.09	-0.05	0.09	0.04	-0.11
RMEP:Z	-0.11	0.11	0.09	0.19	-0.01	-0.04
RWRA:X	0.11	-0.13	-0.03	-0.02	0.13	0.04
RWRA:Y	0.08	0.13	-0.07	0.12	0	-0.07
RWRA:Z	-0.05	0.16	0.05	0.03	-0.13	-0.11
RWRB:X	0.1	-0.14	-0.03	-0.01	0.14	0.05
RWRB:Y	0.07	0.14	-0.07	0.11	-0.02	-0.06
RWRB:Z	-0.06	0.16	0.04	0.04	-0.14	-0.08
RFRA:X	0.11	-0.13	-0.04	-0.03	0.12	0.06
RFRA:Y	0.13	0.1	-0.06	0.1	0.03	-0.1
RFRA:Z	-0.09	0.14	0.07	0.11	-0.08	-0.06
RFIN:X	0.1	-0.14	-0.04	-0.03	0.13	0.06
RFIN:Y	0.05	0.14	-0.07	0.1	-0.05	-0.05
RFIN:Z	-0.05	0.16	0.02	-0.01	-0.19	0
RFHD:X	0.08	-0.09	0.18	0.04	0.11	0.06
RFHD:Y	0.16	0.08	0	-0.08	-0.03	-0.06
RFHD:Z	-0.07	-0.02	0.15	0.23	0.27	0.07
LFHD:X	0.07	-0.07	0.24	0	0.01	0.02
LFHD:Y	0.17	0.03	0	-0.06	0.06	-0.04
LFHD:Z	0.01	0.01	0.24	-0.02	0.2	-0.1
LBHD:X	0.04	0	0.26	-0.06	-0.16	-0.02
LBHD:Y	0.17	0.06	0	-0.09	0	-0.08
LBHD:Z	0.04	0.11	0.17	-0.12	0.01	-0.18
RBHD:X	0.08	-0.07	0.23	0.02	0.02	0.06
RBHD:Y	0.13	0.11	0	-0.11	-0.11	-0.1
RBHD:Z	-0.1	0.07	0.03	0.26	0.1	0.09

Figure 11.4: Loadings for the First Six Principal Components, Performer 1, Prelude 6 with top ten loadings in the first component highlighted in red and the second component in blue

Marker	PC1	PC2	PC3	PC4	PC5	PC6
C7:X	0.14	-0.09	0.1	-0.08	0.04	-0.03
C7:Y	0.14	0.11	-0.1	-0.04	-0.03	-0.02
C7:Z	0.08	0.17	0.09	-0.08	0.04	0.16
T10:X	0.13	-0.13	0.06	-0.05	0.03	-0.08
T10:Y	0.14	0.09	-0.06	-0.02	-0.01	-0.05
T10:Z	-0.07	0.2	-0.12	0	-0.04	0.12
CLAV:X	0.14	-0.09	0.09	-0.04	0.04	-0.03
CLAV:Y	0.14	0.11	-0.07	-0.01	-0.03	0
CLAV:Z	0.1	0.05	0.2	-0.11	0.08	0.15
STRN:X	0.14	-0.11	0.05	0	0.02	-0.07
STRN:Y	0.14	0.12	-0.02	0.03	-0.03	0.01
STRN:Z	0.1	0.02	0.22	-0.1	0.09	0.14
LSHO:X	0.13	-0.12	0.07	-0.09	0.07	-0.05
LSHO:Y	0.14	0.11	-0.08	-0.02	-0.02	0
LSHO:Z	0.11	0.14	-0.01	-0.14	0.06	0.13
LUPA:X	0.12	-0.14	0.06	-0.11	0.02	-0.08
LUPA:Y	0.14	0.11	-0.03	0.06	0	-0.03
LUPA:Z	0.12	0.07	-0.05	-0.13	0.14	0.1
LUPB:X	0.11	-0.13	0.07	-0.16	-0.05	-0.08
LUPB:Y	0.15	0.09	-0.02	0.06	0	-0.04
LUPB:Z	0.1	0.11	-0.1	-0.17	0.07	0.14
LUPC:X	0.13	-0.11	0.09	-0.09	-0.05	-0.11
LUPC:Y	0.13	0.1	0.02	0.13	0.01	-0.06
LUPC:Z	0.13	0.01	-0.03	-0.07	0.22	0.07
LELB:X	0.06	-0.02	0.13	-0.11	-0.27	-0.13
LELB:Y	0.13	0.04	0.05	0.16	0.04	-0.12
LELB:Z	0.12	0.02	-0.08	-0.06	0.21	0.06
LMEP:X	0.06	-0.06	0.09	-0.17	-0.25	-0.1
LMEP:Y	0.12	0.06	0.07	0.18	0.03	-0.11
LMEP:Z	0.11	0.01	-0.11	-0.11	0.2	0.1
LWRA:X	0.11	-0.08	0.02	0.2	0.07	-0.15
LWRA:Y	0.07	0.14	0.11	0.2	-0.04	-0.08
LWRA:Z	0.05	-0.06	-0.13	0.04	0.32	0
LWRB:X	0.11	-0.1	0.03	0.16	0.05	-0.16
LWRB:Y	0.08	0.11	0.09	0.23	0.01	-0.09
LWRB:Z	0.05	-0.05	-0.13	0.03	0.33	0.01
LFRA:X	0.11	-0.1	0.08	-0.01	-0.14	-0.16
LFRA:Y	0.1	0.1	0.09	0.22	0.02	-0.1
LFRA:Z	0.07	-0.03	-0.13	-0.03	0.31	0.05
LFIN:X	0.11	-0.1	0	0.19	0.09	-0.14
LFIN:Y	0.07	0.12	0.1	0.23	0	-0.08
LFIN:Z	0.04	-0.06	-0.13	0.04	0.32	0
RSHO:X	0.14	-0.07	0.1	-0.04	0.03	-0.01
RSHO:Y	0.14	0.1	-0.09	-0.03	-0.02	-0.03
RSHO:Z	-0.09	0.07	0.21	0.06	0.08	0.19
RUPA:X	0.14	-0.07	0.08	0.03	-0.03	0.1
RUPA:Y	0.14	0.1	-0.06	-0.02	-0.02	-0.04
RUPA:Z	-0.11	0.05	0.18	0.1	0.06	0.14
RUPB:X	0.13	-0.08	0.01	0.09	-0.09	0.21
RUPB:Y	0.14	0.11	-0.03	0	-0.02	-0.04
RUPB:Z	-0.11	0.05	0.17	0.1	0.05	0.17
RUPC:X	0.14	-0.07	0.06	0.05	-0.05	0.13
RUPC:Y	0.14	0.13	-0.06	0	-0.04	0.01
RUPC:Z	-0.07	0.06	0.26	-8.354556078	0.15	-0.01
RELB:X	0.1	-0.1	-0.03	0.13	-0.12	0.27
RELB:Y	0.11	0.14	0.01	0.04	-0.03	0
RELB:Z	-0.08	0.05	0.24	0.02	0.13	0.02
RMEP:X	0.1	-0.06	-0.02	0.15	-0.13	0.31
RMEP:Y	0.11	0.16	-0.01	0.03	-0.04	0.01
RMEP:Z	-0.1	0.06	0.22	0.02	0.11	0.02
RWRA:X	0.11	-0.1	-0.04	0.13	-0.11	0.21
RWRA:Y	0.04	0.2	0.07	0.09	0.01	-0.09
RWRA:Z	-0.08	0.19	0.05	-0.08	-0.01	0.05
RWRB:X	0.11	-0.11	-0.04	0.15	-0.1	0.22
RWRB:Y	0.04	0.21	0.07	0.08	0	-0.09
RWRB:Z	-0.08	0.19	0.06	-0.08	-0.01	0.04
RFRA:X	0.1	-0.05	-0.03	0.14	-0.14	0.3
RFRA:Y	0.09	0.2	0.01	0.04	-0.04	-0.01
RFRA:Z	-0.1	0.17	0.12	-0.06	0.03	0.04
RFIN:X	0.11	-0.12	-0.04	0.14	-0.1	0.22
RFIN:Y	0	0.2	0.08	0.06	0.03	-0.1
RFIN:Z	-0.08	0.19	0.05	-0.08	-0.03	0.08
RFHD:X	0.13	-0.08	0.13	-0.09	0.02	-0.02
RFHD:Y	0.12	0.12	-0.13	-0.08	-0.07	-0.02
RFHD:Z	0.09	-0.01	0.25	-0.01	0.08	0.05
LFHD:X	0.13	-0.07	0.13	-0.12	0.01	-0.02
LFHD:Y	0.12	0.1	-0.12	-0.06	-0.06	-0.03
LFHD:Z	0.08	0.01	0.2	-0.19	-0.02	0.04
LBHD:X	0.12	-0.07	0.13	-0.14	0.01	-0.01
LBHD:Y	0.12	0.11	-0.13	-0.08	-0.07	-0.02
LBHD:Z	0.05	0.1	0.08	-0.26	-0.09	0.07
RBHD:X	0.13	-0.07	0.13	-0.1	0.02	-0.02
RBHD:Y	0.11	0.13	-0.13	-0.1	-0.08	-0.01
RBHD:Z	0.08	0.02	0.22	0.09	0.1	0.08

Figure 11.5: Loadings for the First Six Principal Components, Performer 2, Prelude 6 with top ten loadings in the first component highlighted in red and the second component in blue

Marker	PC1	PC2	PC3	PC4	PC5	PC6
C7:X	0.13	-0.04	0.16	-0.01	0.03	0
C7:Y	0.15	-0.06	-0.08	0.01	0.05	0.06
C7:Z	0.02	-0.17	0.11	-0.13	-0.14	0.09
T10:X	0.14	0.02	0.13	0.02	0.09	0.01
T10:Y	0.15	-0.08	-0.08	-0.01	0.06	0.05
T10:Z	-0.01	-0.09	-0.18	-0.03	-0.1	0.21
CLAV:X	0.14	0	0.15	0.01	0.02	-0.01
CLAV:Y	0.16	-0.05	-0.08	0.01	0.01	0.04
CLAV:Z	0.01	-0.14	0.17	-0.13	-0.13	0.01
STRN:X	0.15	0.05	0.11	0.03	0.04	0
STRN:Y	0.16	-0.05	-0.08	0	-0.02	0.03
STRN:Z	0.03	-0.15	0.18	-0.11	-0.09	0
LSHO:X	0.13	-0.05	0.16	-0.01	0.09	0
LSHO:Y	0.16	-0.05	-0.08	0.03	0	0.06
LSHO:Z	0.09	-0.17	0.1	0	-0.04	0.09
LUPA:X	0.12	-0.02	0.15	-0.09	0.22	0
LUPA:Y	0.16	-0.04	-0.08	0	-0.04	0.04
LUPA:Z	0.06	-0.18	0.08	0.14	-0.04	0.11
LUPB:X	0.08	-0.03	-0.08	-0.03	0.04	-0.03
LUPB:Y	0.15	-0.03	-0.11	0	-0.11	-0.01
LUPB:Z	0.04	-0.03	0.18	-0.01	0.04	0.07
LUPC:X	0.11	0.04	0.11	-0.18	0.26	-0.01
LUPC:Y	0.16	-0.04	-0.09	-0.03	-0.04	0.01
LUPC:Z	0.07	-0.17	0.07	0.14	-0.06	0.15
LELB:X	0.09	0.11	0.04	-0.26	0.19	0.01
LELB:Y	0.15	-0.01	-0.1	-0.07	-0.08	-0.06
LELB:Z	0.05	-0.18	0.08	0.17	-0.1	0.14
LMEP:X	0.1	0.08	0.07	-0.23	0.24	0
LMEP:Y	0.15	-0.02	-0.08	-0.04	-0.13	-0.02
LMEP:Z	0.03	-0.18	0.09	0.21	-0.06	0.12
LWRA:X	0.12	0.11	0.02	-0.2	-0.04	-0.05
LWRA:Y	0.15	-0.01	-0.08	-0.04	-0.14	-0.11
LWRA:Z	-0.06	-0.17	0.03	0.16	0.13	-0.18
LWRB:X	0.11	0.12	0.02	-0.23	-0.01	-0.01
LWRB:Y	0.14	-0.02	-0.08	-0.03	-0.15	-0.11
LWRB:Z	-0.04	-0.18	0.04	0.18	0.12	-0.16
LFRA:X	0.1	0.11	0.04	-0.25	0.15	-0.01
LFRA:Y	0.15	-0.01	-0.08	-0.04	-0.14	-0.07
LFRA:Z	0	-0.19	0.08	0.23	0.01	0.02
LFIN:X	0.12	0.11	0.02	-0.19	-0.06	-0.02
LFIN:Y	0.14	-0.02	-0.07	-0.02	-0.16	-0.12
LFIN:Z	-0.05	-0.16	0.01	0.15	0.2	-0.17
RSHO:X	0.13	0.01	0.16	0.01	-0.02	-0.04
RSHO:Y	0.16	-0.05	-0.07	0.01	0.04	0.06
RSHO:Z	-0.1	-0.07	0.11	-0.16	-0.19	0
RUPA:X	0.12	0.06	0.17	0.06	-0.03	-0.11
RUPA:Y	0.15	-0.07	-0.09	0.02	0.03	0.02
RUPA:Z	-0.12	-0.11	0.04	-0.16	-0.15	-0.1
RUPB:X	0.11	0.08	0.16	0.08	-0.03	-0.13
RUPB:Y	0.15	-0.07	-0.1	0.04	0.03	-0.05
RUPB:Z	-0.08	-0.16	0.04	-0.2	-0.1	-0.03
RUPC:X	-0.03	-0.03	-0.03	-0.03	-0.13	-0.39
RUPC:Y	0.14	-0.06	-0.04	0.03	0.02	-0.12
RUPC:Z	0.05	0.05	0.03	0.04	0.1	0.32
RELB:X	0.07	0.15	0.13	0.13	-0.07	-0.17
RELB:Y	0.12	-0.09	-0.12	0.03	0.01	-0.17
RELB:Z	-0.09	-0.15	0.01	-0.2	-0.11	-0.06
RMEP:X	0.07	0.14	0.14	0.13	-0.05	-0.19
RMEP:Y	0.13	-0.09	-0.12	0.04	0.03	-0.15
RMEP:Z	-0.09	-0.16	0	-0.17	-0.06	-0.11
RWRA:X	0.09	0.14	0.1	0.13	-0.1	-0.15
RWRA:Y	0.05	-0.09	-0.21	-0.05	-0.02	-0.09
RWRA:Z	-0.05	-0.17	0.03	-0.04	0.24	-0.15
RWRB:X	0.08	0.16	0.1	0.11	-0.14	-0.12
RWRB:Y	0.05	-0.09	-0.21	-0.05	-0.02	-0.1
RWRB:Z	-0.06	-0.17	0.04	-0.04	0.23	-0.17
RFRA:X	0.07	0.15	0.12	0.12	-0.09	-0.16
RFRA:Y	0.1	-0.1	-0.16	0.01	0.01	-0.15
RFRA:Z	-0.08	-0.18	0.02	-0.12	0.08	-0.15
RFIN:X	0.09	0.16	0.09	0.12	-0.14	-0.13
RFIN:Y	0.02	-0.09	-0.21	-0.07	-0.03	-0.04
RFIN:Z	-0.05	-0.16	0.02	0	0.26	-0.2
RFHD:X	0.1	-0.06	0.16	-0.05	-0.08	0.04
RFHD:Y	0.16	-0.06	-0.06	0.02	0.05	0.07
RFHD:Z	-0.02	-0.16	0.1	-0.13	-0.2	0.03
LFHD:X	0.14	-0.05	0.14	-0.02	0	0.05
LFHD:Y	0.15	-0.06	-0.06	0.02	0.03	0.09
LFHD:Z	0.05	-0.18	0.11	-0.09	-0.1	0.06
LBHD:X	0.14	-0.05	0.12	0	0.05	0.02
LBHD:Y	0.16	-0.05	-0.07	0.02	0.06	0.06
LBHD:Z	0.04	-0.16	0.17	-0.04	0.01	-0.01
RBHD:X	0.11	-0.06	0.17	-0.03	-0.03	0.03
RBHD:Y	0.16	-0.05	-0.06	0.03	0.08	0.05
RBHD:Z	-0.06	-0.13	0.14	-0.1	-0.15	-0.07

Figure 11.6: Loadings for the First Six Principal Components, Performer 3, Prelude 6 with top ten loadings in the first component highlighted in red and the second component in blue

Appendix B
Extra weighted principal components graphs from Chapter 8

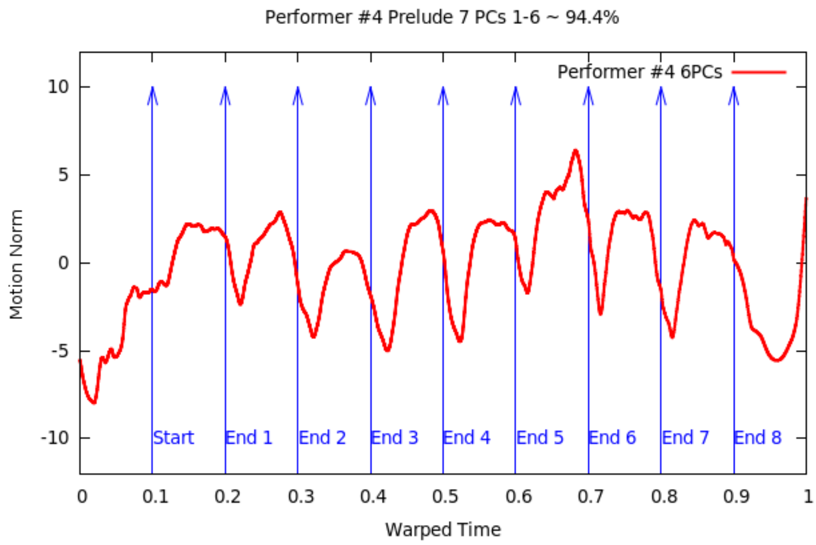


Figure 11.7: Weighted Principal Components for Performer 4, Prelude 7

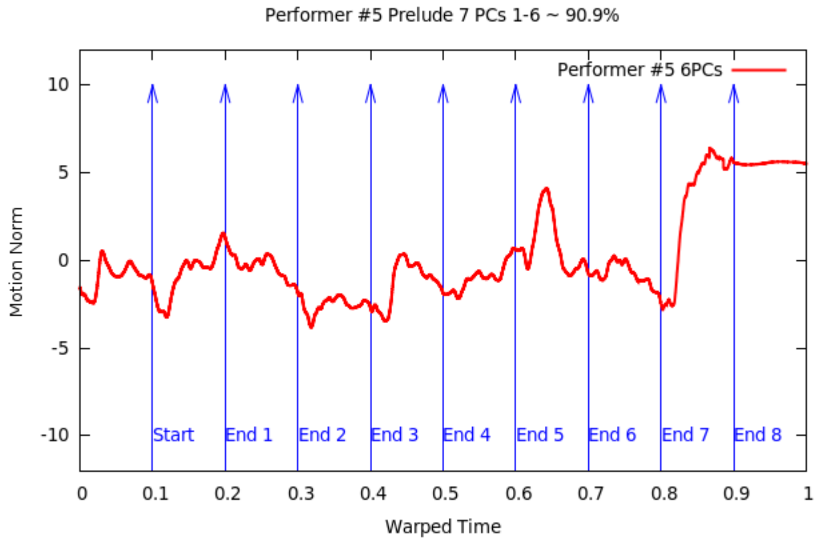


Figure 11.8: Weighted Principal Components for Performer 5, Prelude 7

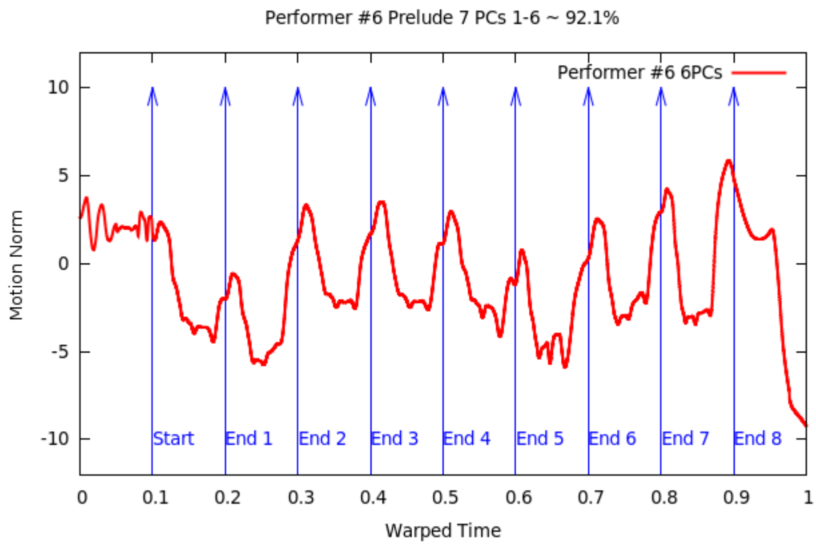


Figure 11.9: Weighted Principal Components for Performer 6, Prelude 7

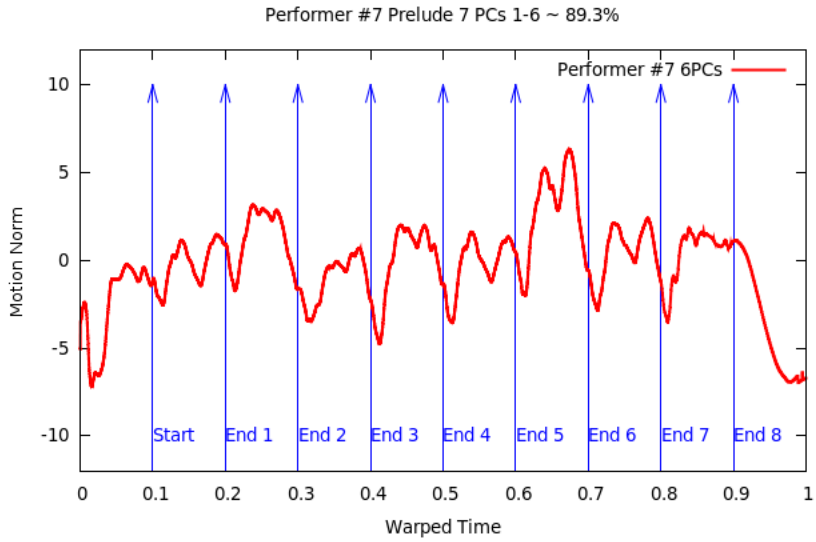


Figure 11.10: Weighted Principal Components for Performer 7, Prelude 7

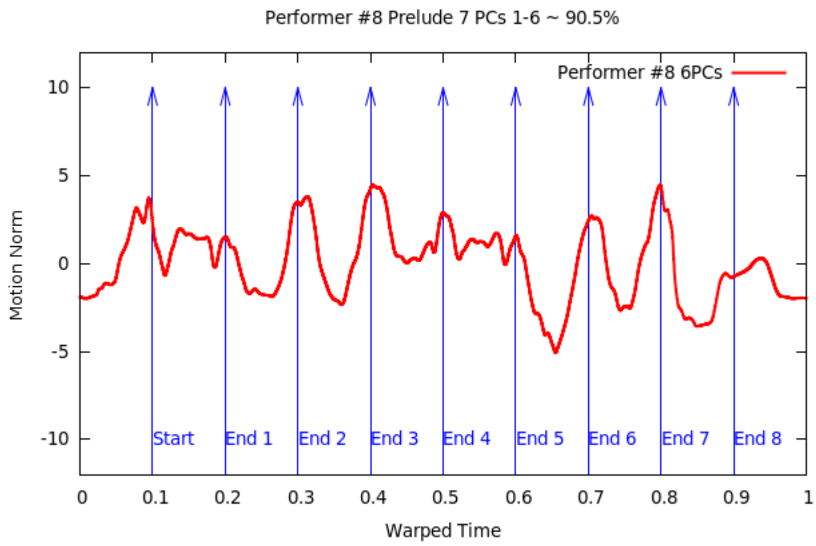


Figure 11.11: Weighted Principal Components for Performer 8, Prelude 7

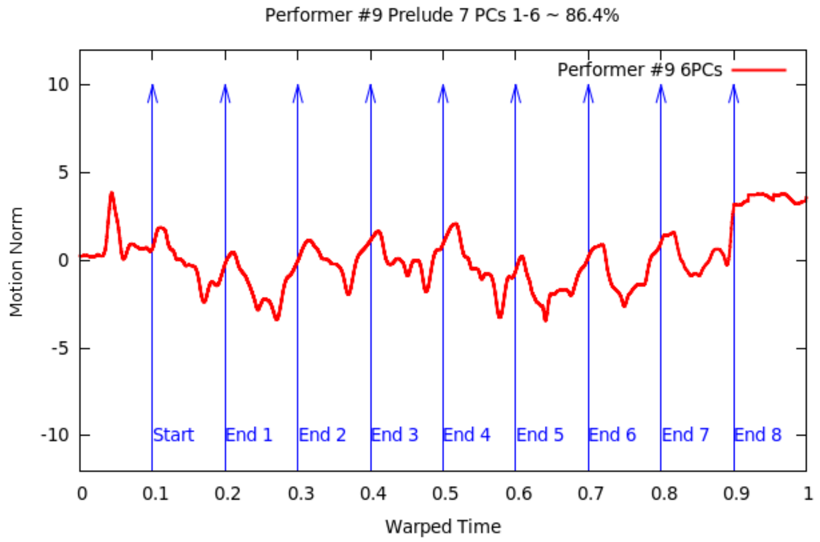


Figure 11.12: Weighted Principal Components for Performer 9, Prelude 7

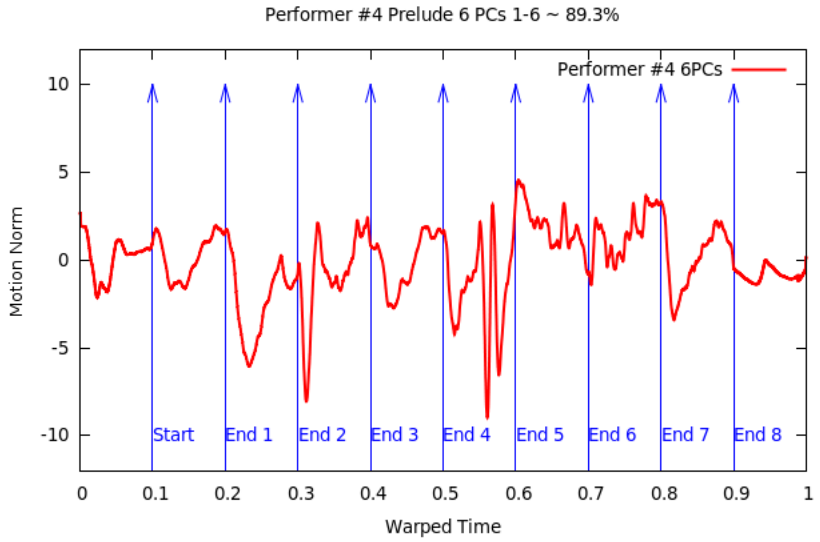


Figure 11.13: Weighted Principal Components for Performer 4, Prelude 6

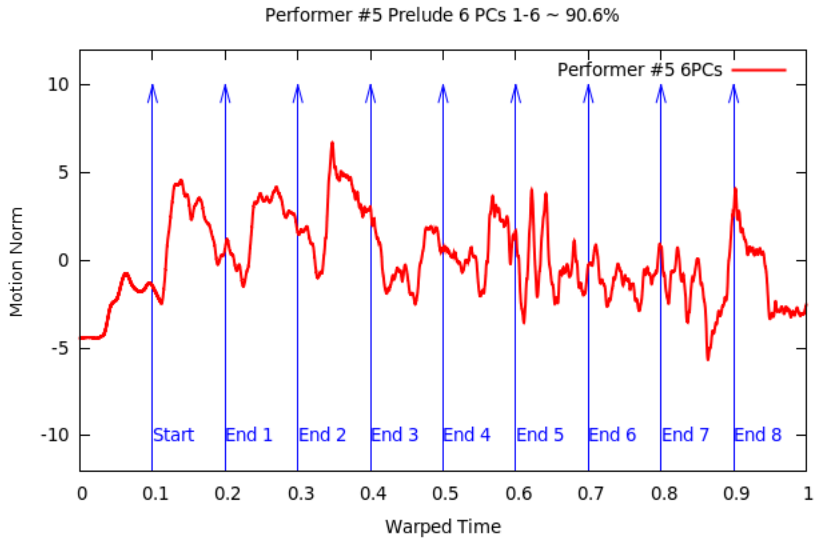


Figure 11.14: Weighted Principal Components for Performer 5, Prelude 6

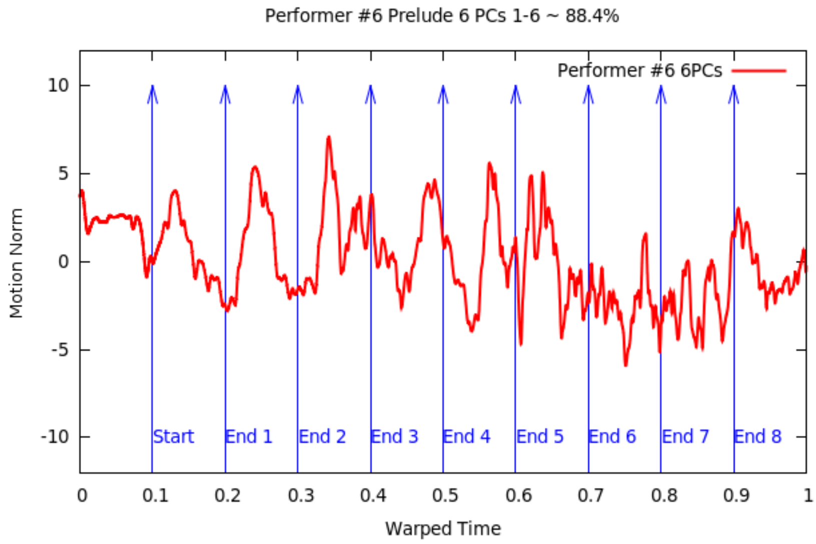


Figure 11.15: Weighted Principal Components for Performer 6, Prelude 6

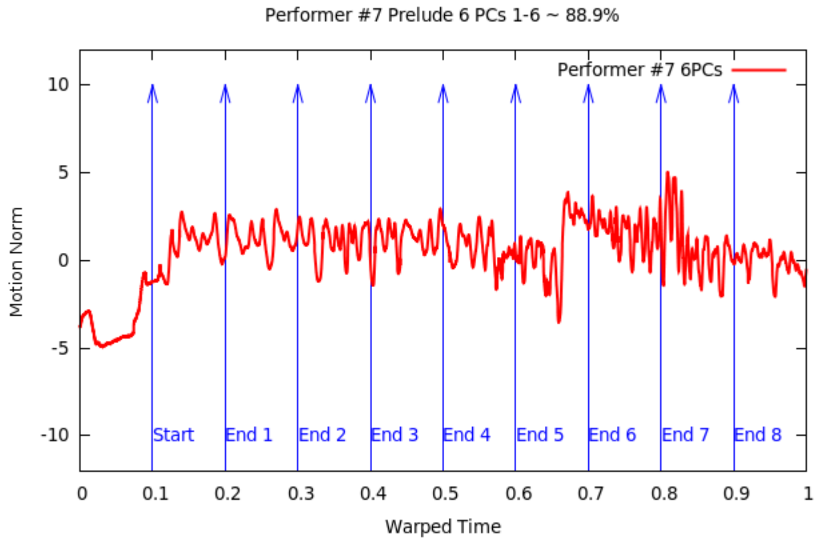


Figure 11.16: Weighted Principal Components for Performer 7, Prelude 6

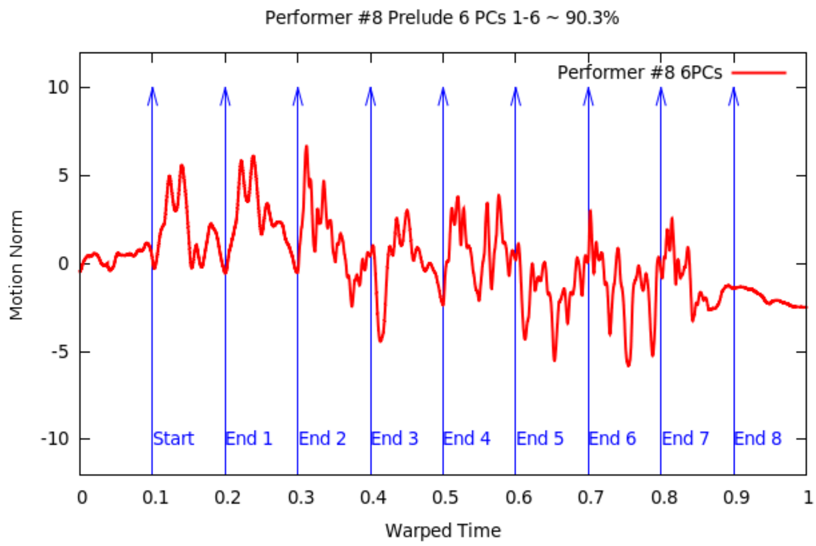


Figure 11.17: Weighted Principal Components for Performer 8, Prelude 6

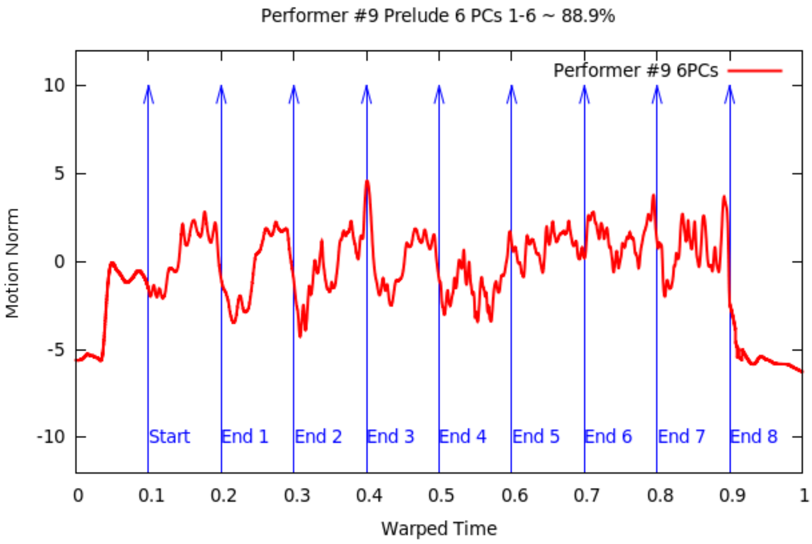


Figure 11.18: Weighted Principal Components for Performer 9, Prelude 6

Appendix C
Extra multi-modal graphs from Chapter 8

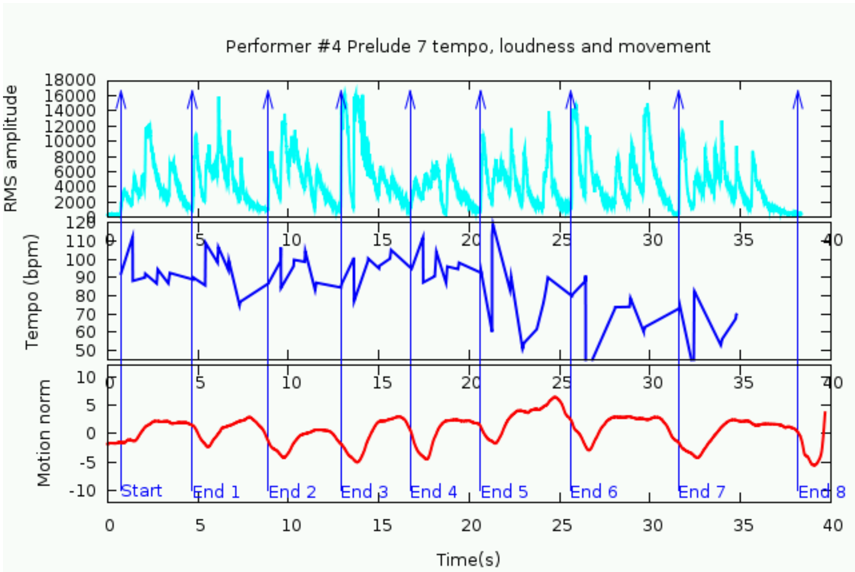


Figure 11.19: Motion, Tempo and Dynamics for Performer 4 , Prelude 7

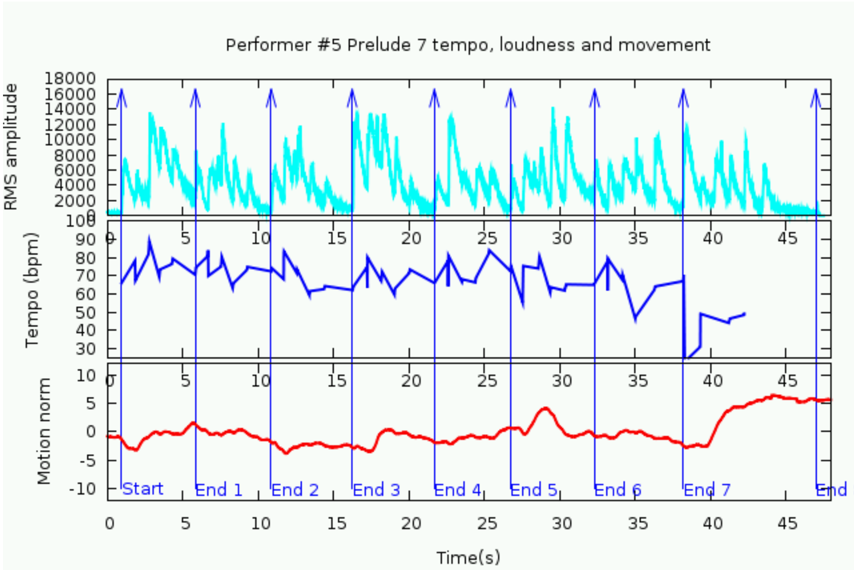


Figure 11.20: Motion, Tempo and Dynamics for Performer 5 , Prelude 7

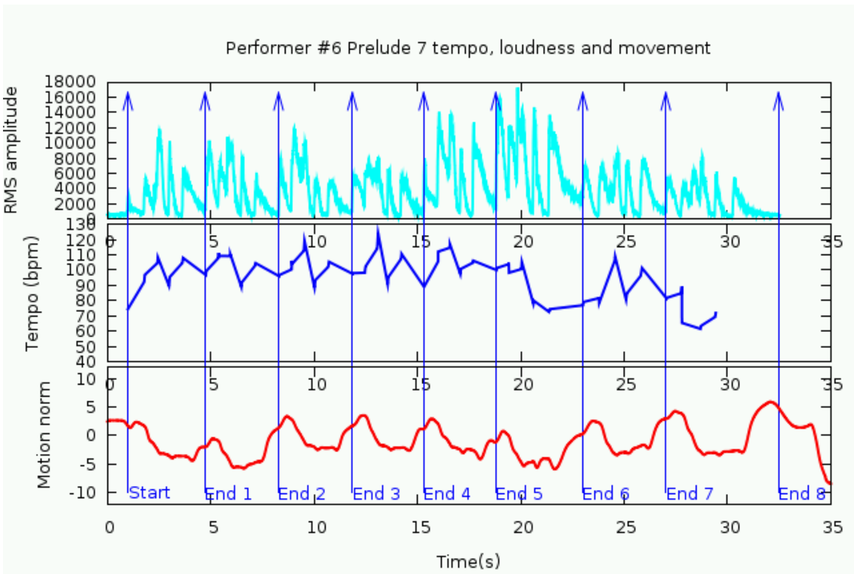


Figure 11.21: Motion, Tempo and Dynamics for Performer 6 , Prelude 7

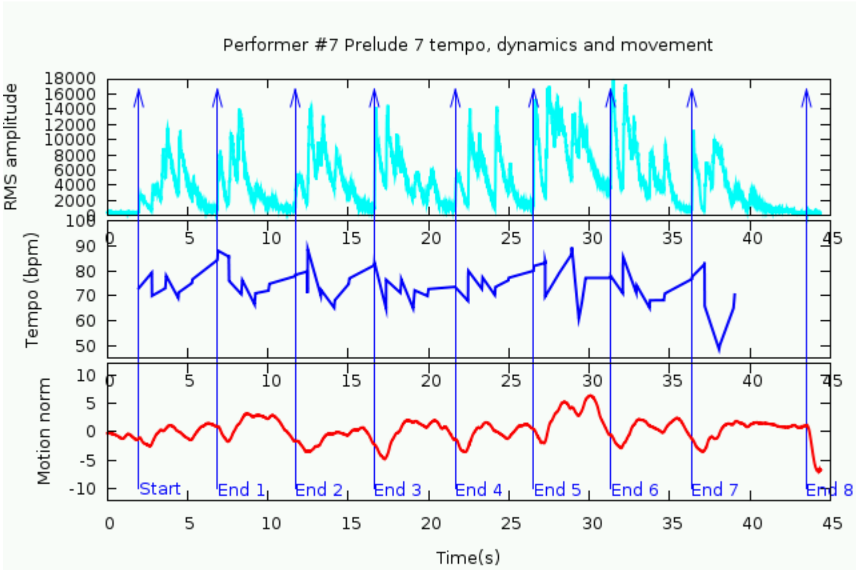


Figure 11.22: Motion, Tempo and Dynamics for Performer 7 , Prelude 7

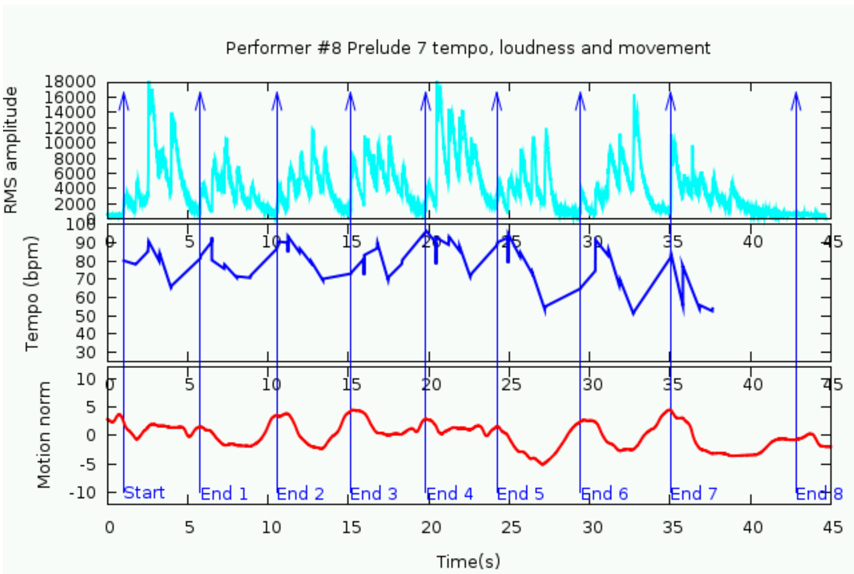


Figure 11.23: Motion, Tempo and Dynamics for Performer 8 , Prelude 7

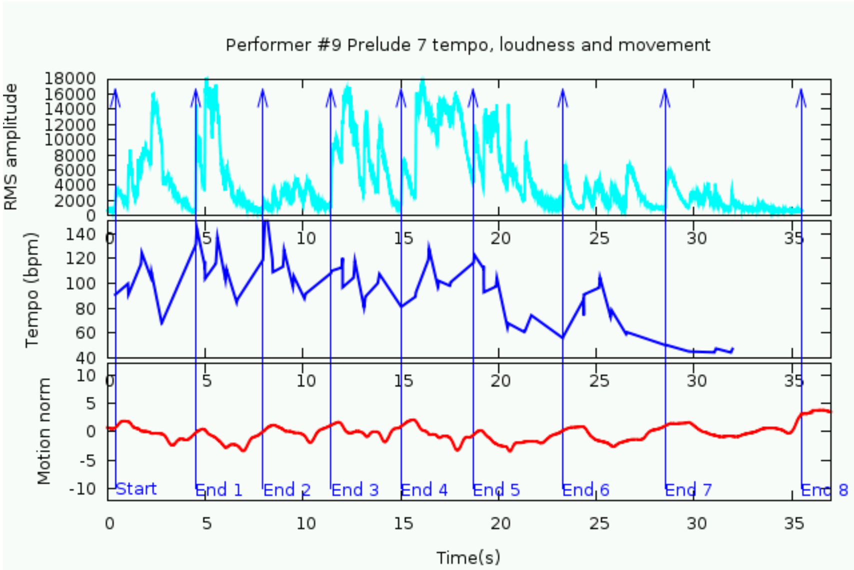


Figure 11.24: Motion, Tempo and Dynamics for Performer 9 , Prelude 7

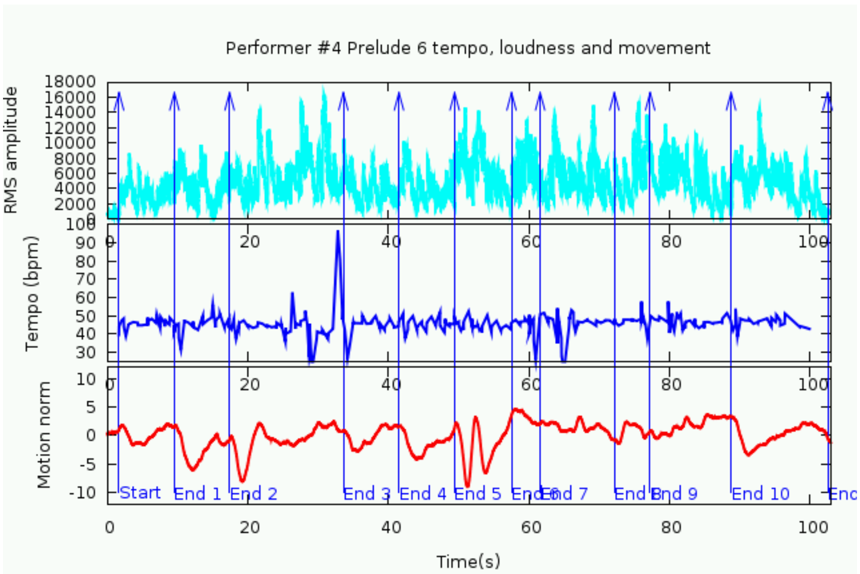


Figure 11.25: Motion, Tempo and Dynamics for Performer 4 , Prelude 6

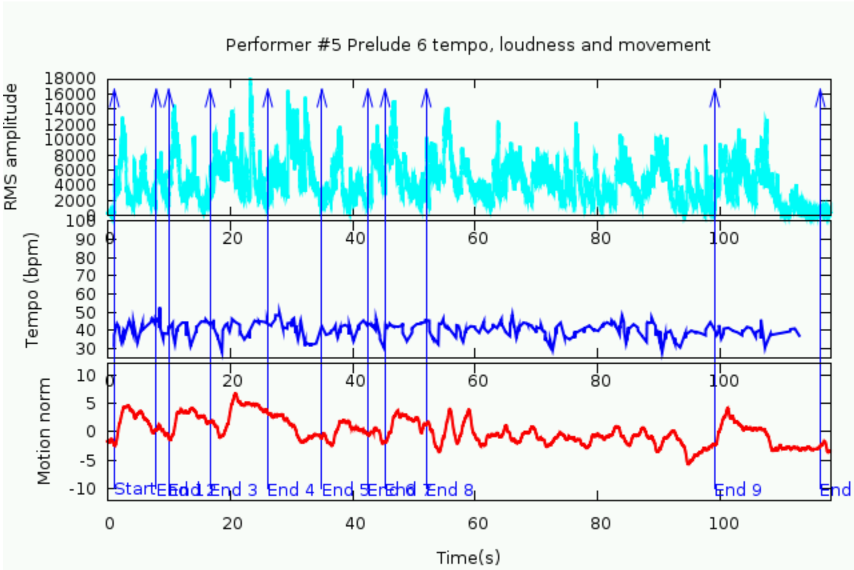


Figure 11.26: Motion, Tempo and Dynamics for Performer 5 , Prelude 6

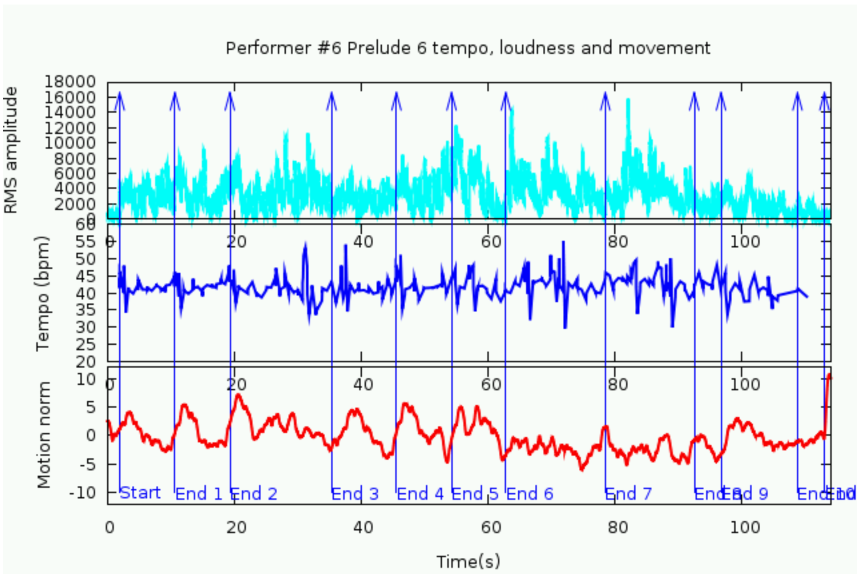


Figure 11.27: Motion, Tempo and Dynamics for Performer 6 , Prelude 6

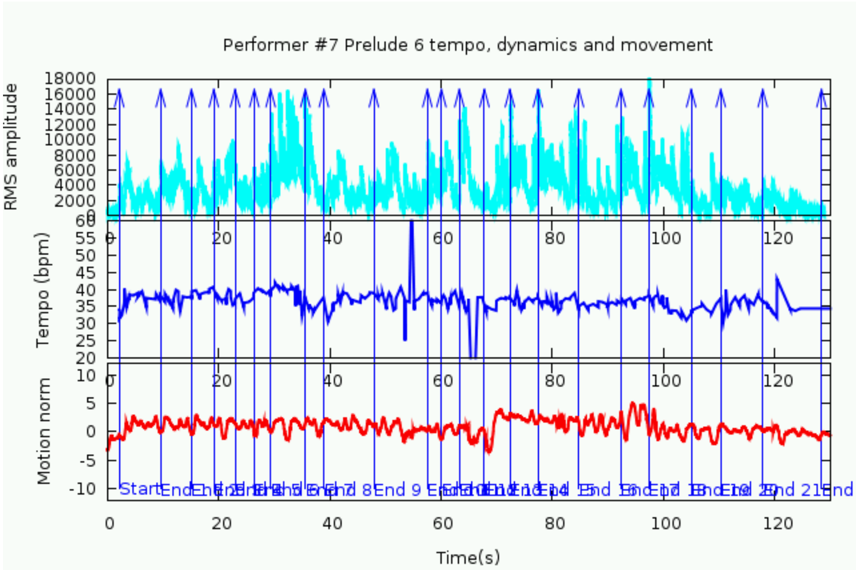


Figure 11.28: Motion, Tempo and Dynamics for Performer 7 , Prelude 6

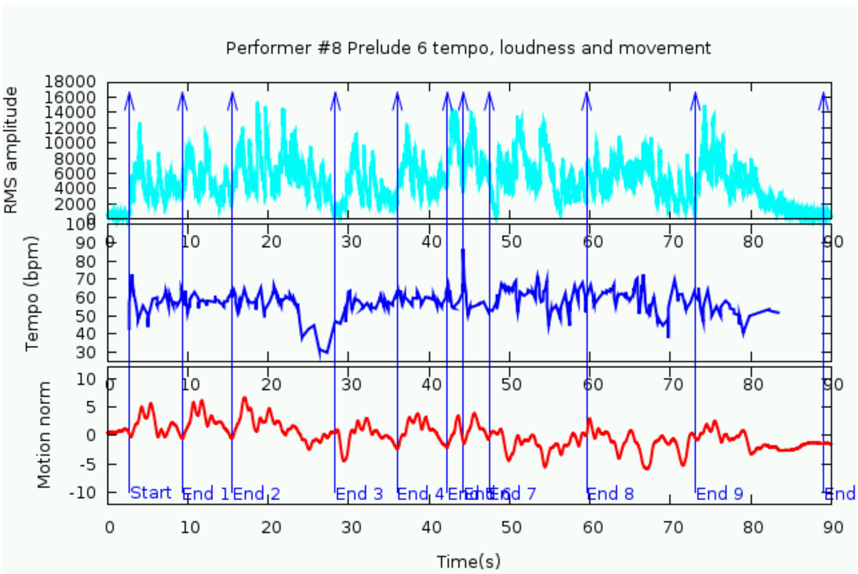


Figure 11.29: Motion, Tempo and Dynamics for Performer 8 , Prelude 6

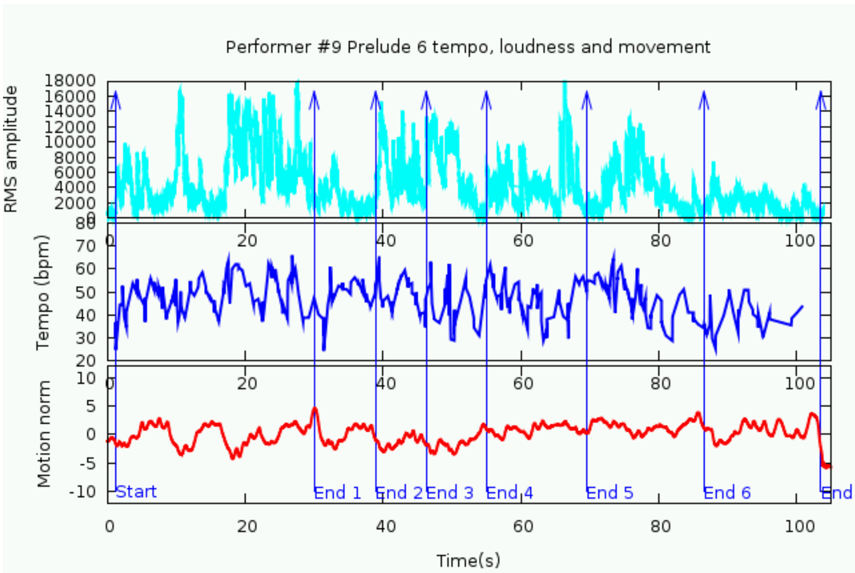


Figure 11.30: Motion, Tempo and Dynamics for Performer 9 , Prelude 6

Appendix D
Extra multi-modal graphs from Chapter 9

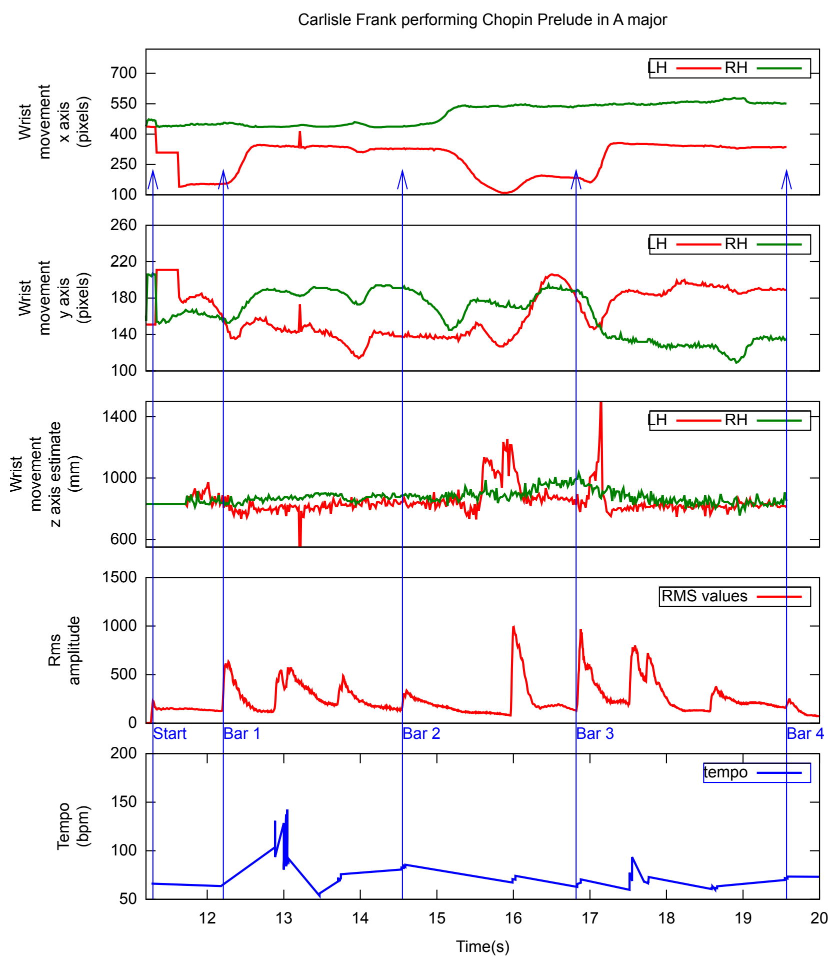


Figure 11.31: Wrist Motion, Tempo and Dynamics for Carlisle Frank, Prelude in A Major

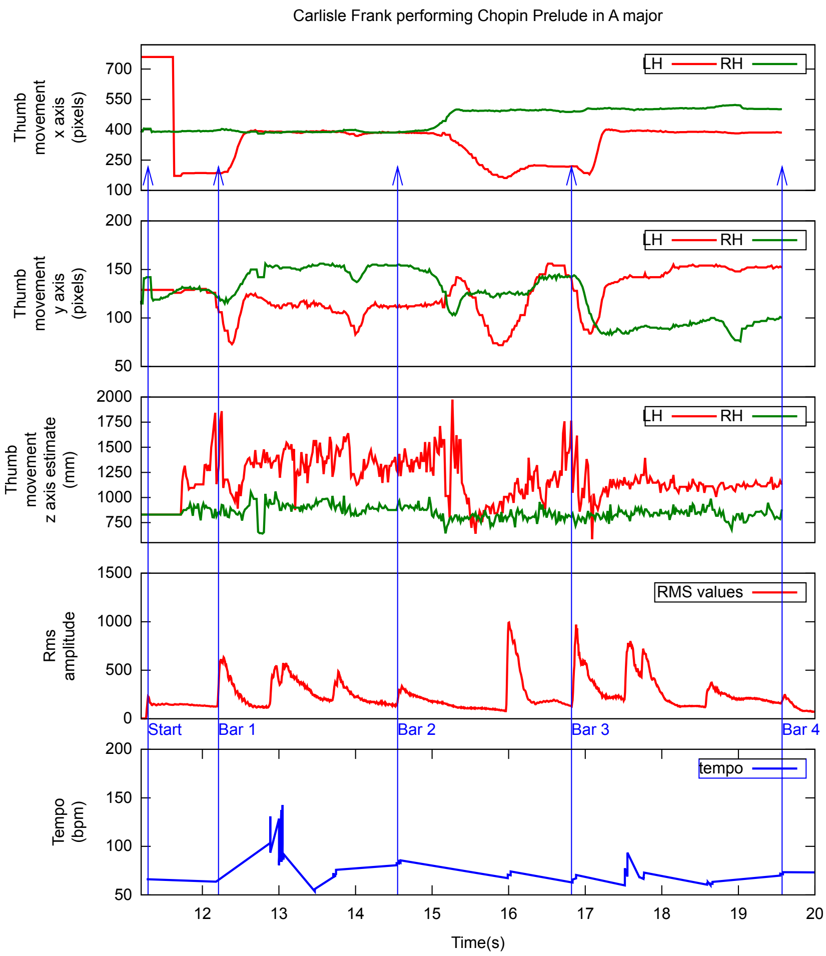


Figure 11.32: Thumb Motion, Tempo and Dynamics for Carlisle Frank, Prelude in A Major

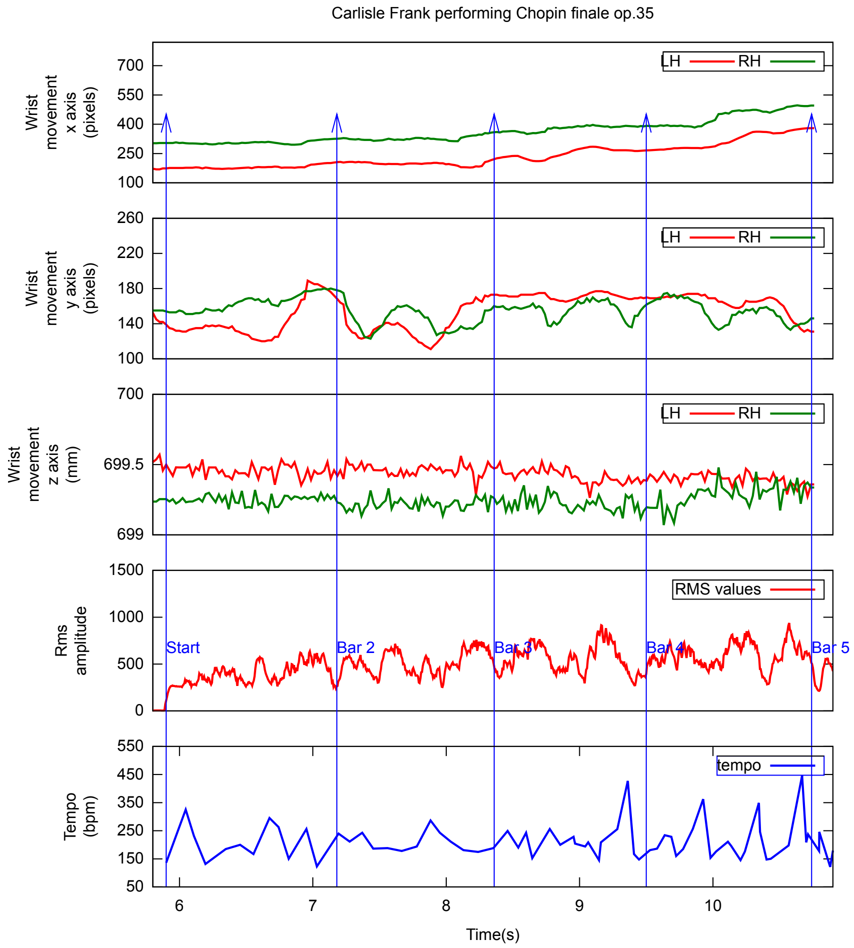


Figure 11.33: Wrist Motion, Tempo and Dynamics for Carlisle Frank, performing the Chopin finale

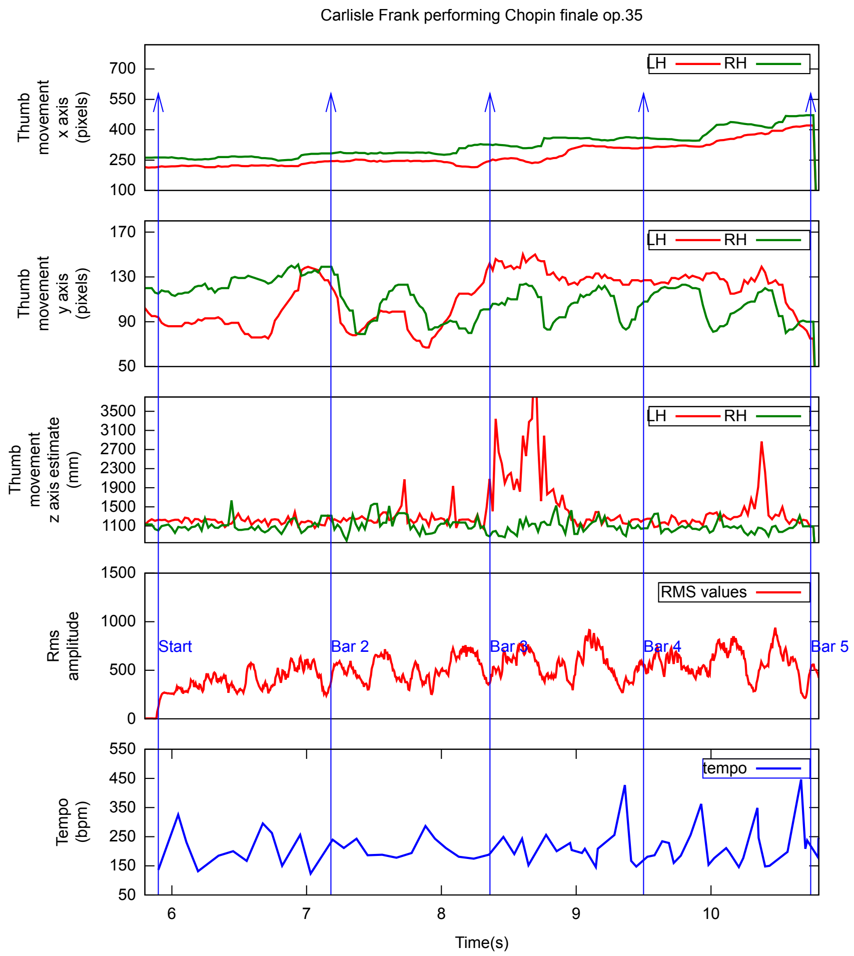


Figure 11.34: Thumb Motion, Tempo and Dynamics for Carlisle Frank, performing the Chopin finale

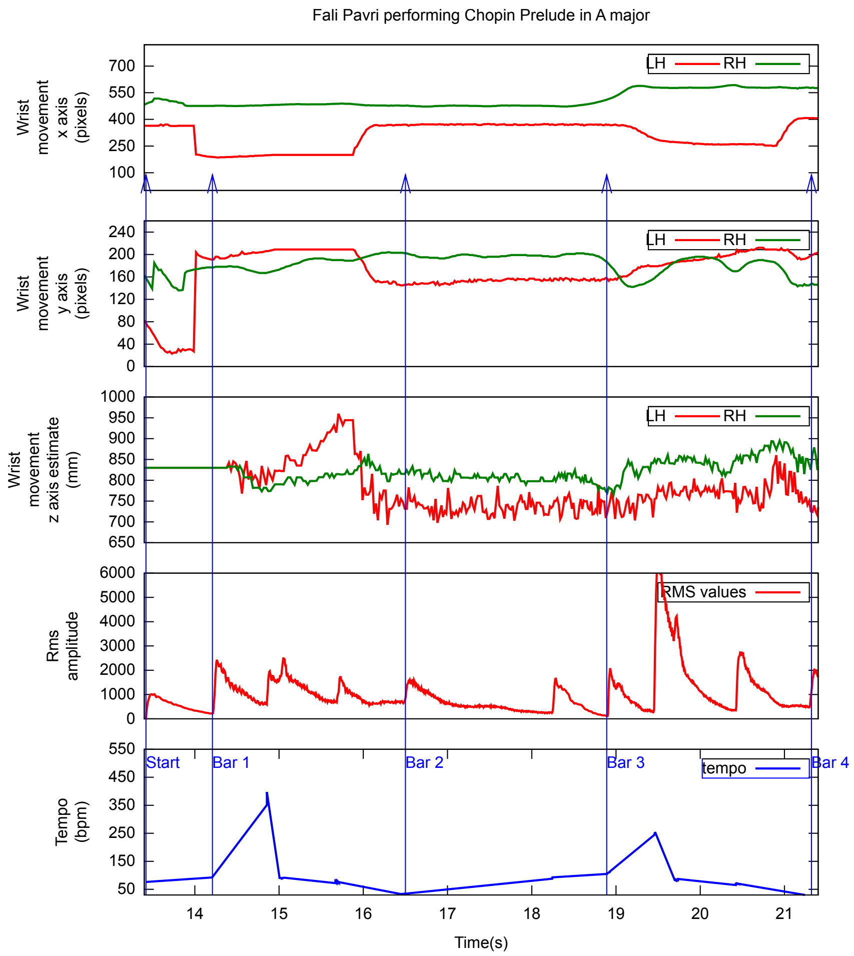


Figure 11.35: Wrist Motion, Tempo and Dynamics for Fali Pavri, Prelude in A Major

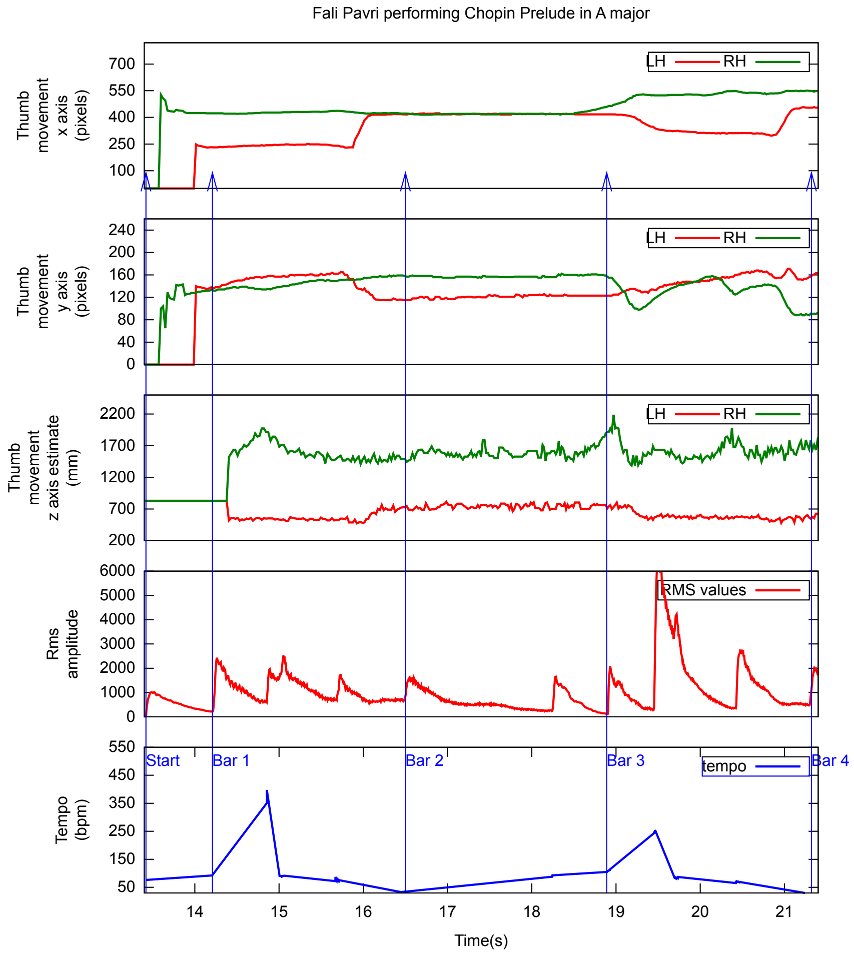


Figure 11.36: Thumb Motion, Tempo and Dynamics for Fali Pavri, Prelude in A Major

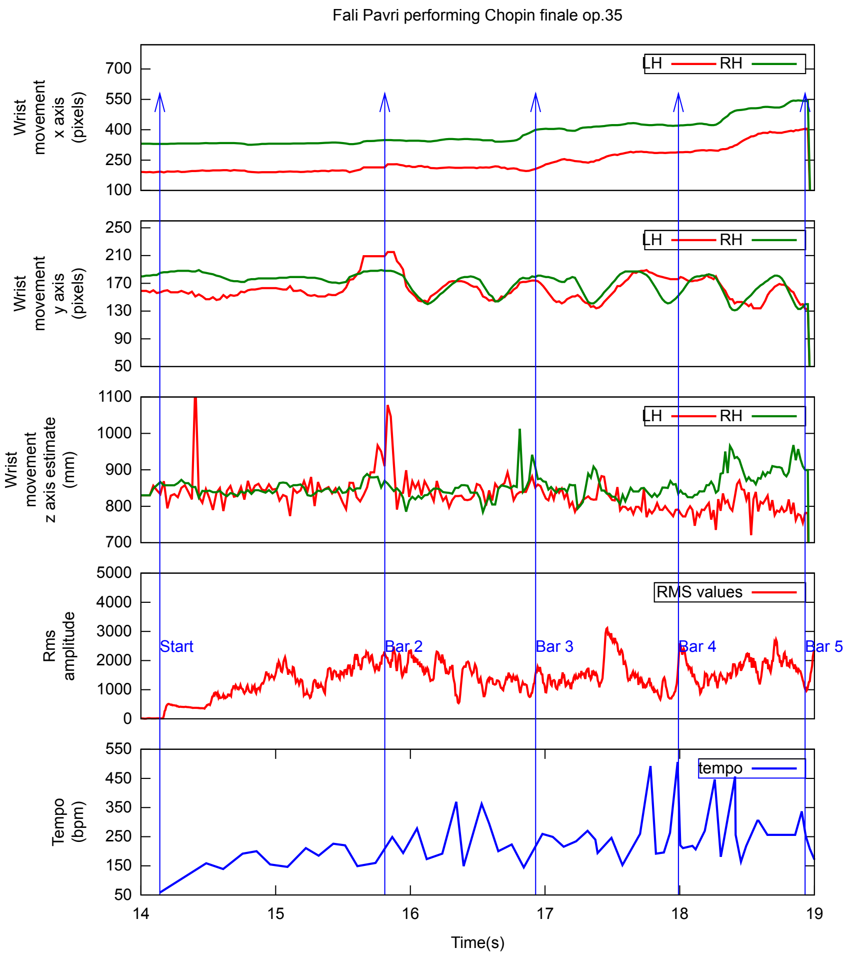


Figure 11.37: Wrist Motion, Tempo and Dynamics for Fali Pavri, performing the Chopin finale

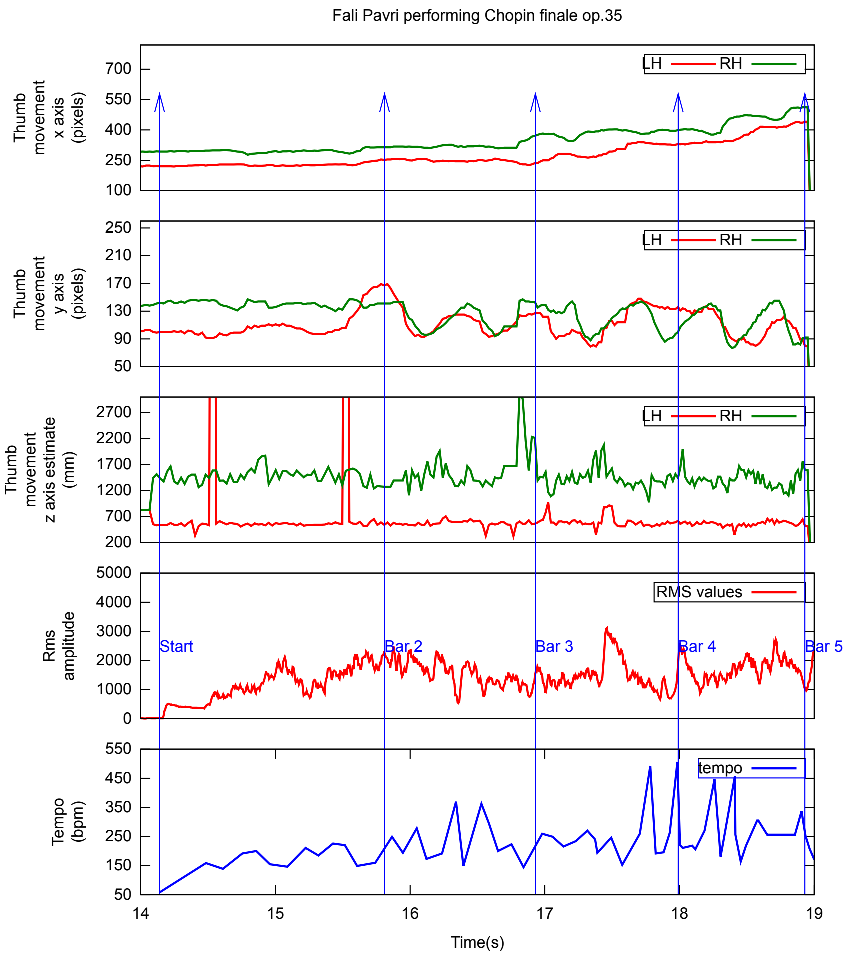


Figure 11.38: Thumb Motion, Tempo and Dynamics for FPavri, performing the Chopin finale

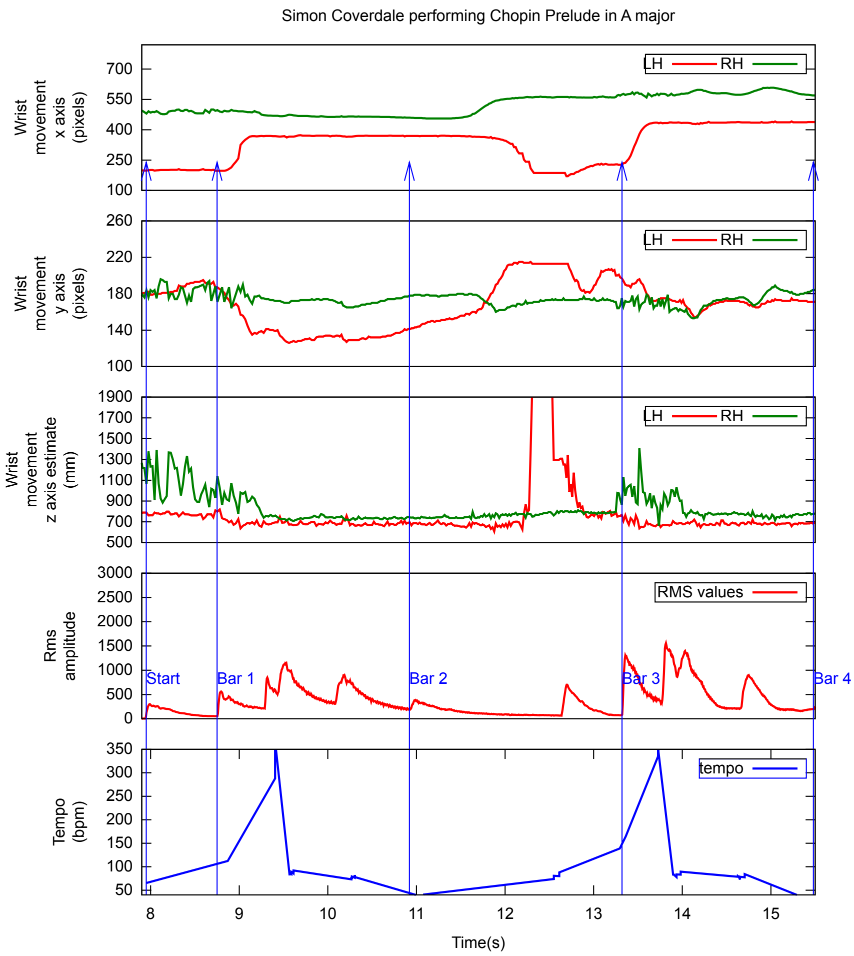


Figure 11.39: Wrist Motion, Tempo and Dynamics for Simon Coverdale, Prelude in A Major

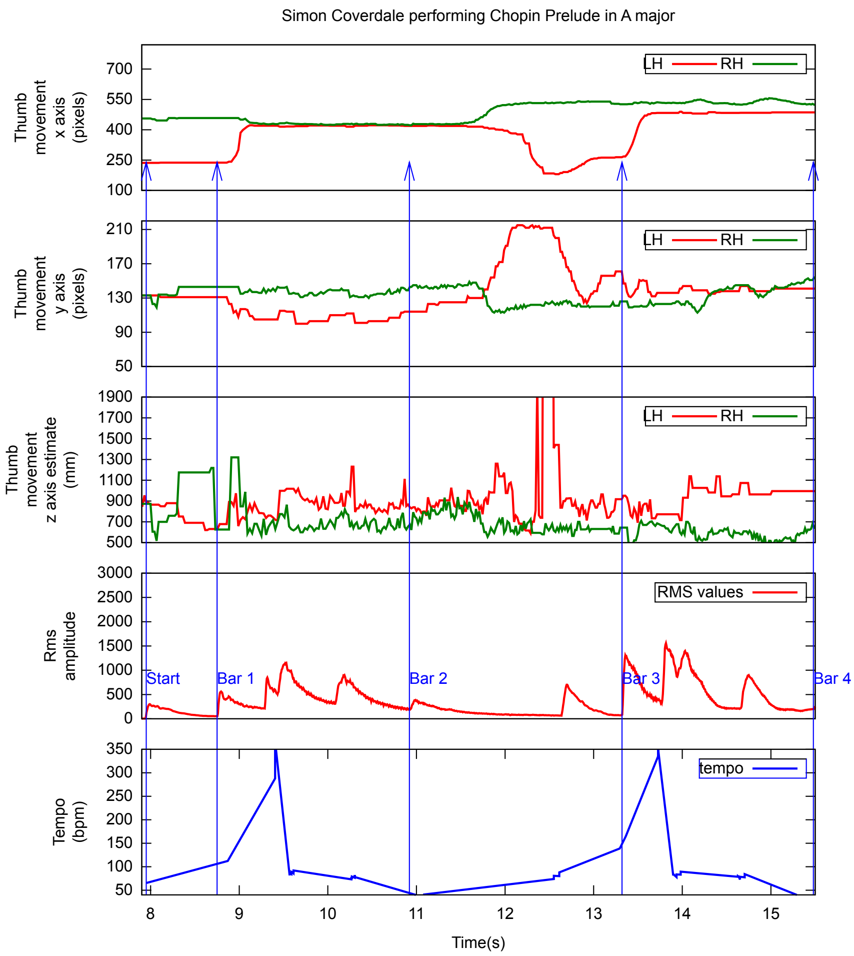


Figure 11.40: Thumb Motion, Tempo and Dynamics for Simon Coverdale, Prelude in A Major

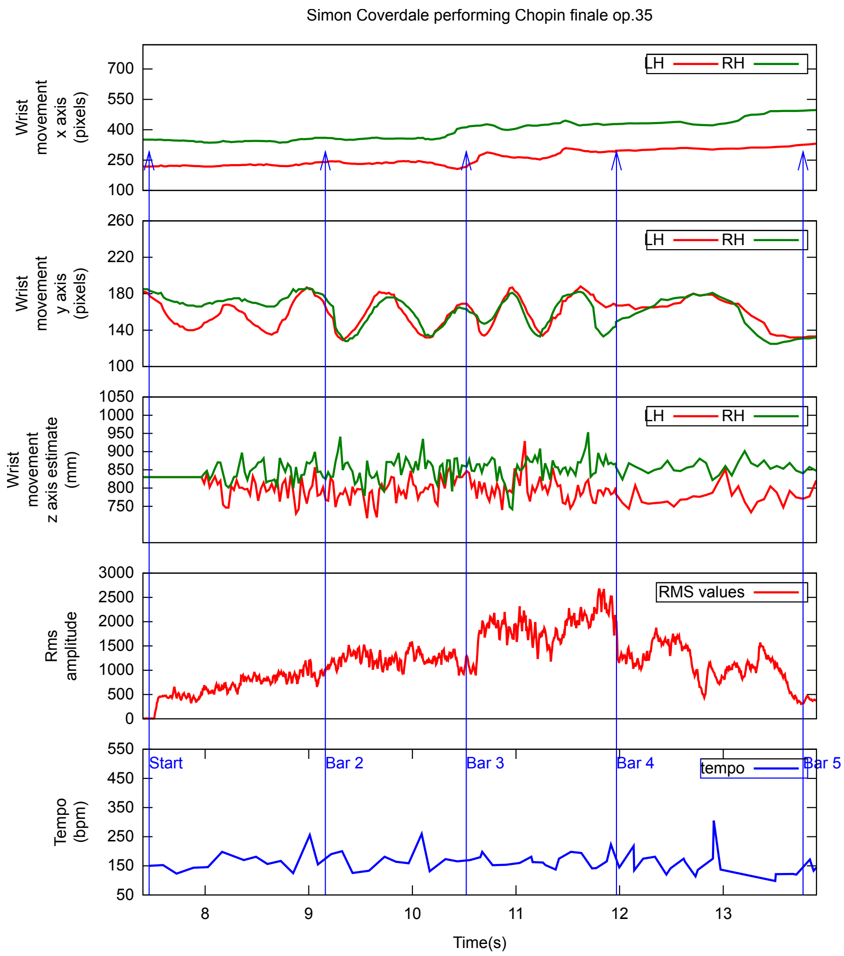


Figure 11.41: Wrist Motion, Tempo and Dynamics for Simon Coverdale, performing the Chopin finale

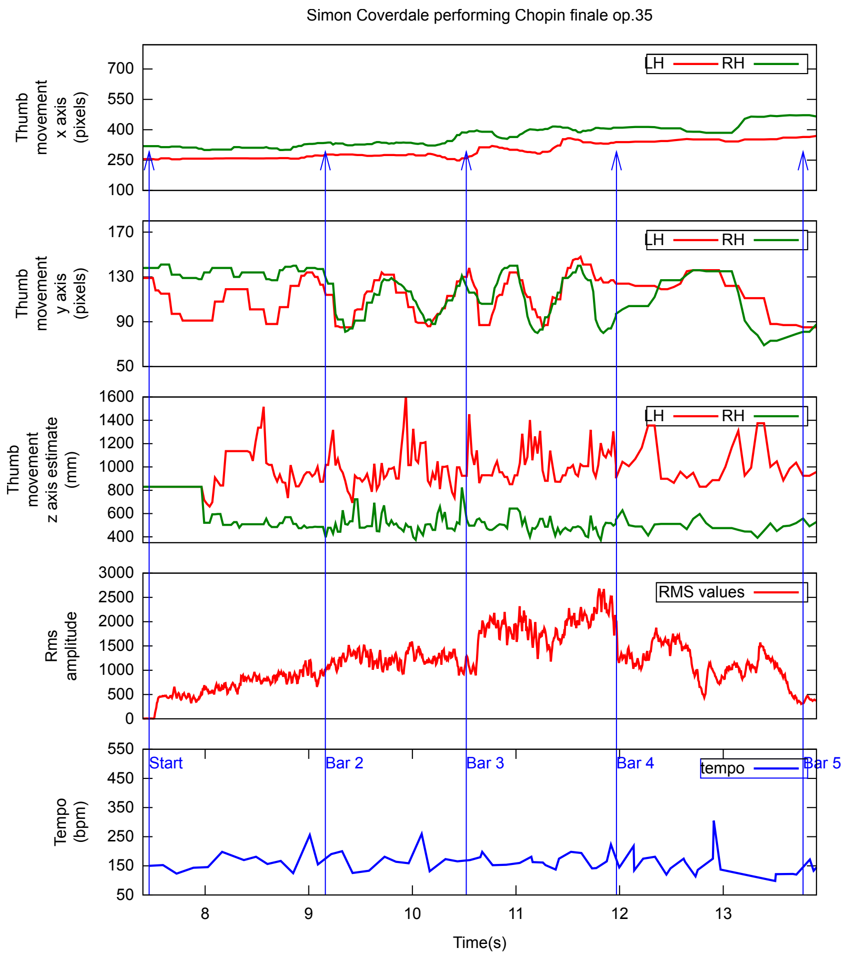


Figure 11.42: Thumb Motion, Tempo and Dynamics for Simon Coverdale, performing the Chopin finale

Appendix E
Extra finale database results from Chapter 9

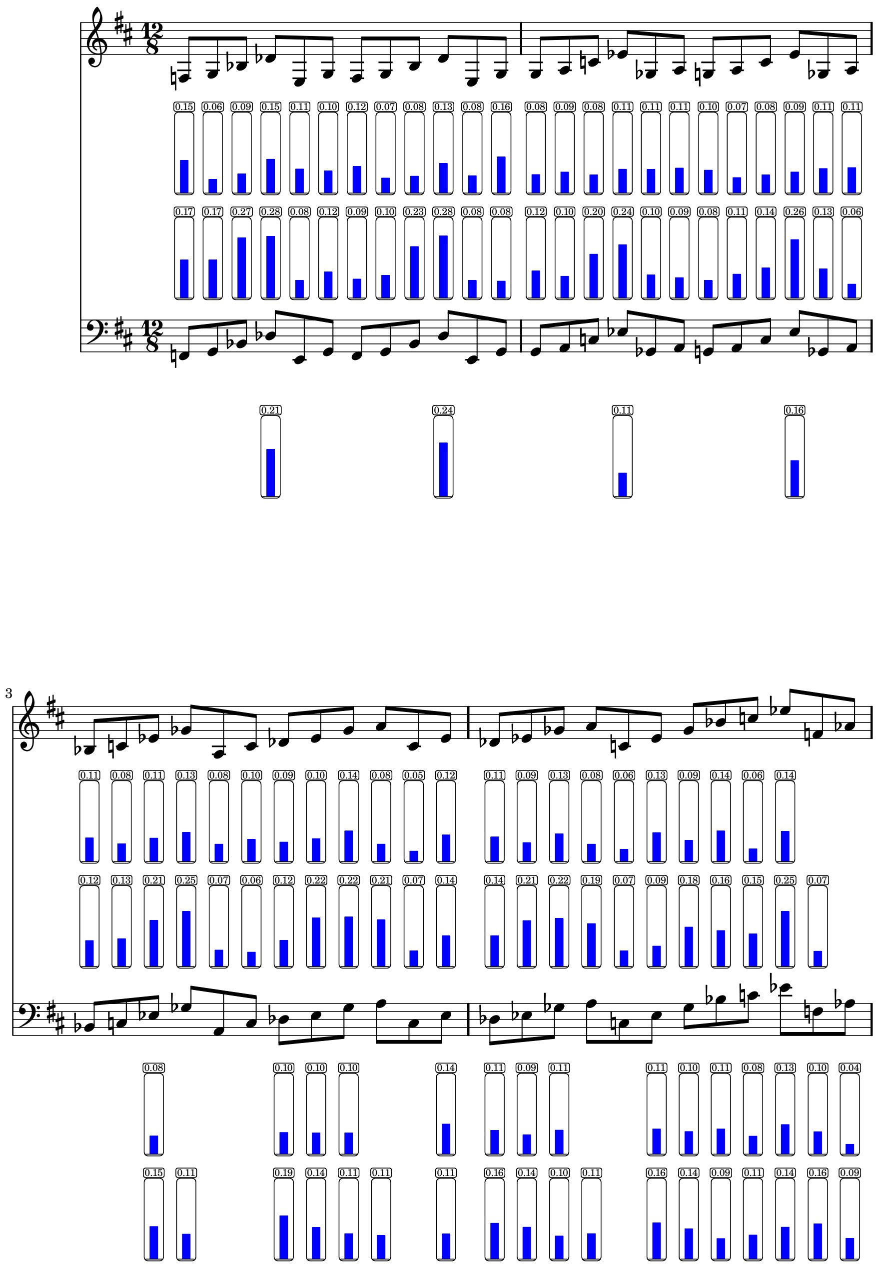


Figure 11.43: Database Results Page 1 for Carlisle Frank, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

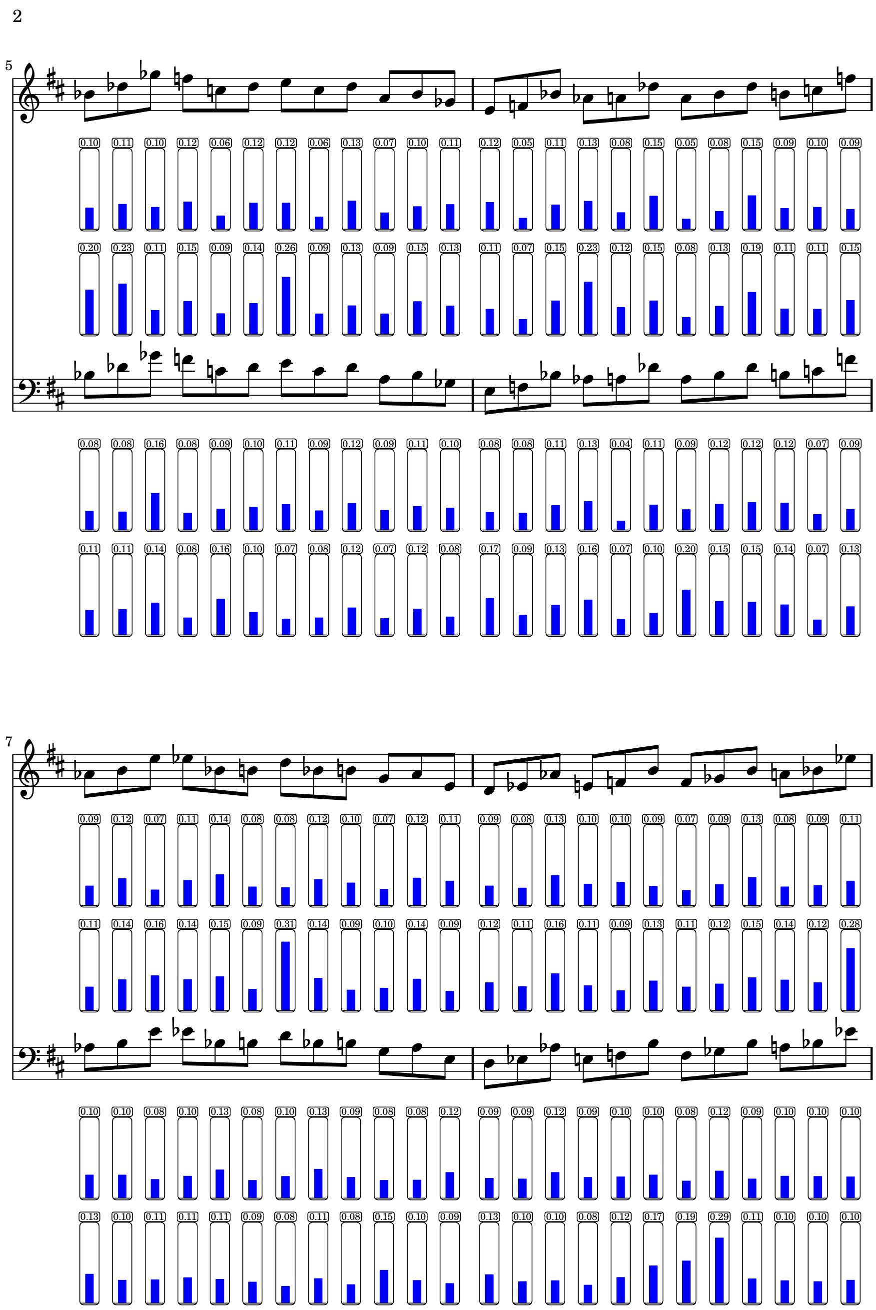


Figure 11.44: Database Results Page 2 for Carlisle Frank, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

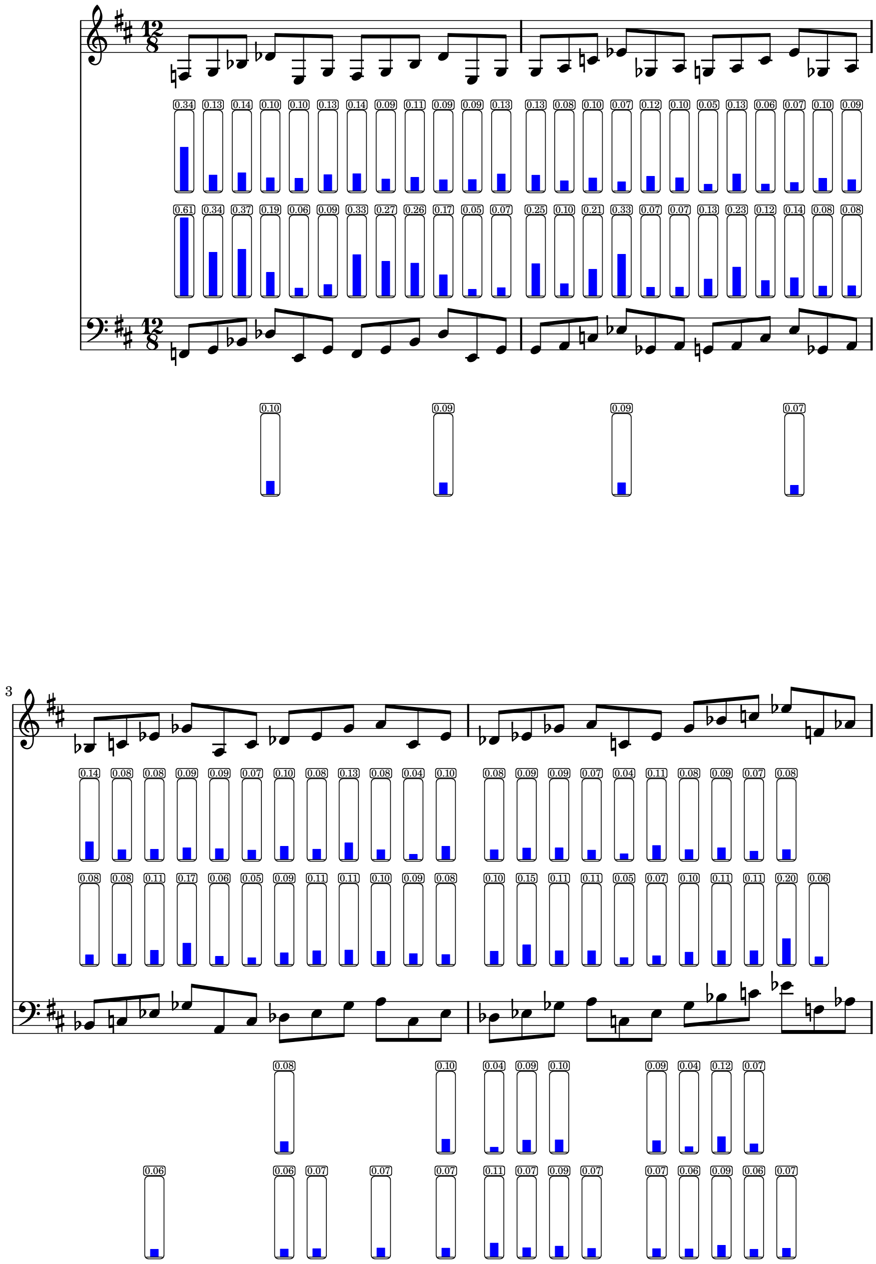


Figure 11.45: Database Results Page 1 for Fali Pavri, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

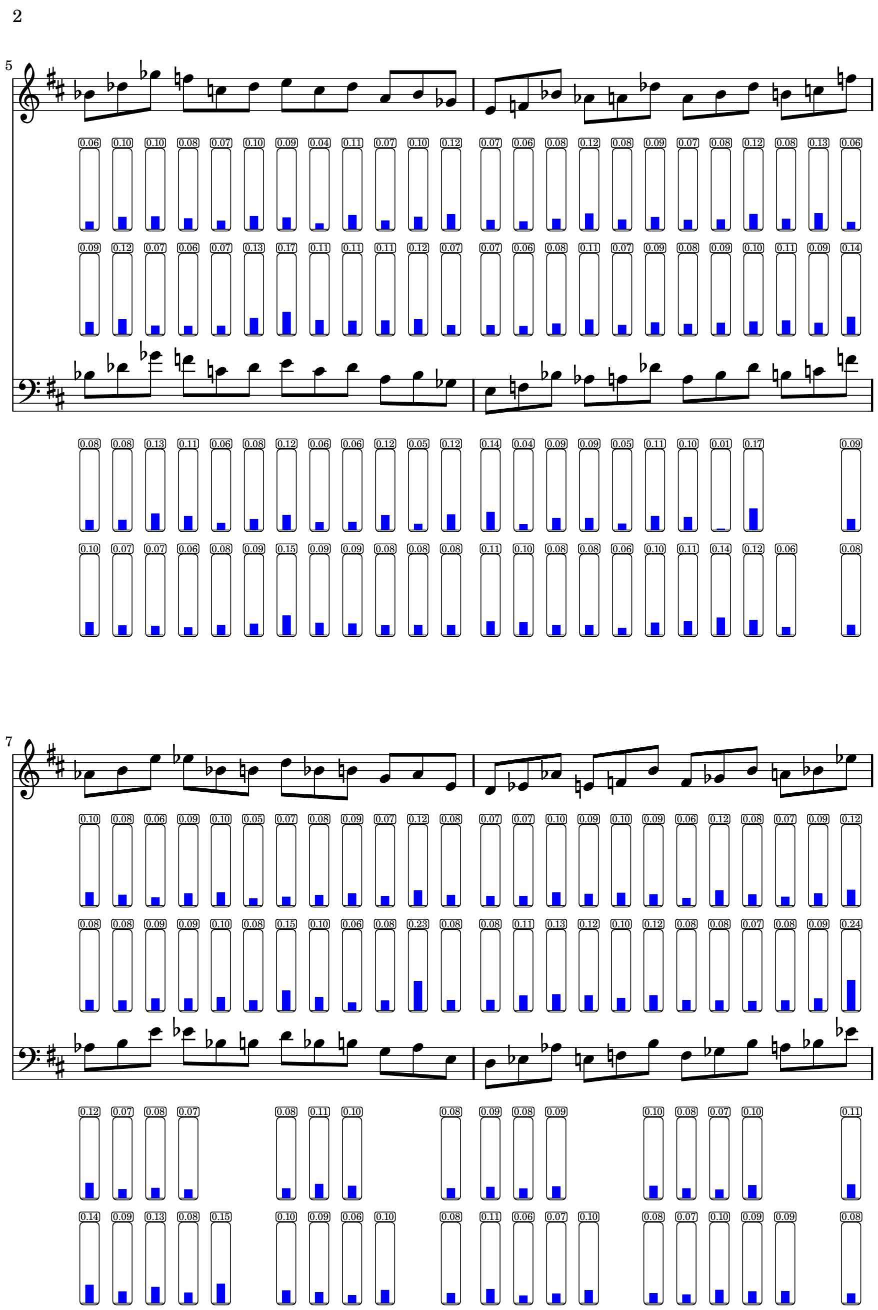


Figure 11.46: Database Results Page 2 for Fali Pavri, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

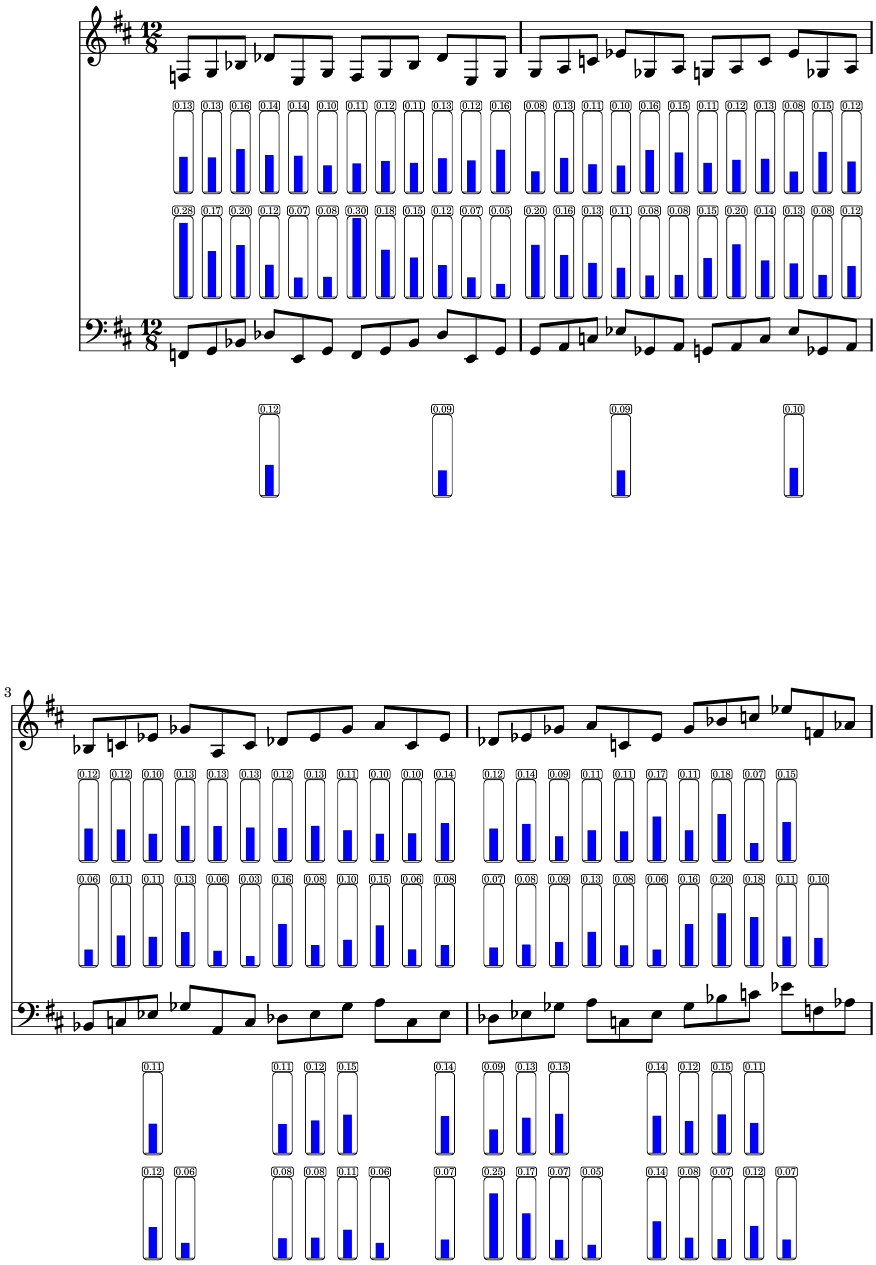


Figure 11.47: Database Results Page 1 for Simon Coverdale, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

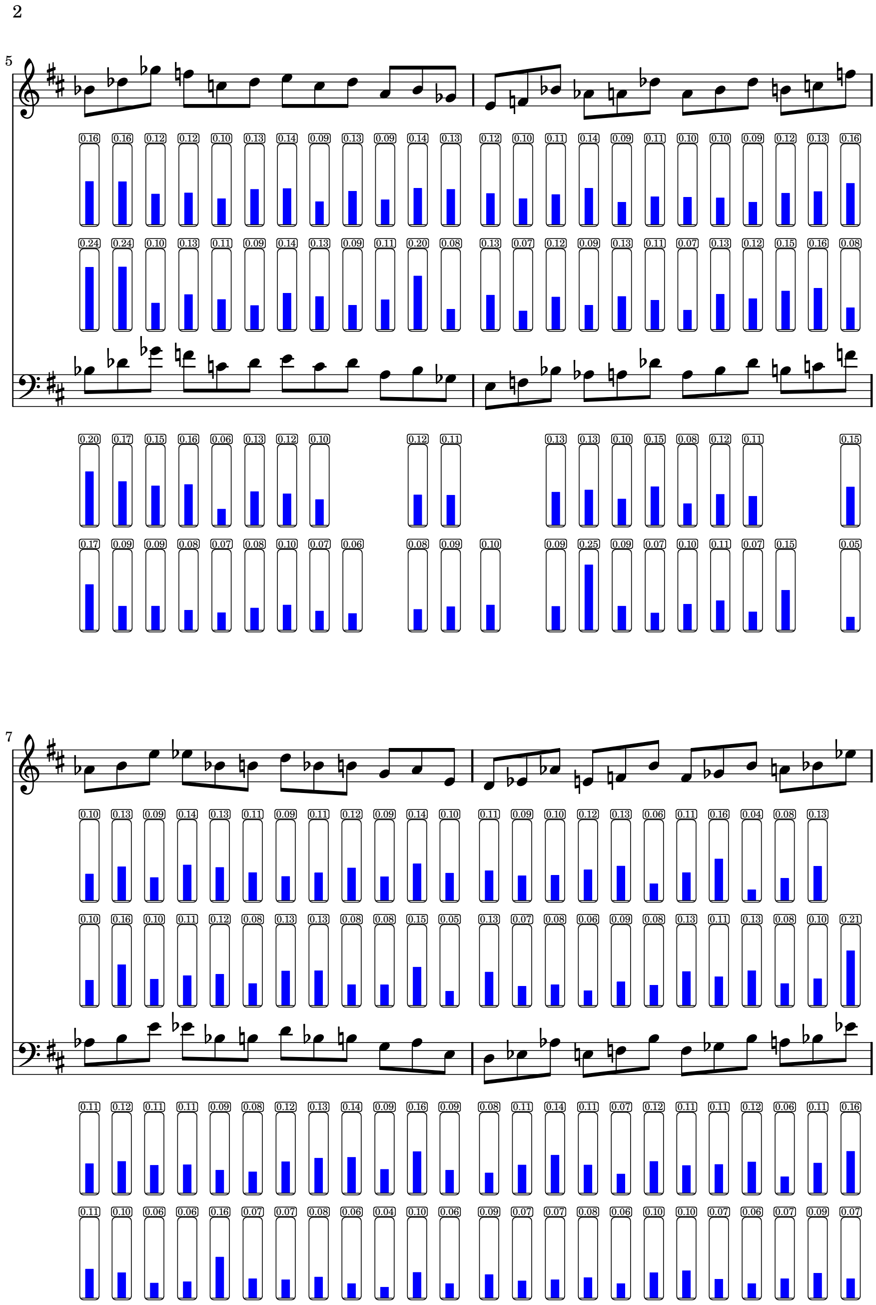


Figure 11.48: Database Results Page 2 for Simon Coverdale, B Flat minor Sonata finale, the first row of columns detailing inter-onset intervals and the second row of columns detailing the keypress durations

Appendix F
Publications Arising From Work Described Herein

MacRitchie, J. Bailey, N.J. And Hair, G.(2006) "Multi-modal Acquisition of Performance Parameters for Analysis of Chopin's B Flat Minor Piano Sonata Finale Op.35" DMRN+1 Workshop, Queen Mary University, London.

MacRitchie, J., Hair, G., Bailey, N.J and Pullinger, S. (2008) "Extracting Musical Structure from Multi-Modal Performance Analysis" Conference for Interdisciplinary Musicology, Thessaloniki, Greece.

Buck, B. MacRitchie, J and Piwek, L (2009) "Perceptual Recognition of Embodied Musical Structure" The Musical Body: Gesture, Representation and Ergonomics in Musical Performance, London, UK.

MacRitchie, J. Buck, B and Bailey, N.J (2009)"Visualising Musical Structure through Performance Gesture" International Society for Music Information Retrieval, Kobe, Japan.

MacRitchie, J. ,Pullinger, S. , Bailey, N.J. And Hair, G. (2009) "Communication of phrasing structure through multi-modal expressive techniques in piano performance" International Conference on Music Communication Science, Sydney, Australia.

MacRitchie, J. Buck, B and Bailey, N.J (2009) “Gestural Communication: Linking the Multi- Modal Analysis of Performance to Perception of Musical Structure” International Symposium on Performance Science, Auckland, New Zealand.

MacRitchie, J., Bailey, N.J, and Hair, G. (2010) “Highlighting Structural Issues in Piano Performance with Optical Finger Tracking.” International Conference on Music and Gesture, Montreal, Canada.

Buck, B., Bailey, B., MacRitchie, J. and Parncutt, R. (2010) “Performers’ Body Motion and Phrase Structure: The role of velocity magnitude” International Conference on Music and Gesture, Montreal, Canada.